Leeghwaterstraat 44
2628 CA Delft
P.O. Box 6012
2600 JA Delft
The Netherlands

www.tno.nl

T +31 88 866 22 00
F +31 88 866 06 30

TNO report

TNO 2021 R12377 | Final report

Wind Field Estimation by Machine Learning regression.

Date	2 December 2021
Author(s)	M. Turrini
Copy no	
No. of copies	
Number of pages	17 (incl. appendices)
Number of appendices	
Sponsor	
Project name	Wind Field Estimation KIP WE1 2021
Project number	060.47428

Contents

Executive summary

The project involved the design of a methodology to both estimate real time and predict future wind speeds at a wind turbine rotor location leveraging only the knowledge of live on-field measurements (i.e. LiDAR, anemometer/wind-vane, other…). A method capable of providing live predictions as such would have the potential of being used for predictive control and possibly increase energy production and/or reduce structural loads on the turbines.

In preparation, the offline LAWINE dataset has been used. A reduced dataset of a few hours of measurement from both a two-beam LiDAR and a Meteorological mast was selected. SCADA data was also used to compute (offline) an estimated rotor speed as a pseudo-signal. This estimated rotor wind speed has been used as the objective signal to be mapped by a number of regression algorithms. The algorithms that have been tested in the scope of this project were two: (1) a Gradient Boost ensemble method and (2) a Feed Forward Neural Network.

Using these two approaches, it was possible to obtain both live and (time) predictive estimates of the wind speed at the turbine rotor location using only far-field LiDAR data (measurement range: 80-440m from the turbine). It was observed that the prediction accuracy was best for 5s ahead in time predictions and would then decrease until reaching a minimum for any attempt to predict 30s in advance, or further ahead in time. Furthermore, the second and third measurement ranges of the LiDAR (120-160m, or 1.5D-2D) were observed to have the highest correlation with the objective wind speed estimate. Therefore the range, or more generally the location, of the input measurements has an impact on the accuracy of the prediction and should be smartly chosen.

In brief, the project has achieved to both presents the potential of using Machine Learning (ML) algorithms as reliable tools for predictive estimation of the expected wind speeds at the turbine rotor plane, as well as set the scene to study the influence of the input measurements location to obtain the best wind speed estimates. The computation time of the wind speed predictions was not a limitation, and could be integrated into a live data processing stream. Furthermore, this computation time of the predictions is independent of how far ahead in time the prediction is done but is influenced by both the dataset preparation method and the type of offline training of the algorithms.

The tested methodology has potential for improvement along two pathways. On one hand, the algorithms used can be improved, or even changed. Currently, recursiveness is not a prior knowledge given to the algorithms. In the case of time-series prediction, deploying algorithms that do consider recursiveness, or put in other words, that have memory, is generally expected to have a positive impact on the accuracy of the results. On the other hand, the input signals' complexity and information quantity could be increased to leverage the power of the ML tools. This could be done by both/either moving from a two-beam LiDAR to a 4+ beams or scanning LiDARS, or by adding in the data stream the use of gaussian process regression to reconstruct the multi-dimensional wind field and use the latter as input instead of the raw measurements.

1 Introduction

The proprieties of the wind (speed, direction, eddies size and strength) that either a single wind turbine or a wind farm is subject to, have an influence on the system's functioning and therefore also their control strategy. As a direct consequence, new technologies are continuously developed with the goal of increasing the quality and the quantity of information obtained through on-line measurements. These measurements are centre-point for the continuous improvement of the wind power industry. Nonetheless, measuring the wind field directly in front of a wind turbine rotor, despite a few technological attempts [1], remains to this day a challenge. The presented project aims to propose a methodology to rely on both LiDAR measurements and numerical algorithms to obtain a reliable estimation and forecast of the rotor wind speed.

This documents presents an assessment of the impact of the measurement location onto the obtained wind field estimation. Hereafter, a two beam lidar is used as support, allowing for a one dimensional Horizontal Wind Field (HWS) reconstruction. It will be seen that given a specific wind condition; measurements ranges will be more of importance than others. To achieve this analysis, machine learning tools such as Ensemble Methods and Artificial Neural Networks are used.

