Universidade do Minho
Escola de Engenharia

Maria Sampaio Magalhães

Park wind field reconstruction using real-time turbine data

Dezembro de 2021

Universidade do Minho
Escola de Engenharia

Maria Sampaio Magalhães

Reconstrução do campo de ventos de um parque eólico utilizando dados de turbinas em tempo real

Dissertação de Mestrado
Mestrado Integrado em Engenharia Mecânica

Trabalho efetuado sob a orientação do
Professor Doutor Luís António de Sousa Barreiros Martins e Doutor Gonçalo Artur Duarte Pereira

Dezembro de 2021

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

ACKNOWLEDGMENTS

First of all, I would like to thank my mentors, Doctor Gonçalo Artur Duarte Pereira of Vestas, and Professor Luís António de Sousa Barreiros Martins, for their guidance, availability, and support throughout this process. Without their help, the development of this dissertation would have been considerably harder.

I want to thank the help and guidance of Doctor Gonçalo Lucas Marcos and Doctor Jens van Schelve from Vestas, for always showing me their sympathy and help, whenever I needed.

I want to thank all of my friends that helped me stay motivated and focused throughout several moments in the making of this dissertation.

I would also like to thank my boyfriend, José Miguel Maia Ribeiro, for the unstoppable support and motivation he was able to provide, even in the most difficult moments. And a special thank you to my family, for always being there, supporting me through the good and the bad days. Thank you for all your affection and love, that helped me pursue this journey with the strength I needed.

Integrity Declaration

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledge the Code of Ethical Conduct of the University of Minho.

ABSTRACT

Wind field reconstruction is a method established to better understand park level aerodynamics, and to be able to control specific turbines in a wind park. It can be used in case anything happens with the turbine (for example, a sensor failing) that will cause the lack of information about the conditions of the wind reaching that turbine. This is critical since every turbine is monitored to continually guarantee that they are facing the wind in order to produce an optimal amount of energy. When all the available wind sensors fail, the production needs to be stopped. Because of this, flow reconstruction and forecasting values of wind characteristics such as wind speed and wind direction on a turbine, are methods that have been studied and validated for a long time. This dissertation aims to study how flow reconstruction in a wind park can aid in understanding turbines' conditions and interaction in a wind park and the wind characteristics on a specific location. Specifically, this dissertation tests seven statistical models developed in Matlab to understand the ability of the models to accurately represent the values of wind speed and wind direction of a turbine, given the dataset of the other remaining turbines in the park.

To test the hypothesis that adding more specific wind-related parameters to the models would improve their results, the first model tested was a more straightforward approach, where it was tested a regression analysis. Afterward, parameters as the distance between turbines, possible wake with some of the turbines in the park, and statistical analysis results were added or experimented in the models to understand if the results improved. One of the models constructed followed a different approach, and it was based on wind particles, and considering a viscosity condition and the last model developed was a machine learning model.

The results suggest that investing in statistical parameters, combined with wind park information will lead to optimal results. Also, machine learning models, possibly developed even further, are algorithms that, combined with a low computational cost, will lead to a very accurate representation of the real wind speed and wind direction values that a turbine is measuring. Another possible conclusion was that every turbine has different behaviour and will present a different error for the same model since they are scattered in a wind park. In some cases, other turbines can surround them and cause disturbances in the wind or cause difficulties for the models to estimate some variations on these turbines, while others are isolated.

KEYWORDS: Flow reconstruction, wind speed, wind direction, wind park.

Resumo

A reconstrução de parâmetros do vento é um método estabelecido de forma a auxiliar um maior conhecimento da aerodinâmica de um parque de turbinas eólicas e ser possível ter controlo de turbinas específicas num parque, em caso de uma falha de sensores, por exemplo, que poderá levar a uma falta de informação sobre as condições de vento às quais a turbina está exposta e esta é uma informação crítica sendo que o controlo da turbina é feito de forma a garantir que a turbina esteja sempre alinhada com a direção predominante do vento, para produzir uma quantidade de energia otimizada. Quando os sensores de vento falham, a produção da turbina precisa de ser travada. Dessa forma, a reconstrução do vento e estimativa dos valores de características como velocidade e direção, são métodos que têm sido estudados e testados ao longo do tempo. Esta dissertação tem como objetivo estudar e investigar de que forma a reconstrução do vento num parque de turbinas eólicas pode auxiliar a perceber as condições a que as turbinas estão expostas, interações entre estas num parque e as características de vento em localizações específicas. Especificamente, esta dissertação testa sete modelos estatísticos diferentes desenvolvidos no Matlab de forma a perceber quais dos modelos conseguiria representar as característias do vento com a menor margem de erro, tendo como dados de entrada dados reais das restantes turbinas do parque.

Para testar a hipótese de que acrescentar um grau de complexidade aos modelos, considerando parâmetros de vento mais específicos, iria melhorar os seus resultados, o primeiro modelo testado foi uma abordagem mais simples, tendo sido testada uma análise de regressão. Posteriormente, foram adicionados parâmetros como a distância entre turbinas, possível efeito de wake entre estas e análises estatísticas foram acrescentados e testados em novos modelos, para verificar se os resultados eram otimizados. Um dos modelos construídos foi baseado em partículas de vento e considerando viscosidade entre estas, e o último modelo desenvolvido foi relativo a machine learning.

Os resultados sugerem que combinar parâmetros estatísticos com informação específica do parque pode levar aos melhores resultados. Além disso, modelos de machine learning, possivelmente ainda mais complexos, são algoritmos que, com um custo computacional baixo, garantem uma representação de velocidade do vento e direção muito realista. Outra conclusão possível foi que cada turbina tem um comportamento diferente para o mesmo modelo. Em alguns casos, outras turbinas perto podem causar distúrbios e mudanças no vento que chega a turbina que está a ser estimada e isto poderá causar dificuldades na forma como os modelos preveem os valores de velocidade e direção.

