Predicting Wind Farm Performance Using Machine Learning

Srinikaeth Thirugnana Sambandam (SUID: 06123236, nikaeth@stanford.edu)
Jason Hu (SUID: 06106924, jhu7@stanford.edu)
Yuran Zhang (SUID: 05884547, yuranz4@stanford.edu)

1. Background & Motivation

As one of the most rapidly growing renewable energy resources, wind energy possesses significant values in economic, environmental and societal aspects. This has led to collaborative efforts to increase US electricity supplied from wind from 4.5% in 2013 to 10% by 2020, 20% by 2030, and 35% by 2050. (Wiser et al. 2015) This growing trend, therefore, requires new tools and technologies to optimize wind power generation and grid integration.

One key challenge regarding the utilization of wind energy is the uncertainty and intermittency associated with wind power because wind energy depends on fluctuating atmospheric conditions such as wind speed, temperature, humidity, etc. This makes it challenging to integrate wind power into the grid while maintaining the balance between power supply and power demand. One solution to this is the development of reliable wind speed/power forecasting models. Besides reducing operating costs and improving grid integration reliability, such models are also useful in effectively reducing the number of sensors need to be deployed at different locations.

Physical approach is a common method for wind farm performance prediction, which is based on physical principles for conservation of mass, momentum, and energy in air flows. However, physics-based wind farm simulation not only requires extensive physical data of the terrain including orography, roughness, obstacles etc., but is also extremely computational expensive. In addition, the complex inter-turbine interaction is hard to be accounted for. Physics-based wind farm simulation is usually advantageous in very short-term horizon where the influence of atmospheric dynamics becomes more important. (Jung et al. 2014)

An alternative approach to physical methods is to use data-driven approach, which requires historical data to train a statistical model and then perform wind speed/power predictions. The data-driven approach circumvents the complex physics and has a much lower computational cost. Various machine learning methods have been tested by researchers. Treiber et al. 2016 used simulated wind turbine power production data based on real-world wind measurements available at NREL and used different regression techniques including linear regression, k-nearest neighbors and support vector regression to predict performance both for individual turbines and for entire wind parks. Spatial correlations between turbines were captured by adding the wind power data of four neighboring turbines of a target turbine to the input dataset of the time series model. Liu et al. (2014) developed a hybrid model that preprocessed one-month wind speed signal of a wind farm in North China with WT (wavelet transform) to remove random fluctuation, then used SVM to forecast the WT-treated wind speed approximation. Liu et al. (2016) used nine machine learning models to predict wind power generation using the data from 7 wind farms in Ontario, Canada. The relative performance of the various machine learning methods were evaluated and compared, showing that SVM had the best overall performance, k-NN works for longer ahead predictions, and that deep learning methods showed potential for more abstract predictions such as spatial correlation predictions.

In this work, we explored different wind farm prediction schemes to not only provide power prediction at a time horizon ahead, but also effectively reduce the number of sensors that need to be deployed at a wind farm by using a subset of turbines to predict the behavior of the others. Specifically, we applied linear regression and non-linear machine learning algorithms to dig the information contained in historical wind speed or power time-series data and train the model to forecast future power generation. In addition, we used the wind speed /power data at multiple turbines to train a regression /machine learning model, so that the data at one turbine could be used to predict the performance of other turbines present in the wind farm. Different training algorithms and prediction approaches were compared and discussed.

2. Data and Data Preprocessing

The data used in this analysis is obtained from La Haute Borne Wind Park in France. The wind park is located approximately 300 km east of Paris and 150 km south of Luxembourg as shown in Figure 1. Specific information regarding the turbines are shown in the Table 1.