As a by-product of the latter point, this project sets a baseline for the investigation of the available potential in methods combining machine learning with LiDAR measurements. The tests are aimed at developing tools capable of either estimating the wind at the rotor location live using remote measurements or predict future states. The outcome of positive results are, among others, predictive control techniques to improve the Annual Energy Production (AEP) of turbines and minimize the loads on the system.

The hereby presented Wind Field Estimation project also contributed in large part to the development of the in-house tool GPyDAR. The latter tool offers a method for Radial Wind Speed (RWS) interpolation (and limited extrapolation) of data coming from LiDAR measurements for three dimensional field reconstruction.

2 Site presentation and measurement campaign

The project uses data obtained by the "LiDAR Application for WINd farm Efficiency", or LAWINE, project, a measurement campaign from 2016 led by ECN together wind Delft University of Technology, Avent LiDAR Technology and XEMC Darwin in the framework of TKI Wind op Zee. The LiDAR data was obtained by NORCOWE.

Figure 1 : Map of the province North-Holland, The Netherlands and a detailed map of the test site [2].

The ECN Wind Turbine test site Wieringermeer (EWTW) is a flat terrain consisting mainly of agricultural area, with few farmhouses and rows of trees. It's location and the detail of the layout can be seen in Figure 1 and Figure 2 respectively.

Figure 2: Layout of the test site indicating the position of the Nordex turbines, the prototypes turbines, the meteorological masts, the measurement office and the LAWINE LiDAR locations [2].

The LiDAR of interest in this project is the Avent Wind Iris LiDAR (WI) placed on the top of Turbine Nordex 6 (N6), and the nearby located Meteorological Mast number 3 (MM3). The relative placement of the Nordex turbines, and the MM3 is visible in Figure 3.

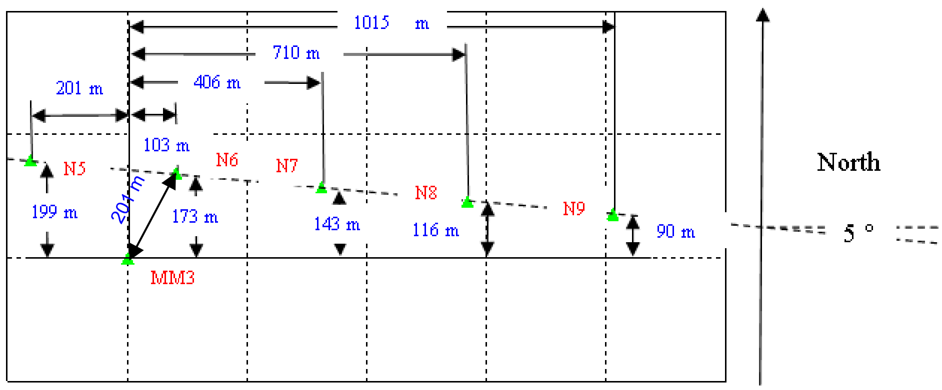

Figure 3: Locations of the five Nordex N80 turbines and Meteorological Mast 3 [2].

The WI LiDAR is a two beam LiDAR with a beam angle of 15 [deg] from to the centre line (see Figure 4). When the wind is coming from SSW (~200 [deg]), the nacelle Line of Sight (LOS), and therefore the WI, are aligned with the MM3. The MM3 is equipped with a cup anemometer at hub height (80m) and a wind vane, used as reference for Wind Speed (WS) and Wind Direction (WD) measurements. The latter will hereafter be the reference for wind speed measurements at the MM3 location.

The WI LiDAR measurements are done on the beam LOS on ranges spanning from 40m to 440m, with a spacing of 40m. The frequency of the signal is of 4Hz.