Palavras-chave: Reconstrução do vento, velocidade do vento, direção e parque eólico.

TABLE OF CONTENTS

FIGURE INDEX

LIST OF TABLES

List of Abbreviations

Nomenclature	Meaning
VAWT	Vertical–axis wind turbine
HAWT	Horizontal–axis wind turbine
LiDAR	Light detection and ranging
RD	Rotor diameter
RMSE	Root mean square error
SD	Standard deviation
FT	Ultrasonic anemometer
RM	Regression model
DWM	Distance weighted model
CWM	Correlation weighted model
WIM	Wake influenced model
MIWM	Mutual information weighted model
MLM	Machine learning model
PM	Particle model

LIST OF SYMBOLS

Symbol	Meaning	Unit
A	Area	(m^2)
a/b	Width of the bell-shaped function	
C_p	Power coefficient	
C	Center of the bell-shaped function	
d	Distance between the turbines	(km)
F	Tangential force	(N)
h	Height of the bell-shaped function	
I	Mutual information coefficient	
L	Longitude	(°)
p	Pressure of the fluid	(Pa)
P_{eff}	Effective power output	(W)
$P_{máx}$	Maximum available power	(W)
$p(x, y)$	Joint probability function	
$p(x)$	Marginal density function for variable x	
$p(y)$	Marginal density function for variable y	
R	Radius of the Earth	(km)
v/u	Wind speed	(m/s)
v_f	Velocity	(m/s)
$y_i/\hat{y}_i$	Estimated and real measured values of the turbine	
φ	Absolute wind direction	(°)
β	Bearing or angle between two points	(°)
θ	Latitude	(°)
η	Dynamic viscosity	(kg/m.s)
γ	Error	
δ	Coefficient of viscosity	(P)
ν	Kinematic viscosity	(m^2/s)
η	Dynamic viscosity	(kg/m.s)
ρ	Fluid density	(kg/m^3)

1. Introduction

1.1. Background

Wind energy is one of the primer drivers of the energy transition towards renewable energy. In Portugal, renewable energy production in 2016 reached 33.3 TWh, and wind energy contributed 12.5 TWh. Because wind energy is renewable and environmentally friendly, systems that convert wind energy into electricity have developed rapidly. However, the implementation of wind energy faces some barriers in reality. Each turbine is equipped with meteorological instruments that monitor the wind and are later used to control the turbines. There are various events that can lead to a turbine shut down, such as meteorological sensors as a wind vane fail, and it is not possible to obtain the wind speed values at a certain point in time. This dissertation is crucial since it will investigate possible resolutions and models that can allow turbines to operate normally, even in cases where the information from sensors is not available. That will be exceptionally useful to optimize production levels in wind parks worldwide, improve wind turbines performance, reduce operation and maintenance costs, and help reduce even further the impact that energy production has on climate change. These measurements' high temporal resolution characteristics and the typically spatial distribution of the wind over a large area such as a wind park are enablers for flow reconstruction and short-term forecasts (for example, 10 minutes).

This dissertation will investigate the accuracy of statistical models designed to estimate wind parameters as wind speed and absolute wind direction. To achieve this, it will be used information from measurements of every turbine generated by meteorological sensoring equipment mounted on the nacelle of every turbine, such as anemometers and wind vanes.

1.2. Goals

The main goal is to develop a model or methodology able to reconstruct wind field characteristics such as wind speed and wind direction. The development of this model is based solely on the real-time measurements of every other turbine in the park. The goal is to understand, while trying to estimate these wind variables, how wind turbines can operate in very distinct conditions such as different ranges of wind speed and wind direction. Understand how every turbine interacts and possibly affect one another in a

wind park and what impact does a wind park have not only in the performance of the turbines but in the way these models can accurately represent every measurement for all turbines.

1.3. Dissertation outline

Firstly, the next chapter of this dissertation presents a background of wind theory and information about wind turbines that are relevant to understand the scope and important of the work that was developed. After this, the third chapter will present and explain the initial dataset of the wind park used for this investigation and the statistical analyses that occurred.

The next chapters, the fourth and the fifth, present all the models developed, including their theoretical explanation and the practical results that was possible to obtain for each one of them.

In these two chapters, after testing the first four models, a new approach is experimented, considering the creation of wind particles that are capable of predicting the wind speed and absolute wind direction of random turbines in the park, and lastly, a machine learning algorithm was experimented, to test the hypothesis that an artificial intelligence branch will show better results.

Subsequently, by the end of this dissertation, conclusions and expectations are drawn of all the possible results with the seven models.

2. Literature Review

2.1. Wind Turbine

2.1.1 Basics of a wind turbine

Fossil fuel sources such as oil and gas have been losing economic feasibility due to increasing scarcity, environmental concerns, and the cost of extraction. Renewable energy sources are considered viable options to replace conventional energy sources. Wind energy is one of the most promising renewable energy sources. [1]

The force of the wind can be powerful, demonstrated after the passage of a hurricane or a typhoon. Historically, humanity has collected this visible force peacefully. One of its most essential usages was the propulsion of ships using sails before the invention of the steam engine and the internal combustion engine. The wind was also used in windmills to grind grain or to pump water for irrigation. This practice is still used in modern days, in some regions with an agriculture purpose. At the beginning of the twentieth century, electricity came into use, and windmills gradually became **wind turbines** as engineers connected the rotor to an electric generator. Unlike windmills used directly to pump water or grain grinding, wind turbines convert wind energy to electricity.