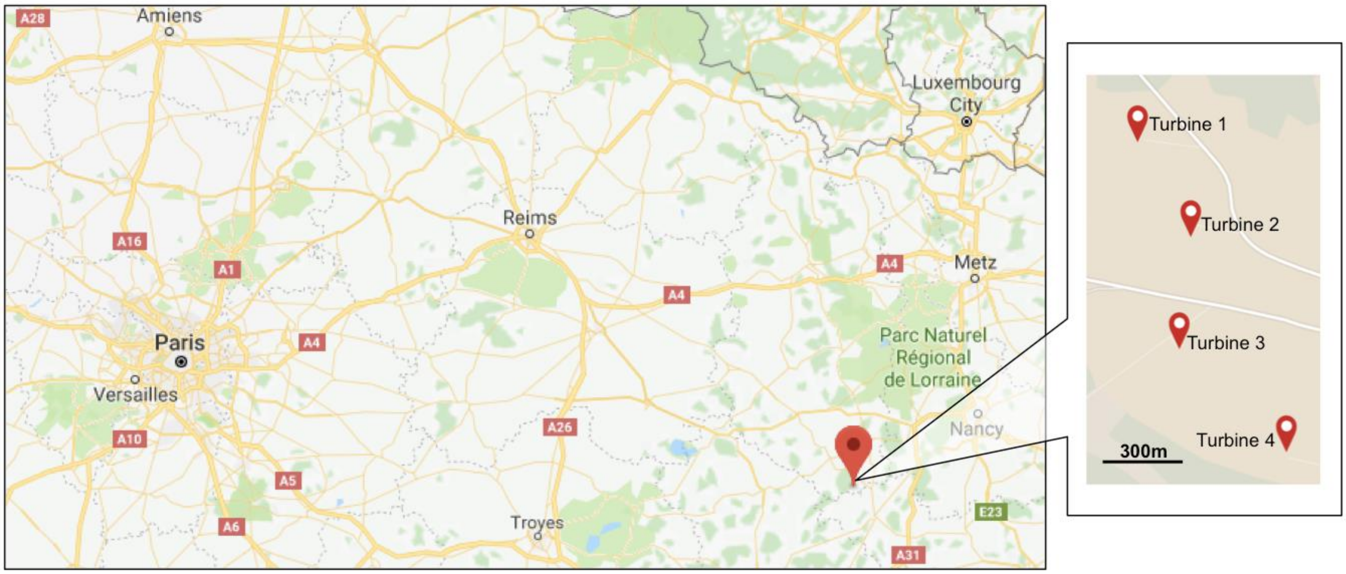


Figure 1. Location of La Haute Borne windfarm (left) and the names & relative locations of the four turbines within this wind farm (right)

Table 1. Distance and position information for the 4 turbines

Turbine No	Latitude	Longitude	Distance b/w 1 and x [km]	Distance b/w 2 and x [km]	Distance b/w 3 and x [km]	Distance b/w 4 and x [km]
1	48.4569	5.5847	0	0.4211	0.8169	1.3316
2	48.4536	5.5875	0.4211	0	0.4359	0.9119
3	48.4497	5.5869	0.8169	0.4359	0	0.5752
4	48.4461	5.5925	1.3316	0.9119	0.5752	0

The wind park includes four wind turbines and information regarding the wind speed, direction, temperature at each specific site are recorded continuously from 2013 to 2016 at a temporal resolution of 10 mins. All wind turbines deployed in this park are Senvion MM82 with 80 m hub height, 82 m rotor diameter and a rated power of 2050 MW. All turbines were at the same altitude of 411 m.

<u>Data preprocessing</u>

The raw data obtained from the open data website was cleaned up to remove infinite-valued and missing data points. As part of the pre-processing, the power curves for each turbine were plotted and outliers were removed (as shown in Figure 2). In addition, the data was sorted in chronological order to enable the machine learning model to extract temporal patterns.

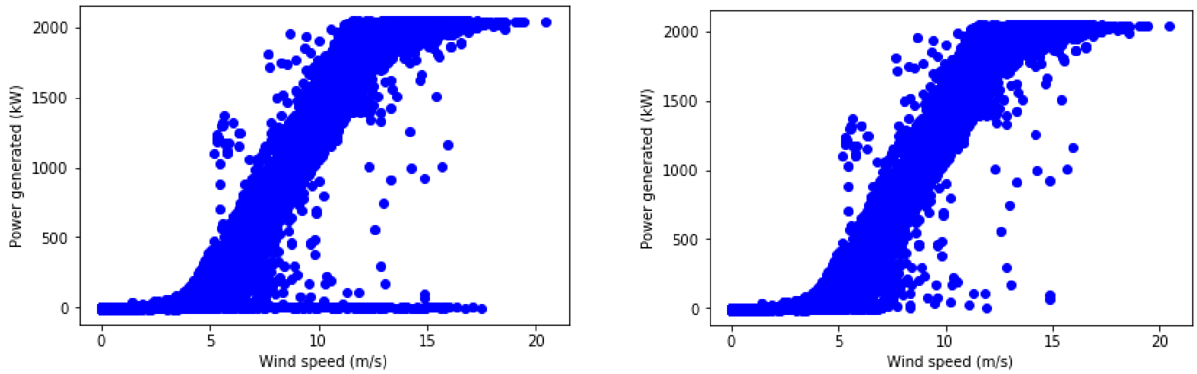


Figure 2. Power curves for turbine 1 before (left) and after (right) outlier removal.