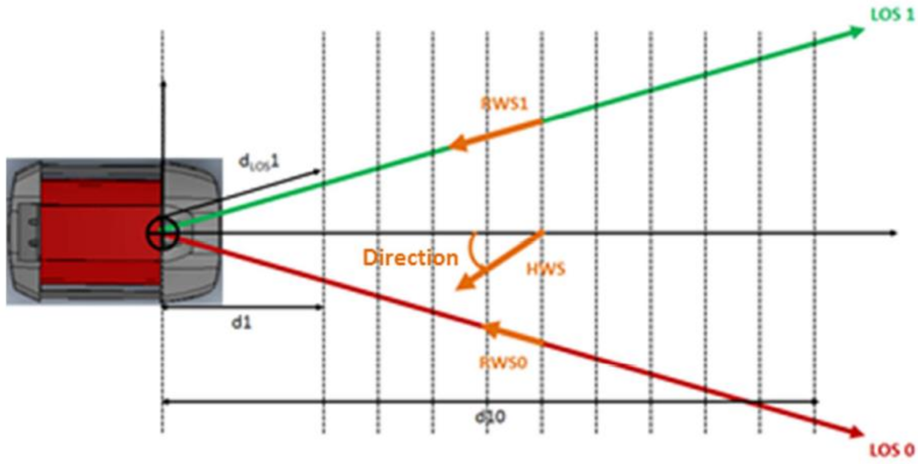

Figure 4 : Schematics of the Wind IRIS beams geometry (adapted from [3]). The measurements are planar (2D).

The MM3 For more information on the site and the measurement campaign, refer to the "*LAWINE Instrumentation report*" internal document [2].

3 Data overview

There are three sources of wind speed mentioned in this report. The first two are the wind speed measurements from the WI LiDAR and the MM3, and the third one is a Wind Speed Estimate (WSE) at the turbine rotor. The latter is a pseudo signal computed using SCADA data from the turbine. A Kalman Filter estimator model generates the expected wind that the turbine rotor is experiencing during operation using high frequency measurements. The so called WSE is, in the presented project, the objective function for the machine learning regression algorithms.

While there is no higher limit in the amount of data that can and should have been used for a project such as the one presented, 6 hours of measurements have been chosen for analysis (2014-01-07 01:00:00-07:00:00). This time window has been selected for its consistent recorded wind speeds and wind directions appropriate for an alignment of the WI LOS and the MM3.

Figure 5 and Figure 6 represent respectively 420 and 30 minutes time series of the HWS as obtained by the three mentioned sources. For or visualisation purposes, the data is filtered using a 60s moving average. Figure 5 presents a visualisation of the HWS as computed by the WI at the MM3 location.

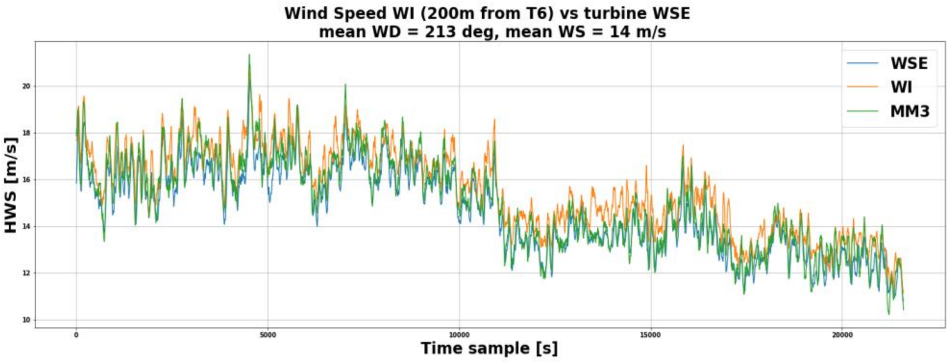

Figure 5 : In green, orange, and blue, 6h time series of the HWS as measured by the MM3 and the WI 200m from the turbine, and the WSE as computed by the estimator at the turbine rotor location.

Figure 6 depicts the HWS WI measurements plotted against the WSE during a shorter time window (30 minutes). This representation allows to visualise the observe lag (11.5s) that appears between the two signals.

This time lag is due to the difference in space of 200m. The lag was computed by cross correlation analysis. In the present case, the mean wind speed was 16 [m/s] (200m~=11.5 [s] *16 [m/s]). The lag in time and space will be hereafter discussed as directly correlated, leveraging the concept of information traveling over space and time.