Progressively, wind power began to be a reliable source of power, which means it began to be a source of power that consistently showed good results. People realized that a country or region where energy production was based on imported coal or oil would become more self-sufficient using wind power alternatives. Electricity produced from the wind does not produce $C0_2$ emissions and, therefore, does not contribute to the greenhouse effect, so it is a clean and renewable energy source. [2]

In order to understand the impact of this renewable energy source on energy production worldwide, it is vital to understand the primary pieces of knowledge of how wind turbines work.

A wind turbine transforms the kinetic energy in the wind to mechanical energy in a shaft and finally into electrical energy in a generator. The maximum available energy, P_{max} (equation (1)) can be obtained if theoretically the wind speed could be reduced to zero:

$$P_{max} = \frac{1}{2}\rho A v^3 \quad (1)$$

Where ρ is the air density, v is the wind speed, and A is the area where the wind speed was reduced. This equation is fundamental since it demonstrates that power increases with the cube of the wind speed

and only linearly with the air density and swept area. The available wind speed at a given site is therefore first measured over a short period of time (1 to 5 years) or using meteorologic models before a project is initiated. [2]

In practice, it is impossible to capture all the kinetic energy that the wind has and reduce its speed to zero since the wind, as a fluid, needs to have some velocity to pass through the rotor. It can never decrease to an absolute zero. If a turbine trapped 100% of the energy in the wind, the wind would stop blowing, and the blades would not be able to turn and produce electricity. Because of this, a power coefficient C_P is a ratio between the actual power obtained and the maximum available power as given by equation (1). Since it is impossible to extract all the energy from the wind, it was defined a theoretical maximum for this power ratio, a fundamental physical law for the aerodynamics of wind turbines, denoted as the Betz Law. [3]

The Betz theory defines that no horizontal axis wind turbine can extract more than 16/27 (59.26%) of the wind's kinetic energy. The factor 16/27 (0.593) is known as the Betz limit. As shown in Figure 1, $\overline{u_1}$ and $\overline{u_4}$ are mean velocities far upstream and downstream from the wind turbine, and it is possible to see that downstream, the area of the airflow will expand, but the velocity is lower since it was not possible to capture all the energy of the wind coming upstream. [4] [5]

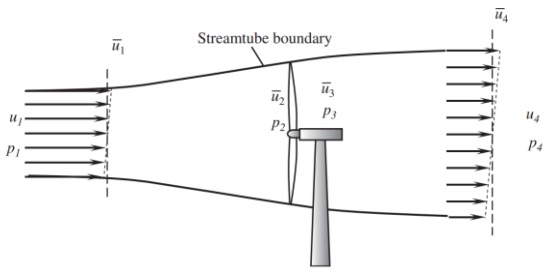


Figure 1 – Example of an airflow through a wind turbine

Modern wind turbines tend to operate very close to this limit, with a power coefficient (C_P) up to 0.5, and are, therefore, more optimized. Lift and drag are force components that can represent the propulsion of the blades: lift is the force component perpendicular to the direction of the relative wind, and drag is the force component parallel to the direction of the relative wind. Theoretically, it is elementary to understand that it is much more efficient to use the force lift rather than drag for the wind turbine to extract power from the upcoming wind, since lift is the opposite force to the gravitational force of the blade (which has the same signal as drag). Because of that, all modern wind turbines consist of several rotating blades with a shape of a propeller aerofoil in order for the lift force to be the leading force

component when the wind reaches the blades. If the blades connect to a vertical shaft, the turbine is called a vertical-axis machine (VAWT), and if the shaft is horizontal, the turbine is called a horizontal-axis wind turbine (HAWT). The advantages of this type of wind turbine include the high turbine efficiency, high power density, low cut-in wind speeds, and low cost per unit power output. All the wind turbines that the dissertation focuses on are horizontal-axis turbines, the most common type on the market. In Figure 2, this type of wind turbine is represented, as long with the wind velocity $\bar{u}$ and the swept area from the rotor A. [2] [4]

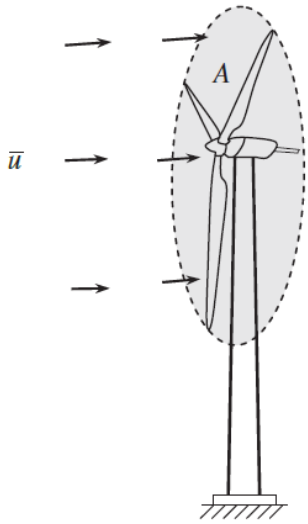


Figure 2 – Example of a horizontal-axis wind turbine

Based on the configuration of the wind rotor concerning the wind-flowing direction, the horizontal-axis wind turbines can be further classified as upwind and downwind wind turbines. Most horizontal-axis turbines used today are upwind turbines, in which the wind rotors face the wind. The main advantage of upwind designs is to avoid the distortion of the flow field as the wind passes through the wind tower and nacelle. The wind turbines further evaluated in this paper are classified as upwind turbines.

The tower height is also an essential parameter since wind speed increases with height above the ground. The rotor diameter is also critical since this gives the area A (rotor swept area), used in equation (1). The ratio between the rotor diameter and the hub height is often approximately one. The rated power is the maximum power allowed for the installed generator. The control system will then guarantee that this power does not exceed in high winds.