3. Methodology

Time series model

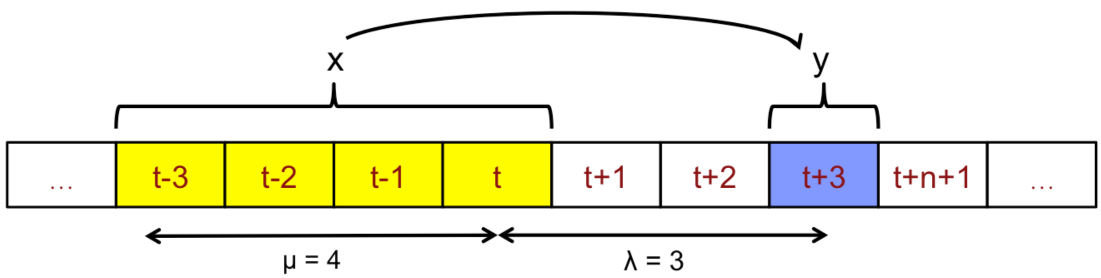


Figure 3. Section of a time series. The pattern x is mapped to label y. The time horizon of the prediction is λ., and the number of past measurements is μ.

The wind speed /power measurement $x = p(t)$ (pattern) is mapped to the wind speed /power production at a target time $y = p_T(t + \lambda)$ (label) with $\lambda \in N+$ being the forecast horizon (Figure 3). In our case we took $\lambda = 3$, meaning that the forecast is 30 min into the future. For our regression model we used one month of such pattern-label pairs (4320 data points) that are the basis of the training set $T = \{(x_1, y_1), \ldots, (x_N, y_N)\}$ and allow, via linear regression or SVM, to predict the label for a unknown pattern x. Pattern x was extended to 3 additional past measurements to feed more information to the model ($\mu = 4$). Note that λ was chosen to be 3 based on an initial evaluation of appropriate time horizon, that is, a λ that is too small does not provide sufficient prediction into the "future", while too large of a λ renders the prediction inaccurate because the further apart the two data points, the less correlation they have. An example of a bad fit because of too large λ value is shown in Figure 4.

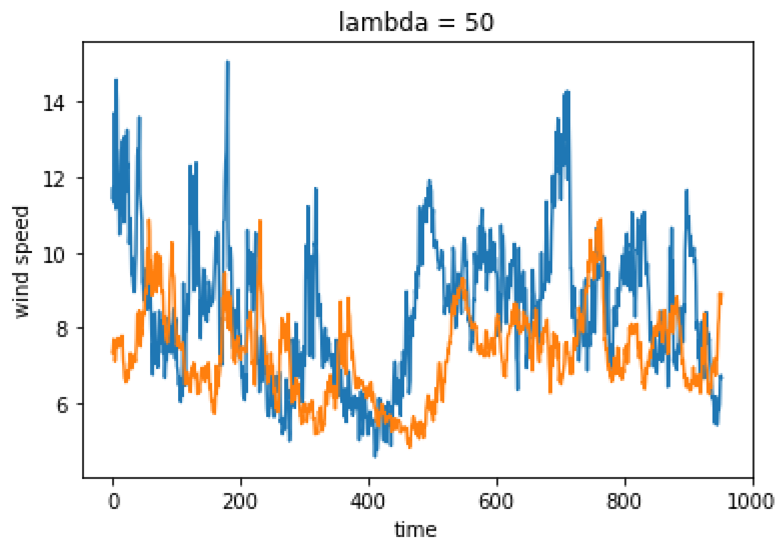


Figure 4. Illustration of bad fit when λ is too large. x axis (time) in [min], y axis (wind speed) in [m/s]

Predicting wind speed at other turbine site & power curve

In this project we explored using machine learning algorithms combined with power curve to estimate power generation for the nearby wind turbines in the wind farm. The method can be split into 3 different parts:

First, we train our model on a month of wind speed data at turbine 1, and then use the model to predict the wind speed at turbine 1 for a week following the month we trained on. Different machine learning algorithms including Linear Regression, Support Vector Machine (SVM) are used to perform this task and the performance of these algorithms are compared to determine a best practice. Linear regression algorithm produces models in the following form by minimizing the residual sum of squares:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n$$

Linear regression is simple and computationally inexpensive, but is incapable of capturing nonlinear patterns. Generally raw wind power data series may contain nonlinear and non stationary features. In anticipation to the nonlinear relationship between features and predictors we applied Radial Basis Function (RBF) kernel to map the original features into higher dimensional space and thus separate non-linear data.. This kernel trick has the following general formula:

$$K(X', X) = exp(-\frac{||X' - X||^2}{2\sigma^2})$$

Second, we use the predicted wind speed data at turbine 1 to estimate wind speed at turbine 2, turbine 3 and turbine 4. Linear Regression is selected to perform this task as the model is trained on more than 2 years of data points and advanced deep learning algorithm requires more computational power than we can offer.

Third, we generate the characteristic power curve for each wind turbine by plotting turbine power generation against wind speed at the turbine site. Then we feed predicted wind speed at each site to the respective power curve function to acquire the power generation data at that specific site.

This hybrid method has the advantage of learning large amount of data at a low computational cost. Linear regression methods are capable of learning efficiently on large datasets and adjust model parameters to make accurate wind speed predictions. Power curve demonstrate instantaneous energy production at different wind speed, so by providing wind speed to the power curve function we are able to achieve good estimate of power generation at that specific moment.

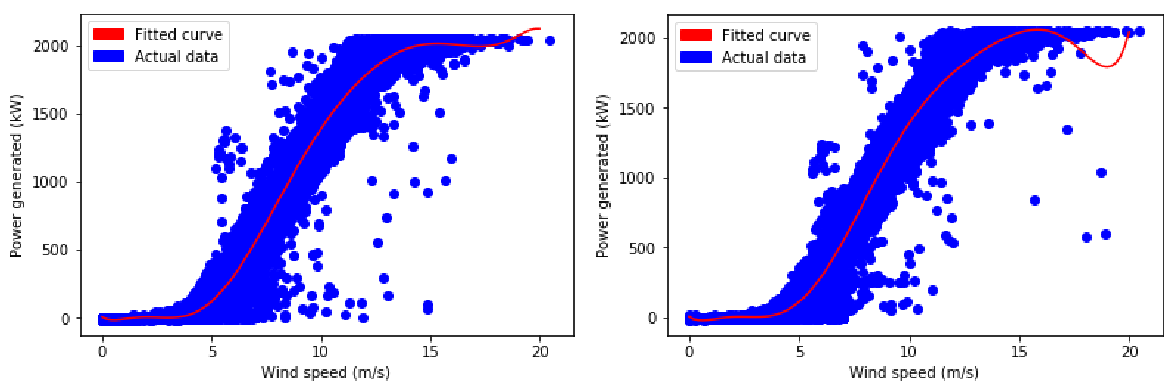
Predicting power generation directly from historical power generation data

Power curve is an useful ancillary tool to assist energy production forecast whenever turbine-specific power production data is readily available. However, in reality turbine-specific power generation data can be hard to collect, and we may often find we have to make predictions without the help from power curves. In this project our team also attempted to make forecast of energy production based on the historical energy production data. This method is compared with the first method we introduced to determine the best practice.