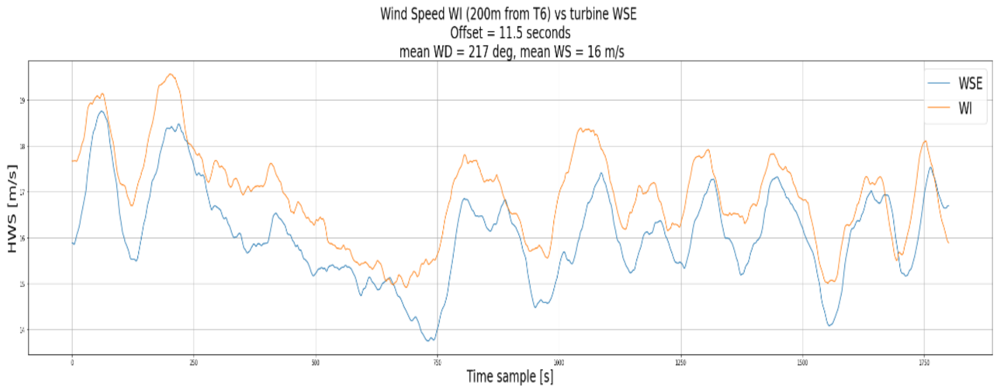

Figure 6 : 30 minutes time series of the HWS reconstructed by the WI LiDAR at 200m from the N80 wind turbine and the WSE at the rotor.

To quantify the linear correlation (similarity in temporal evolution) between two signals X, Y, the Pearson coefficient, or R coefficient, is used hereafter and in sections to come of this document. A high Pe coefficient is synonym of signals in phase, with correlated dynamics. The definition is given by the following equation:

$$Pe = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

Where $Cov(.,.)$ is the covariance function and σ_i is the standard deviation of signal i=[X, Y].

Figure 7 offers an overview of the correlation between input signals (MM3, WI) and objective signal (WSE). The signals are compared between filtered and unfiltered (noisy), using a moving average of 60s. As expected, the correlation decreases as the range increases. This is due to (1) the time delay induced by the shift in space (2) the non-frozen nature of the wind travelling downstream but instead subject to changes as it travels. It is to be noted that the correlation between the HSW measured by the WI at 200m and the WSE is comparable to the correlation between the MM3, also located at 200m, and the WSE. This was expected as the WI accuracy has been in the past validated against the MM3 measurements [4]. A noticeable increase in the correlation is also visible after smoothening of the signals suing the presented moving average approach.

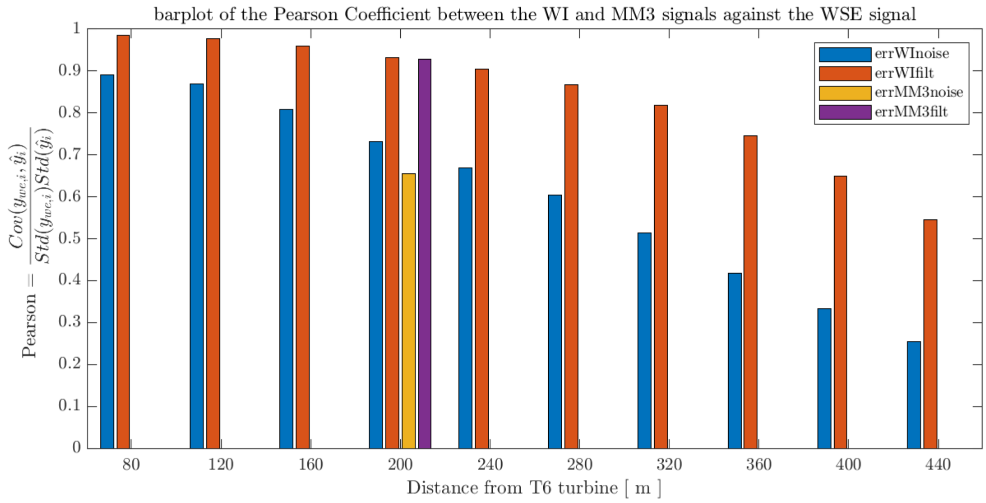

Figure 7 : Correlation analysis between the measurement signals coming from the MM3 and the WI compared against the WSE. For each range, the filtered and unfiltered signals are compared against the also filtered and unfiltered WSE.

4 Wind estimation methodology

Linking input signals to output signals using empirical model and machine learning model is nowadays common practice. The problem can be approached in numerous ways, from simple regression to recursive algorithms taking into account past information for the estimation of future states. For this preliminary study, the base approach of a non-linear regression is set up and investigated.