Ideally, a wind turbine rotor should always be perpendicular to the direction of the wind to obtain maximum energy. On most wind turbines, a wind vane is therefore mounted somewhere on the turbine

to measure the direction of the wind. With the information coming from this signal, it is possible to adjust the yaw angle of the turbine (through yaw motors mounted on the nacelle) and continuously turn the nacelle into the wind. For high wind speeds, the yaw angle of the nacelle is constantly being adjusted to the direction of the wind, but for lower wind speeds, this angle is not adjusted so frequently. The sensors used are mounted in the nacelle on the wind turbines that will be analysed. The sensors used are an FT sensor, one wind vane, and one wind cup. Their location on the wind turbine is represented in Figure 3. [2]

Figure 3 – Example of wind turbine from Vestas and the location of the sensors on the nacelle

Another vital part of the wind turbine's components is the nacelle. In Figure 4, it is possible to see a typical wind turbine configuration, with the nacelle being the structure that can hold and house components such as the main shaft, the gearbox, and the generator. As for the blades, using a pitch control system (Figure 4), each blade is pitched individually to optimize the angle of attack of the blade for allowing a higher energy capture in regular operation and for protecting the turbine components such as the blades, the nacelle, and the tower, from damaging in emergencies. With the information from the wind vane, such as measured instantaneous wind direction and speed, the yaw control system provides the yaw orientation control for ensuring that the turbine is constantly against the wind. [4]

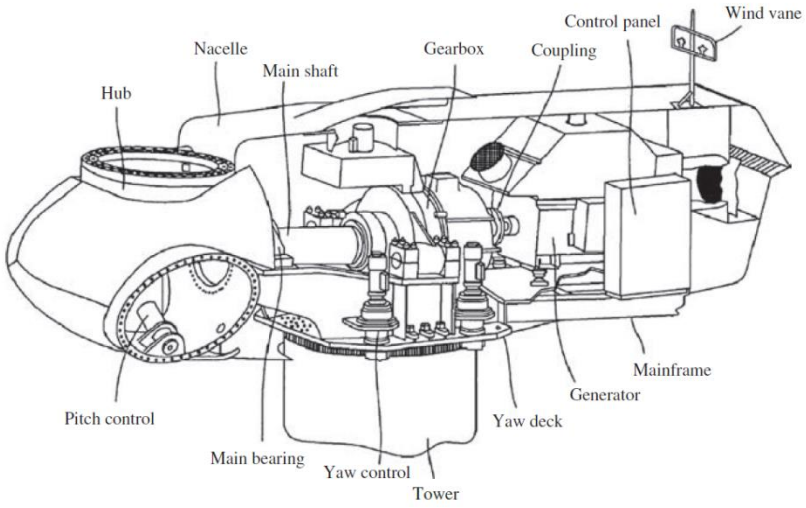


Figure 4 – Example of a horizontal-axis wind turbine configuration

A wind turbine converts wind energy into electricity by using the aerodynamic force in the rotor blades, and these blades work as an airplane wing or helicopter rotor blade. When wind flows through one of the blades, the air pressure decreases on one side of the blade. The difference in pressure for both sides of the blade will create lift and drag forces. The force in the lift is more substantial and causes the rotor to spin. The rotor connects to a generator (directly or through a shaft and gearbox), which speeds up the rotation and allows for high rotation values in the generator, creating electricity. For maximizing the capture of wind energy, wind blades need a careful design. [6]

Mechanical energy captured by wind blades converts into electrical energy via wind generators. However, three parameters can influence this converting efficiency: the gearbox efficiency, the generator efficiency, and the electric efficiency. Therefore, the power output from a wind turbine is slightly different from equation (1) since we need to consider the efficiency of these three parameters. So, the actual power output from a wind turbine becomes:

$$P_{eff} = \frac{1}{2}(\eta_t \rho A \bar{u}^3) \times C_p \qquad (2)$$

As shown in equation (2), the actual electrical power output from a wind turbine is directly proportional to the available wind power and the total effective wind turbine efficiency.

The wind turbine's power curve displays the power output, either the actual electrical power output or the percentage of the rated power, of the turbine as a function of the mean wind speed. Field measurements usually determine power curves. In Figure 5, it is possible to understand that the wind

turbine starts to produce usable power at a low wind speed, defined as the **cut-in speed**. The power output increases continuously with the wind speed increase until reaching a certain point, representing the value where the power output reaches its maximum. This point is defined as the **rated power output**. Correspondingly, the wind speed at this point is the **rated speed**. So, from this point forward, even if there is an increase in wind speed, the power output will remain the same. That is possible because when a wind turbine reaches this point, power control mechanisms are activated. These control mechanisms will be discussed further in chapter 2.1.2. Eventually, when the wind speed becomes too high and can potentially damage the wind turbine, it needs to be shut down immediately to avoid damaging its components and structure. This value of wind speed is the **cut-out speed**. Thus, the cut-in and cut-out speeds are the values that define the operating limits of the wind turbine. [4]