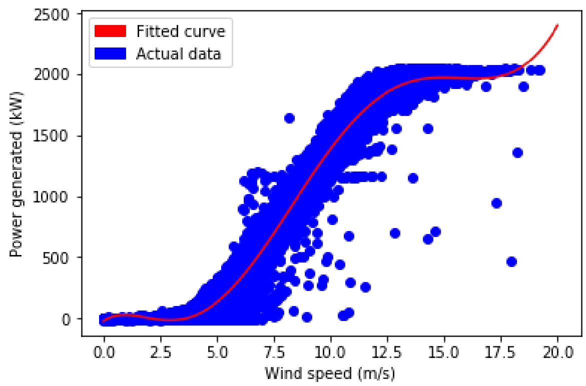
4. Results and Discussion

Power curve estimation

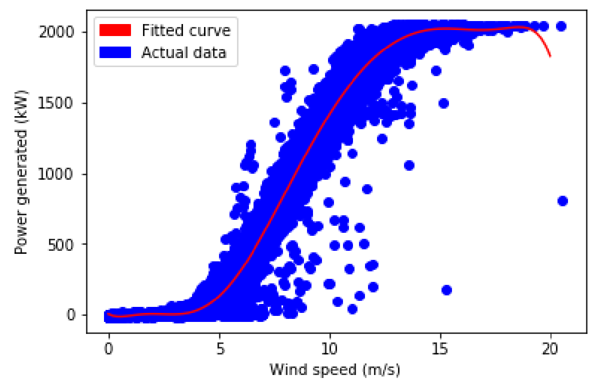
The power curve for the 4 different turbines was obtained by fitting polynomials of different degrees onto the available data. The fitted curves and their corresponding degrees are presented in Figure 5 below.



(a) (b)



(c)



(d)

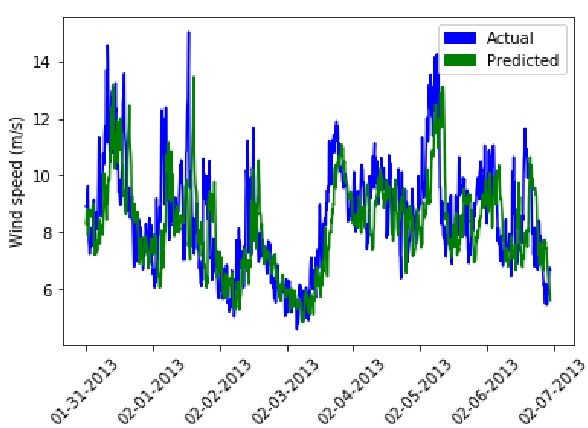
Figure 5. Power curves for Turbine 1 (a), 2 (b), 3(c) and 4 (d) shown above.

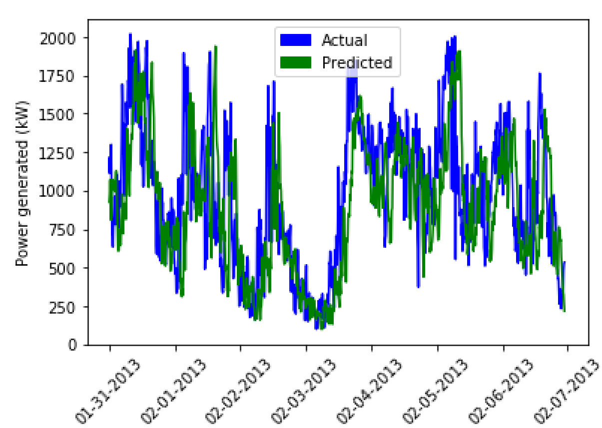
The power curves for turbine 1,2 and 4 used a polynomial of order 9 to obtain the fitted curves. For turbine 3, a polynomial of order 6 was used. The order was chosen by iteratively fitting polynomials of increasing order and choosing the one with the lowest error.

These power curves are used to estimate the power generated by a turbine when the wind speed at its location is provided. In this analysis, it is assumed that the power curve for each turbine is known beforehand.

<u>Forecasting wind speed at turbine 1</u>

The results of the wind speed forecast at turbine 1 for a period of one week and a comparison with the actual wind speed during that time is presented in Figure 6a. The power curve for turbine 1 discussed in the previous section is used with this forecast to estimate the power generated by the turbine (Figure 6b) in this time period. Since this forecast method only requires historical wind speed data, this provides a convenient and accurate way to estimate the power generated by turbine 1 in the future.





(a) (b)

Figure 6. Wind speed (a) and estimated power generated (b) at turbine 1 compared to actual values.

The wind speed forecast presented above are from a linear model that showed the best performance. The performance of other methods employed are shown in Table 2 below:

Table 2. Comparison of machine learning models on wind speed forecast (MSE: $(m/s)^2$).

Method	MSE	MAPE
Linear regression	2.53	13.74
SVR – Gaussian kernel	2.86	24.90
SVR – Polynomial kernel (deg 2)	3.76	27.47
SVR – Polynomial kernel (deg 3)	5.51	30.13
SVR – Polynomial kernel (deg 5)	~16500000	17656.53

The support vector regression (SVR) with a polynomial kernel of degree 5 performed unusually poorly and showed large non-physical predictions as shown in Figure 7 below. We expect that this is due to an overfitting with a large degree polynomial.