As the objective is to see the influence of the input measurements spatial origin onto the output estimate, the regression problem is set as follows:

- The regression input are 10 values, one for each range of the WI LiDAR.
- The regression output is 1 value, the WFE at the turbine.

The problem is to be seen as an observer scanning time step by time step 10 time series simultaneously, moving forward in time and outputting the time step of the objective time series at the queried time instant. The queried time instant can be before (negative lag), after (positive lag) or at the same instant (null lag). It is clear to see that as the problem is set up, the model used is not given the opportunity to learn time correlations in recursive events.

As a negative time lag is not of interest here, the analysis of the estimation of the wind speed will be done with increasing lag, representing how far in the future it is possible to predict.

The regression models used in this project are two: an Ensemble Technique of Gradient Boosting (GB), and a Deep Feed Forward Neural Network (DFFNN).

4.1 Gradient Boosting

Gradient boosting is a type of ensemble methods, or ensemble of regression trees (models/learners…, see Figure 8). It relies on the intuition that the best possible next tree, when combined with previous ones, minimizes the overall prediction error. The algorithm relies on a number of learners and trains them in sequence so that the subsequent trees are built with knowledge of the precedent trees. At the end of the iterative process, the final model is a weighted sum of the n existing trees.

The subset of parameters used is:

- n_estimators : the number of boosting stages that will be performed.
  - 300
- max_depth : limits the number of nodes in the tree. The maximum depth is the depth of the decision tree estimator in the gradient boosting regressor.
  - 9
- min_samples_split : the minimum number of samples required to split an internal node.
  - 500
- learning_rate : how much the contribution of each tree will shrink.
  - 0.1
- loss : loss function to optimize.
  - Mean Squared Error

Where the values indicated have been chosen combining empirical best guesses and a grid search for optimizing the accuracy of the results.

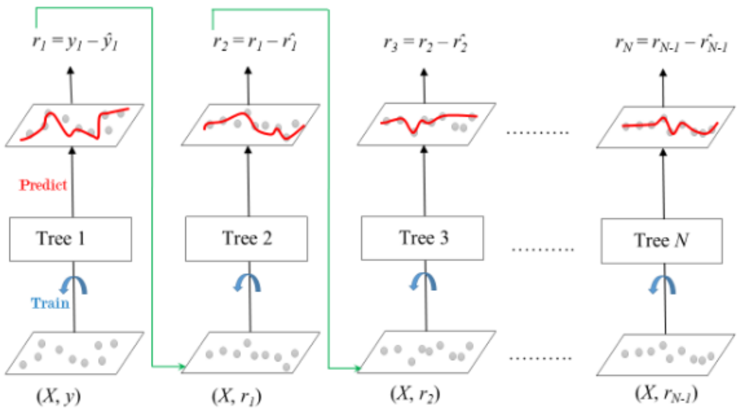

Figure 8 : Construction of a Gradient Boost ensemble of N trees.

4.2 Deep Feed Forward Neural Network

Neural Networks are a type of machine learning that relies of a structure of artificial neurons called processing units or nodes. These processing units linked by activation functions. The input units receive structures of information based on an internal weighting system. Recursively, the neural network attempts to link input structures to output structures using a loss function as an learning indicator. Several type of nodes and activation functions exist. In the presented case, a feedforward structure is used. This is the most accessible type of NN and can be considered the baseline of most applications involving NN.

The structure used is visible in Figure 9. The structure of two hidden layer of 64 nodes is chosen as a compromise between a perceptron (1 layer, or ensemble of nodes) and a more deep structure which would risk to underfit or overfitting the training data, respectively. The structure was approximated by a best guess and optimised iteratively. Layers are interconnected using relu functions.