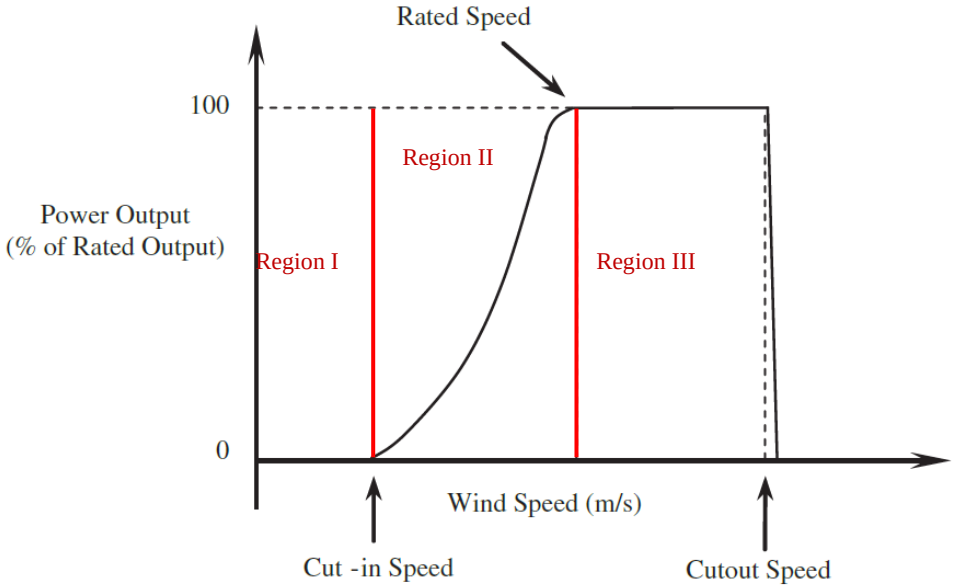


Figure 5 – Example of a typical wind turbine power curve

2.1.2 Control and loads

In a wind turbine power curve, it is possible to distinguish different operating regions. Each region has its control strategy and is typically determined based on a generator speed feedback signal. The ideal power curve for a variable pitch/speed wind turbine is shown in Figure 5 and exhibits three central regions with distinct control objectives. In **region I**, the wind turbine produces energy since wind speed will turn the rotor. However, the amount of energy produced is lower than the amount of energy spent to rotate a wind turbine. When wind speed reaches a particular value, the power produced starts to compensate for the spent power. In **region II**, as the wind increases, the power output increases much more rapidly (proportionally to the cube of the wind velocity). When the power produced by the wind turbine reaches

a value that could compromise the safety of all the equipment, it needs to be controlled and kept to a constant value. So, in this region, as the wind speed increases, the power output by the wind turbine remains the same, and that is only possible with control mechanisms that guarantee the safety and protection of the wind turbine. In **region III**, the wind turbine is designed to keep working at the rated power. Even for values of wind speed above the rated wind speed, this power cannot exceed. To maintain the same power output, even when wind speed values reaching the rotor are increasing, the pitch angle of the blades is continuously adjusted to keep the power at a constant level. [7]

Control systems are not only important for a single turbine but also for conditions where multiple turbines work at the same time and space in wind farms. This dissertation mentions this type of control in the next chapter 2.2.

2.2. Wind Farms

In this dissertation, one of the analysis subjects is how turbines work together and produce energy in a wind farm. A wind turbine is seldom located alone and, to produce maximum energy, it is necessary to place with other turbines, i.e., a wind farm or wind park. Placing wind turbines together has many benefits, such as reduced deployment costs of the turbines and reduced operation and maintenance costs.

Wakes can reduce performance in wind farms. As a wind turbine extracts energy from the wind, it causes a change in the wind flow downstream from that turbine. The altered flow is called the wake of a turbine. The wake characteristics are space-, time-, and parameter-dependent. A wake is space-dependent because, e.g., far downstream of a turbine, the wake is much less intense than closer downstream of a turbine. The wake is also time-dependent because the operation of a wind turbine changes over time and the surrounding flow. Correct wind farm control is critical because wind direction and speed do not remain constant over time. Finally, the wake is parameter-dependent, as external variables, such as temperature, that influence the wake's behaviour.

As a result of increased turbulence in the wake, fatigue loads on the downstream turbine can increase and shorten its lifetime, and, in most cases, the center of the wake will not coincide with the center of a downstream rotor, and the wake is skewed. Because of this, there is more thrust on one side

of the rotor which can lead to accelerated structural degradation of waked turbines. Studying and modelling a wake is a broad research topic by itself, and there is still ongoing research.

The wake created by every turbine will lead to a time-varying interaction between the individual turbines. A common practice in the industry has been controlling turbines individually and ignoring this effect while optimizing the power and loads of the individual turbines. Turbines that are affected by the wake of others experience reduced wind speed and increased turbulence, leading to reduced energy extraction and increased dynamic loads on the turbine, respectively. Neglecting the dynamic interaction between turbines in control is not the best approach since it will lead to suboptimal performance of the total wind farm. Maximizing the energy production in wind farms has been the most important issue until now and is likely to remain the same in a near future. [8][9]

2.3. Wind Field Reconstruction

Wind field reconstruction is a challenge in the wind industry to understand better the conditions in which turbines operate or some wind field characteristics such as wind speed, direction, horizontal (veer) and vertical gradients (shear), and turbulence. This process combines data containing information about wind field parameters. It uses a model to predict or reconstruct the conditions of wind at a given location with the available data.

This type of flow reconstruction can understand and determine some of the conditions that compromise wind turbines, such as the variation of wind speed and wind direction, including wind gusts (a sudden and brief increase of the wind speed). With this information, flow reconstruction models can predict the wind reaching a specific turbine at a specific time.

That type of information can be helpful, for example, in situations where a sensor fails and cannot measure a given variable, and for this flow reconstruction is used to predict the values of wind characteristics when the sensor is not working. With this, the turbine can still receive information from where the wind is coming from and work.

Because of this, wind forecasting becomes even more critical for predictive turbine and wind farm control. [8][10]

There is active research on this topic, highlighting the relevance of this dissertation. Many papers have good results in that they can predict the values of wind speed or direction on one specific turbine with a small amount of error. For example, in [11], the authors use wind speed measurements coming

from LiDAR measurements. LiDAR (Light Detention and Ranging) is an accurate and reliable remote sensing technology for wind speed measurements. It is becoming a mature technology for remote wind speed or direction measurements and rotor effective wind speed predictions. This paper attempts to fit a wind model with the LiDAR measurements using an iterative procedure. They consider a 2D wind field reconstruction with a fixed altitude, a similar situation with this dissertation. Many papers, such as [12] and [13], work with measuring devices such as LiDAR and Doppler radars, which tend to be more reliable and expensive than a normal anemometer. However, this dissertation will be working with measurements from sensors located on top of the nacelle, a part of the wind turbine that may cause some deviations in the measurements since located so close to the rotor. In [14], the authors also use information from sensors located on top of the nacelle and attempt to create a model that is able to predict values of wind speed and wind direction of one specific turbine with the measurements from all the remaining turbines in the park, using a simple Lagrangian advection scheme. In this dissertation, the measurements from every turbine were also used, but it will try to predict and reconstruct the wind flow of all the turbines in the park and not just a specific one.