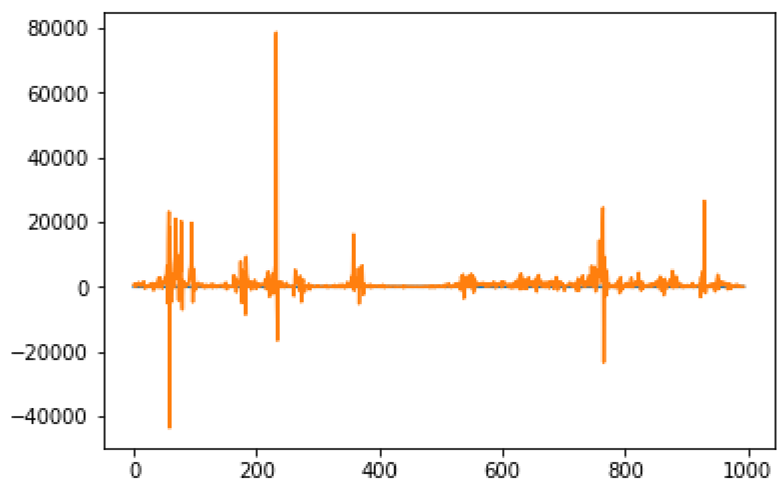
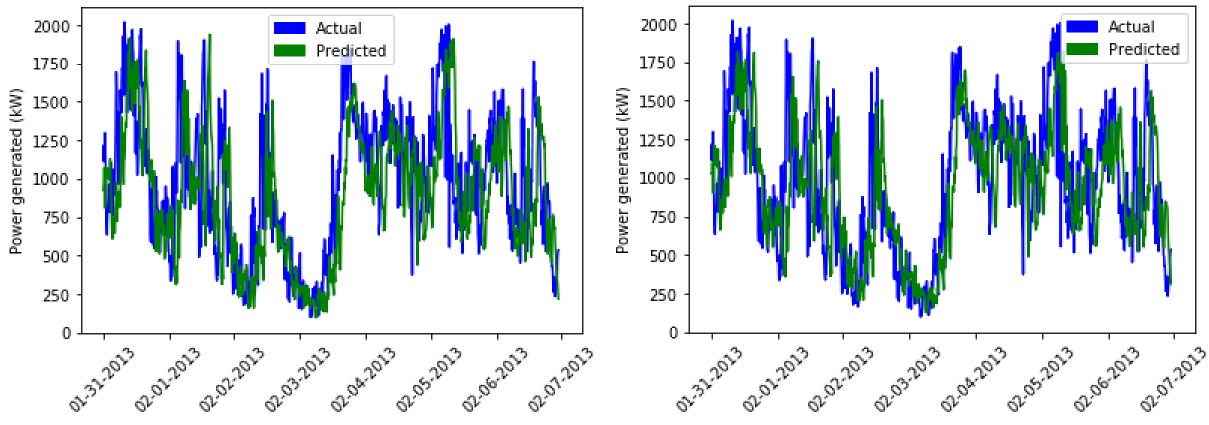


Figure 7. Plot of predicted wind speed for SVR with polynomial kernel of degree 5.

<u>Comparison with direct power forecast</u>

In the previous section, the power generated by turbine 1 was estimated by first forecasting the wind speed at its location and then using its corresponding power curve. In this section, the above result is compared to a *direct* forecast of the power generation at turbine 1 from historic data. As shown in Figure 8 below and quantitatively in Table 3, it is clear that the two methods of estimating power generation by turbine 1 based on historic information are close and accurate. This is useful because the actual wind speed information at a location can be easily obtained and using this model it is possible to predict the potential power generation by a turbine of a given power curve.



(a) (b)

Figure 8. Power generation estimate using power curve (a) and direct forecast (b).