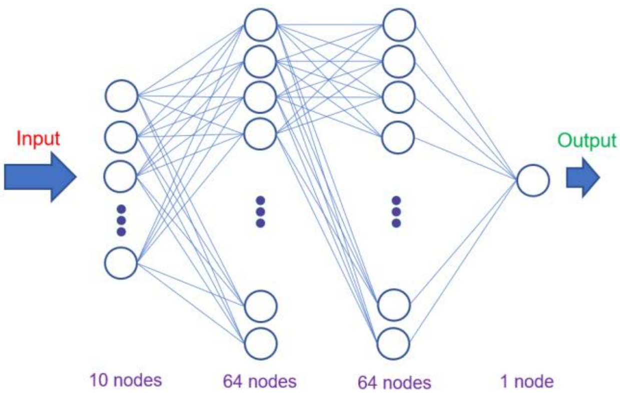

Figure 9 : Diagram of the DFFNN used. The arrow indicate the data flow direction

5 Results

Both the GB and the DFFNN have been trained on about 290 minutes of HWS measurements signals and evaluated on 72 minutes of input signals. The frequency of the signals being of 4 Hz, this corresponds to 691200 data points for training and 172800 data points for validation.

The index used for the evaluation of the results are the Person factor (Pe), as presented in section 3, and the MSE, computed as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Where n is the number of samples, Y_i is the estimated value and $\hat{Y}_i$ is the true (or objective) value. Figure 10 summarizes the results of the analysis. The Pe and MSE index evolution as the time lag is increased can be observed for the estimations done using (1) the GB model, (2) the NN model and (3) the mean value of the inputs. The latter is used as baseline for the other models, indicating the un-weighted sum of the input values.

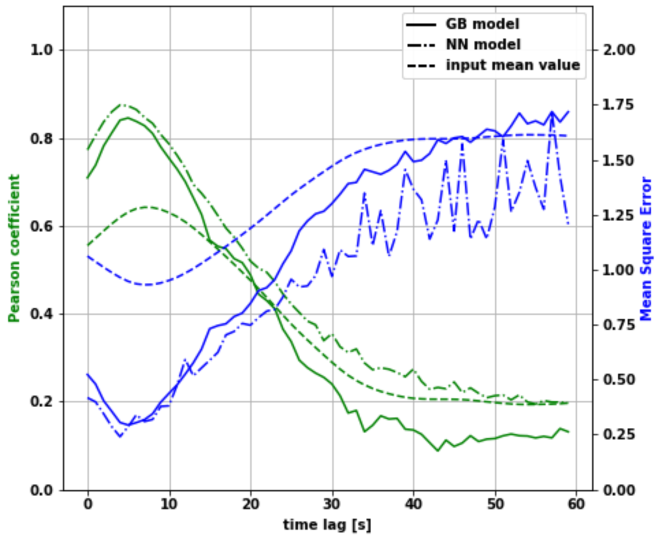

Figure 10: Summary of the results for the NN model (left) and the GB model (left). In blue, the MSE error between the estimated signal and the objective. In green, the correlation factor. In dashed, the results using the mean value of all the inputs as reference.

It is clear that the both the Neural Network and the Gradient Boost models are already a clear improvement from the mean of the input signals. This trend is most strongly visible for a time lag between 0 and 10/20s. Both the Pe and the MSE show record a peak for a time lag lg=5s, with best values of 0.88 and 0.17 respectively. As the time lag is extended beyond the 30s mark, the added value brought by models slowly disappears. It is interesting to notice that the NN seems to produce a better estimate all the way until a lag of 60s. The information of the state of the wind is not available to the model, since 60s of time lag corresponds, on average, to ranges above 700m, while the furthest measurement point is much closer at 440m. Therefore, despite the non-recursive set up of the model, a pattern was most likely learned during training.

Figure 11 illustrates the correlation between inputs and the target function. The structure allows to visualise how certain inputs are expected to have a higher importance in the learning of the models. It can be seen that the mean highest correlation between inputs and the target depends on the time lag. For a null time lag, the first, or closest measurement to the turbine has the higher correlation. As the lag increases, the peak of correlation shifts to measurement ranges further away from the turbine. Nevertheless, we see that the correlation peek, even for time lags of 30s or higher, remains centred around the measurement ranges WI2 and WI3, the third and fourth closest point from the turbine. It can be noted that the WI2 and WI3 ranges correspond to 120-160m, or 1.5D/2D, on the region bordering the end of the induction zone of the turbine.