3. CASE STUDY

3.1. Wind Park

The data for this dissertation belongs to one wind park in flat terrain with very few obstacles. Within this park, the number of turbines that will be analysed is 20 turbines. However, there are 36 turbines in this park. The remaining 16 turbines will not be analysed due to lack of data, but they will be mentioned as an additional parameter to evaluate further results of every turbine. For this wind park, turbines are laid out in positions to obtain the maximum amount of energy. For the protection of information purposes, the location and more detailed characteristics about the park will remain confidential and the location of every single turbine. In Figure 8, it is possible to see the park layout and the display of every turbine within this wind park. The coloured turbines in the figure represent the 20 turbines that will be studied.

Since the location of every turbine needs to remain confidential, it was necessary to anonymize their values of latitude and longitude. Data anonymization is a process to protect private or sensitive information by erasing or encrypting identifiers that reveal the data. As a result, one turbine was chosen as the origin of the referential with a latitude and longitude of 0 (represented by turbine 1 in Figure 8). After this, all the remaining locations of the other turbines were altered, calibrated for turbine 1.

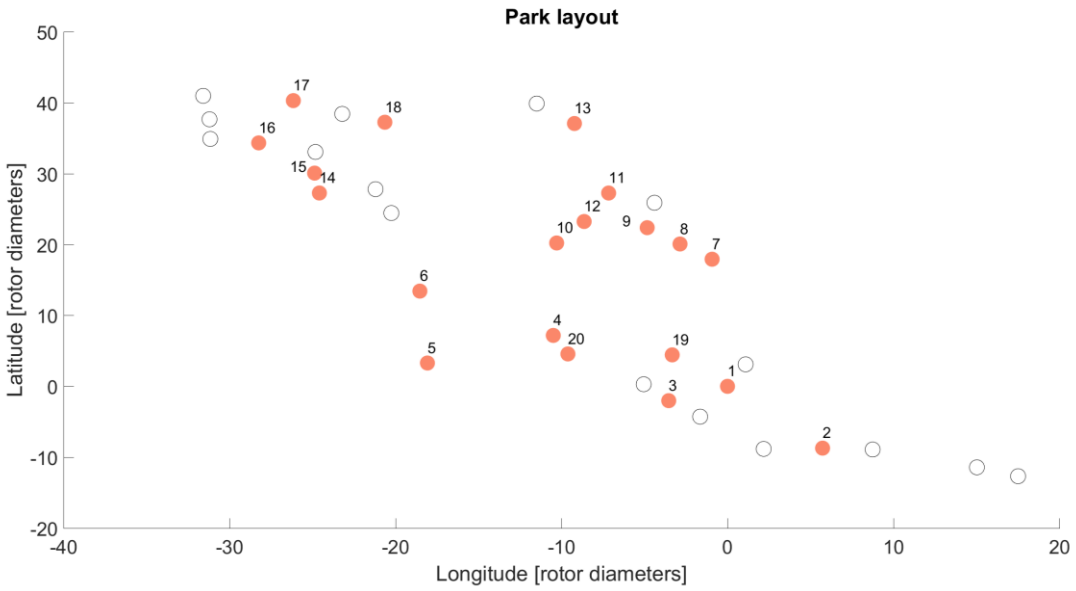


Figure 8 – Graphic representation of the park layout

3.2. Dataset

3.2.1 Characterization of the dataset

The dataset had information from over 100 signals, representing different sensors and different calibrations of measurements throughout a day. The data used was collected for two months (initial data from 26th of October to 6th of February of 2019) and a 10 Hz frequency. These signals can provide information for wind speed, absolute and relative wind direction, nacelle position, and grid power. Because of the amount of information coming from different signals, it was essential to filter the type of signals that would be more useful for the dataset necessary.

Despite having a wide set of initial data, it is important to select and filter this dataset to be able to interpret the data and develop future prediction models. The initial dataset needs to pass a few steps known as pre-processing data, which will be discussed in the next chapter. In Figure 9, it is possible to see an example of the variable wind speed (v) of the initial dataset for turbine 1. The initial data has a high frequency and much noise that can easily change the predicted values of variables such as speed and direction, as well as large periods of missing data.