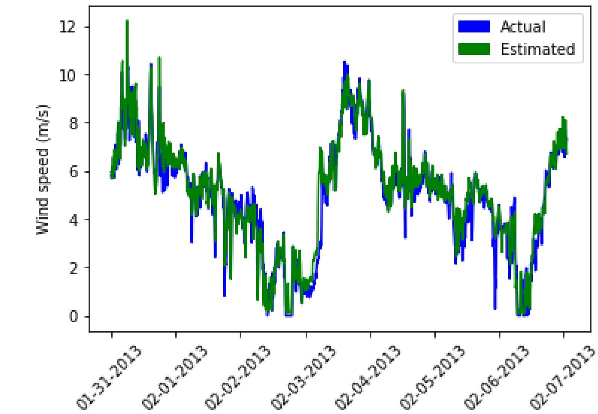
Table 3. Comparison of metrics for both methods of power estimation (MSE: kW^2).

Method	MSE	MAPE
Using power curve	154044.72	34.46
Direct forecast	140224.93	34.97

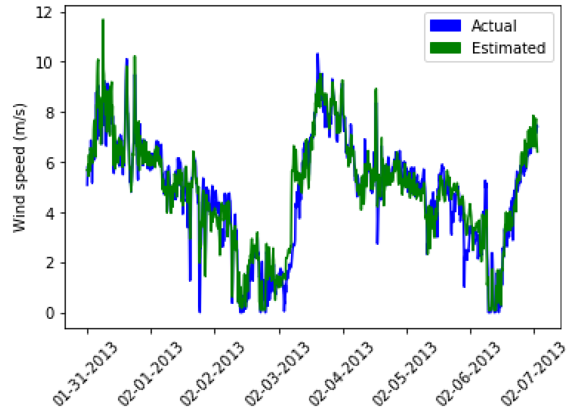
<u>Performance estimate of other turbines in the wind farm</u>

Using the wind speed information at only turbine 1, separate machine learning models were used to estimate the wind speed at the other turbines (2,3,4) as described in the methods section. The results of these wind speed estimates (for the same week shown in the above plots) is shown in Figure 9a.

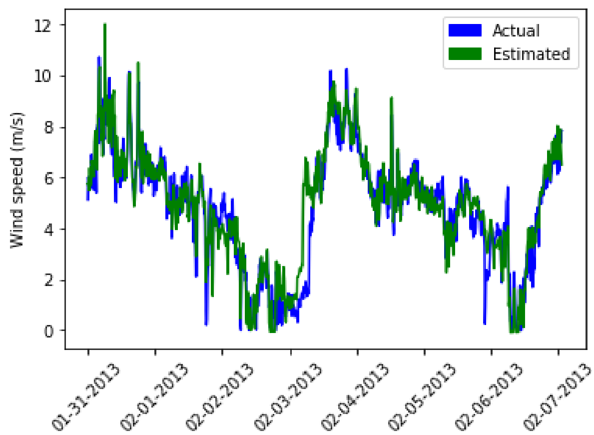
Similar to the approach followed for turbine 1, the power generation by these turbines are also obtained using their respective power curves. These results are presented in Figure 9b.



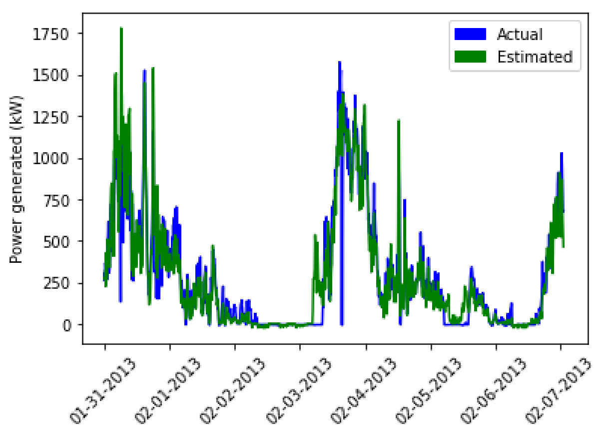
a. (i)



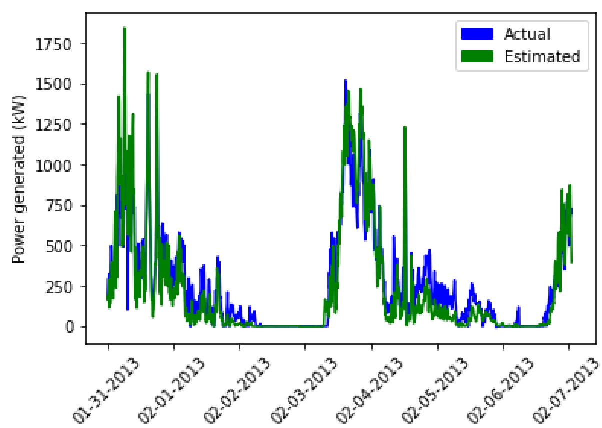
a. (ii)