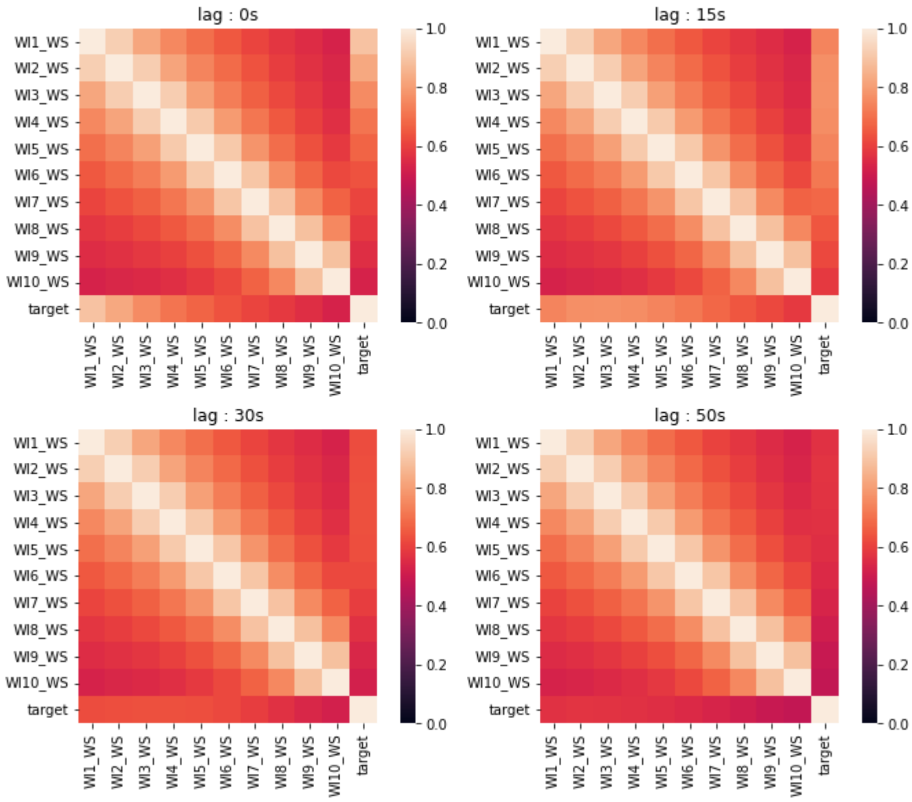

Figure 11 : Correlation matrixes between. The signals WIX_WS indicate the WI measurements at location X = [1;10], and the target signal, the WSE at the turbine rotor.

6 Discussion

From the results presented in section 5, it was shown that:

1- The range (or spatial location) choice of the measurements used as input for live and predictive wind speed estimation at the turbine rotor has an impact on the accuracy of said estimation. Indeed, some ranges have shown to have a higher correlation with the objective function. When the time lag matched on average the distance of these ranges from the turbine, the estimation accuracy was at a maximum (lowest MSE = 0.17, highest Pe = 0.87)

2- Regression tools such as GB or NN have the potential to create smart connection between the measurements as inputs, and the output of interest: the wind speed estimate. Both the model have shown that without physics driving the learning, it is possible to leverage their potential to attempt time series regression for live, but also predictive estimations.

From this observation, it is of interest to ponder on roads for improvement.

On one hand, in this project, the spatial location was only made vary in one dimension, i.e. closer or further away from the wind turbine, in a line of sight parallel to the turbine rotor axis. The use of either 5 beams nacelle mounted lidars or scanning lidars placed nearby the turbine would allow to expand the study to three dimensions. For this, a combination of the LiDAR measurements with the in-house developed GPyDAR tool should be considered. As a matter of fact, the use of the presented alternative LiDAR technologies in combination with the GPyDAR tool would allow to precisely grid the volume of wind flow. Virtual measurements could be precisely located in space, and used as input to models similar to the ones used in this project.

On the other hand, the machine learning tools developed in this project, while already having yield promising results, are basic. The room for improvement is large, from more complex regression models, to advanced techniques that would be capable of capturing the time variance of the input data. A few tools can be considered as suggestion:

1- Recurrent Neural Networks
2- Sliding window regression methods
3- LSTM Networks

The second point is hereby left for future development. As for the first point, it can be noted that this has been attempted in this project. The GPyDAR tool has been set up to use the data gathered by two beam forward looking Wind Iris LiDAR (see Figure 12). Nevertheless, the measurements obtained by the WI were not spatially nor temporally sparse, making the use Gaussian Process interpolation a computationally expensive tool with scares added value. Therefore, the use of the estimation of the RWS from the GPyDAR tool has not been considered valuable and left on the side during the analysis of the results.