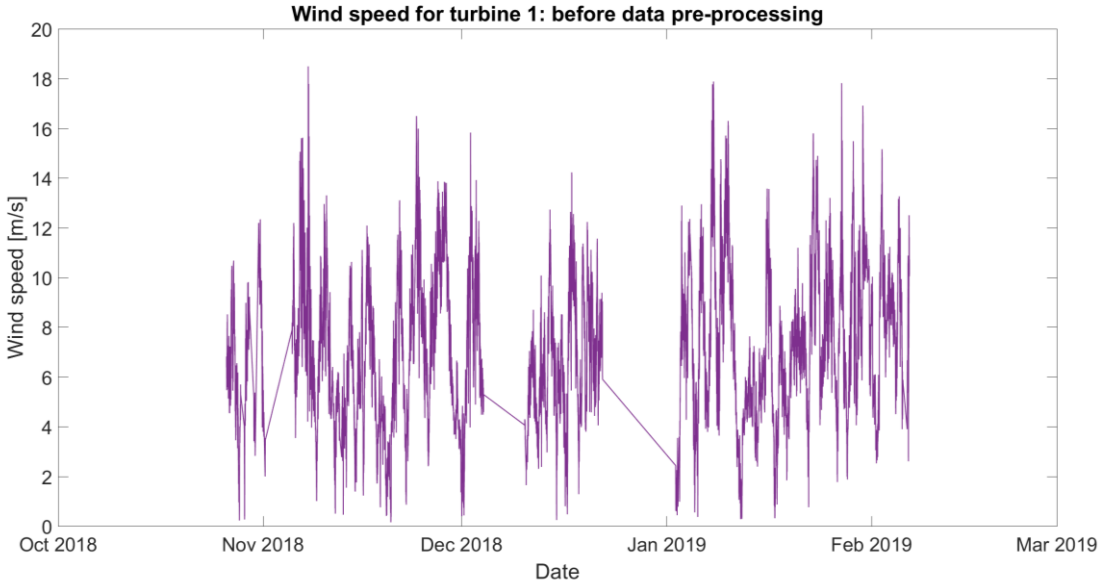


Figure 9 – Example of a time series of wind speed for turbine 1 with the initial dataset

3.2.2 Pre-processing data

In data analysis, especially when working with a software such as Matlab, pre-processing data is often of critical importance if one wants to obtain good results, and in this case, to construct a good prediction model. Un-processed data is generally incomplete, contains noise and outliers, and is inconsistent. Because of this, it is required techniques such as pre-processing, which are designed to compensate for deviations, handle missing values, remove outliers, and resolve conflicts. Because data is often taken from multiple sources, it is simply unrealistic to expect that the data will be perfect. There may be problems due to human error, limitations of measuring devices, or flaws in the data collection process. When there are so many possible sources of error, some conflicts can appear that pre-processing data will fix, such as missing values and inconsistent values. Some datasets may include data objects that are duplicates of one another. [15] [16]

For this dissertation, the data considered are measurements coming from sensors placed in wind turbines from a real wind farm. The inconsistency in measurements coming from sensors affected by temperature, wind, and pressure needs to be treated. The pre-processing steps that were necessary for the original dataset were the following:

1. The information from the initial 127 signals was reduced to 55 signals, according to their relevance for the work
2. Data resample from an initial frequency of 10 Hz to a frequency of 0.0017 Hz (10-minute period) that was used for the development of the models and a frequency of 1 Hz (1 second) used explicitly for the particle model (chapter 4.2)
3. Creation of time-series for every turbine's data
4. Synchronization of all time-series concerning turbine 5 (turbine with the smaller dataset)
5. Calibration of the North position by eliminating the error associated with absolute wind direction and nacelle position measurements)
6. Elimination of periods that had missing sensor data.

The dataset ranged from the dates of 26th of October to 6th of February. For a 10-minute period, the shortest time series ended on the 15th of December. For every variable of the dataset to be synchronized, it was first necessary to create time-series, the step before the synchronization. A time series is a time-oriented sequence of observations on a variable of interest. For this dissertation, the variables chosen were wind speed, absolute wind direction, and the absolute time of the measurements. After guaranteeing that the time vectors of all turbines had equal lengths and that all variables were in a time series format, it was now necessary to proceed with the next step: synchronize all turbines with the same time vector. The synchronization function in Matlab collects the variables from all input data,

synchronizes them to a common time vector, and returns the result as a single timetable. The method used for resampling data in synchronization is a linear interpolation of the data.

The 5th step in these pre-processing methods was the North calibration. A technician is responsible for performing these calibrations using a compass or a mobile phone approximately 200 meters away from the turbine. After that, the values are saved and inserted into the wind park software. That means there are two possible sources of error in these measurements: the technician did not measure correctly, or the equipment used was not precise. Since the calibration of turbines is not entirely accurate, the next logical step was to guarantee that the absolute wind direction measured by the turbines in the park and the nacelle position were calibrated to the same position, i.e., to the same North. For a wind farm, the value 0° of absolute wind direction is considered the geographical North, and the values of wind direction will vary from 0 to 360°. Every turbine position (latitude and longitude) was calibrated, considering that turbine 1 was at position (0,0).

For this portion of pre-processing data, considering a deviation for every wind direction measurement, it was necessary to find the value of that deviation for every turbine in the park. Because of this, it was necessary to calculate the mean of the absolute wind direction (φ) (equation (3)) and execute the same procedure for the values of the nacelle position for every turbine. After that, the error (γ) was calculated, associated to that direction, by subtracting the values of wind direction and nacelle position measured for every turbine in the park with the mean of the wind park (equation (4)), previously calculated. The result is the error associated with the measurements for every turbine, the deviation necessary to find. The next step was to calculate the mean error for the two variables (equation (5)).

$$\bar{\varphi}_{park} = \frac{\sum_{i=20} \varphi_i}{Number\ of\ turbines} \tag{3}$$

$$\gamma_i = \varphi_i - \bar{\varphi}_{park} \tag{4}$$

$$\bar{\gamma} = \frac{\sum_{i=20} \gamma_i}{Number\ of\ turbines} \tag{5}$$

After this, the error found in equation (5) was removed from the measurements of absolute wind direction and nacelle position at every turbine. With these steps, every turbine was calibrated for the same North. The absolute wind direction before (left) and after (right) the calibration of every turbine is represented in Figure 12. As it is visible in the figure, after calibration the turbines present more coherent values of absolute wind direction.