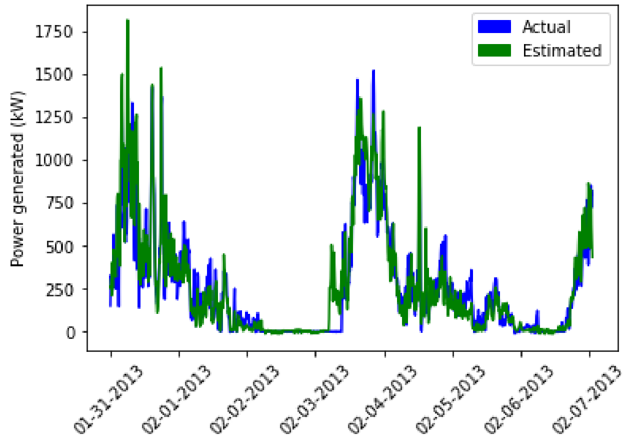
a. (iii)



b. (i)



b. (ii)



b. (iii)

Figure 9.
a. Plots of wind speed estimates at turbine 2 (i), 3 (ii) & 4 (iii).
b. Plots of power generation at turbine 2 (i), 3 (ii) & 4 (iii).

From these plots, it can be observed that the wind speed at the 4 turbine locations are highly correlated and by having information at any one location (such as turbine 1, as assumed in this analysis), one can accurately obtain wind speed information at the other locations.

Furthermore, at a given time frame, there are only slight variations among the wind speed at the different locations. This is because of their relative proximity and a limitation of the dataset used in this analysis.

5. Conclusions

In this analysis, a workflow for estimating the power generated by each turbine in a wind farm was proposed and demonstrated using real data from the La Haute Borne wind farm. Using historic wind speed information at one location, future forecasts of the wind speed at that and the other locations are made using machine learning models. Combining this information with the (known) power curves for each turbine, the performance of each of the turbines (and in turn, the wind farm) can be estimated accurately. A comparison of different machine learning models was presented and the results were also compared to a direct forecast of the turbine performance.

6. Future Work

The potential future work for this analysis can be presented in the following categories:

- Machine learning/predictive models:
    - Several other machine learning techniques such as Deep Learning and Hidden Markov Models can be applied to both forecast the future wind speed and

establish inter-turbine relations. Due to limitations in time and computational power, these were not explored in the current analysis

- Data:
    - The data used in this analysis contained only a limited number of wind turbines in the wind farm and hence, larger effects such as the effect of turbine location on generation and wake effects could not be studied.
    - Features from a quadrant-hole analysis could also not be included due to the lack of wind speed information in the vertical plane.
    - As discussed in the results section, heterogeneity in the wind speed at the turbine locations could possibly reveal more interesting patterns
    - Due to the scarcity of open source wind farm data, the authors believe that an open standard for such would be quite beneficial for sharing data and analysis

7. Contributions

All three authors came up with the idea of working on wind farm energy forecast.

All three authors also worked together in preprocessing the dataset.

Yuran worked on literature reviews and forecast methods, while Jason and Srinikaeth worked on designing and implementing the prediction algorithms. All members worked extensively on the project and presentation.

Overall, all the teammates were happy with each other’s contribution to the project and confident that they made a strong team.

8. References

Jung, J., & Broadwater, R. P. (2014). Current status and future advances for wind speed and power forecasting. Renewable and Sustainable Energy Reviews, 31, 762-777.

Liu, D., Niu, D., Wang, H., & Fan, L. (2014). Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renewable Energy, 62, 592-597.

Liu, Y., & Zhang, H. (2016, December). An Empirical Study on Machine Learning Models for Wind Power Predictions. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on (pp. 758-763). IEEE.

Treiber, N. A., Heinermann, J., & Kramer, O. (2016). Wind Power Prediction with Machine Learning. In Computational Sustainability (pp. 13-29). Springer, Cham.

Wiser, R., Lantz, E., Mai, T., Zayas, J., DeMeo, E., Eugeni, E., ... & Tusing, R. (2015). Wind vision: A new era for wind power in the United States. The Electricity Journal, 28(9), 120-132.