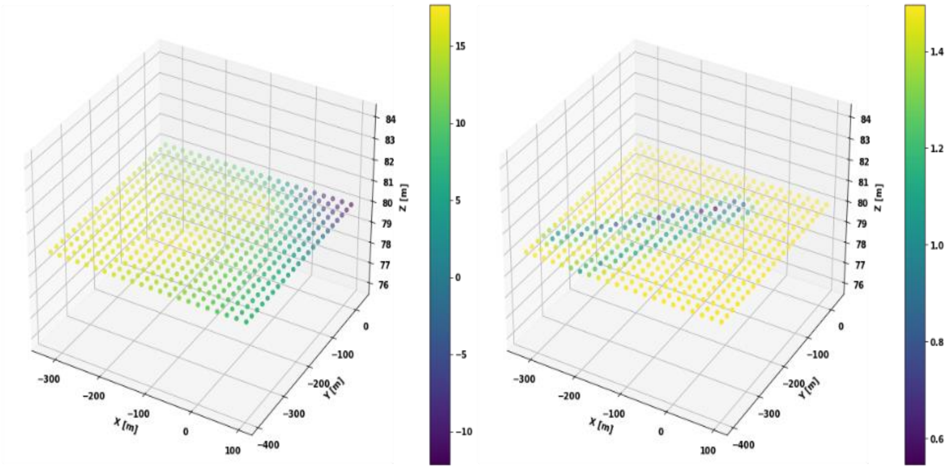

Figure 12 : 2D wind field reconstruction on a grid of 400x400m, placed at hub height. On the left, the RWS field, on the right, the uncertainty field.

7 Conclusion

This project considered the design of a method aimed at producing online and predictive estimation of the wind speed experienced by wind turbines using LiDAR measurements during turbine operation. The set up considered a forward looking lidar, two machine learning regression tools, and an estimator model based on turbine SCADA data.

The study has shown the possibility of achieving accurate estimations of effective wind speed live and as far as 10 seconds ahead in time before experiencing a drop of both the Pearson correlation factor (<0.8) and the MSE (>0.2) computed between the estimated and the true signals. The estimation beyond 10 seconds using the regression models sustained a visible improvement from the best guess baseline models until reaching a lower limit at time lags of 30s and more. The Neural Network model, while not entirely out performing the Gradient Boost model, has shown to be capable of drawing correlations between input and output signals beyond the 30s lag mark, indicating a potential to rely on patterns in the close field rather than information on the far field for predictive estimations. As a matter of fact, the maximum LiDAR range measurements was 440m, corresponding in theory to a capability to father information of winds incoming at up to 30s ahead on average for a mean speed of 13-14m/s.

The spatial influence of the incoming measurements was also observed, noticing that the peak in both Pearson correlation coefficient (0.9) and the minimum in the MSE was found for a 5s lag. Given the mean wind speed of the period used for validation (13m/s), this time lag value corresponded to the 80m range or the first range of measurement of the WI LiDAR. It was seen that the ranges with highest correlation between inputs and outputs, even when the time lag moved beyond 5s, remained centred around the value of 1.5D-2D.

Given the promising results of this project, a few paths for further improvement have been proposed. On one side, the spatial influence of the input measurements points could be directed towards the use of multi beam and scanning LiDAR measurements to further understand the correlation between space and wind speed estimation accuracy, expending from the 1D analysis proposed by this document. On the other side, a number of machine learning tool have been presented as possible improvements to the current set up. The proposed ones along other available, have potential if implemented in online measurement streams with the goal of improving turbine performance.

8 Bibliography

[1] H. Hohlen, “Holistic performance monitoring of wind farms - the iSpin Guardian,” 2016.

[2] G. Bergman, J. Wagenaar and K. Boorsma, “LAWINE instrumentation,” 2016.

[3] U. SAS, *Wind Iris Software 1.1 – User Guide.*

[4] J. Wagenaar, S. Davous, A. Medawar, G. Courbard-Millet and K. Boorsma, “Turbine performance validation; the application of nacelle LiDAR,” 2014.