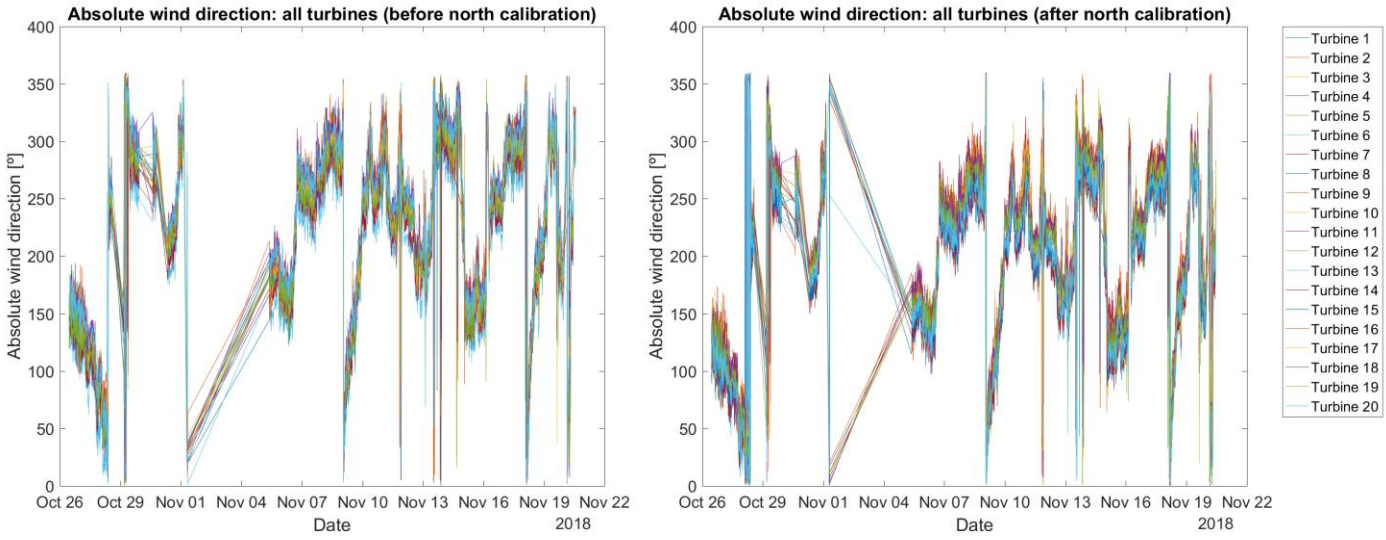


Figure 10 – Graphic representation of the absolute wind direction for every turbine before (left) and after North calibration (right)

Another step of pre-processing data that was crucial to develop some of the models will be explained in chapter 4, and it was the calculus of the distances and angles between every turbine, and merely from the values of latitude and longitude that existed for all turbines. To calculate the bearing and distances between two objects with their geographic coordinates, equations (9) and (12), retrieved from [17] were used.

Consider points A and B as two different turbines where LA represents turbine's A longitude and θA represents its latitude (in degrees). The same thing goes for turbine B, ΔL represents the difference between the values of longitude of the two points and β represents the bearing or the angle between the turbines, the variable that was necessary to calculate:

$$x = \sin(\Delta L) \times \cos(\theta b) \tag{6}$$

$$y = \cos(\theta A) \times \sin(\theta B) - \sin(\theta A) \times \cos(\theta B) \times \cos(\Delta L) \tag{7}$$

$$y = cos(\theta A) \times sin(\theta B) - sin\,(\theta A) \times cos\,(\theta B) \times cos\,(\Delta L) \tag{8}$$

$$\beta = atan2(x, y) \tag{9}$$

For distances between turbines, a value of 6371 km was assumed for the radius of the Earth and applied the following equations:

$$aux1 = sin\left(\frac{\Delta\theta}{2}\right) \times sin\left(\frac{\Delta\theta}{2}\right) + cos(\theta a) \times cos(\theta b) \times 2sin\left(\frac{\Delta L}{2}\right) \tag{10}$$

$$aux2 = 2 \times atan2(\sqrt{aux1}, \sqrt{1 - aux1}) \tag{11}$$

$$d = R \times aux2 \tag{12}$$

In this equation, θa represents the latitude of one of the turbines but in radians, and d represents the final value of distance in kilometres between the two points. The final table contains values of angles and distances between every single one of the turbines in the park is represented in Annex A. This table was very useful for developing some of the models and further understanding the turbines' interactions.

3.3. Dataset analysis

3.3.1 Statistical analysis

Usually, when working with data sets, it is important not only to pre-process it (chapter 3.2.2) but also to uncover all the information from its values. With this, it is possible to improve future work, which in the case of this dissertation, means discovering new information that could improve the results of the models.

Correlation between wind speed and distances

In the wind industry, when working in wind field reconstruction with information coming from various turbines in a wind park, it is often assumed that closer turbines have a higher correlation. However, this is an assumption merely on logic since it is reasonable to consider that turbines closer to the one whose values are estimated are more reliable and have a higher correlation than others who are much more distant.

A correlation analysis was proceeded between the real wind speed values for every pair of turbines to understand if this assumption was correct. A correlation analysis between two variables is a statistical measure that indicates the extent to which two or more variables fluctuate together and determines how strongly the pair of variables is related. So, a higher correlation of the real wind speed values between a pair of turbines means that that pair will be more related to each other and, in that sense, the information coming from one or the other will be more reliable. [18]

After obtaining the correlation values for this analysis, it was necessary to relate these values with the distances between every turbine to conclude from this information. For this comparison, since the wind conditions in the park, such as wind speed and absolute wind direction, will vary through time, different intervals of wind speed were considered and different sectors for wind direction. For wind speed, three intervals were considered: [0-5] m/s, [5-11] m/s and [11-13] m/s; for absolute wind direction, two sectors were considered: [90-180]° and [180-360]°. The wind speed intervals were chosen considering the distinct regions on a wind turbine's power curve (Figure 5). The cut-in speed would be 5 m/s, the interval [5-11] m/s represents the median values, and 13 m/s would be the rated speed. These values were reached after analysing all the wind speed measurements of every turbine. The wind direction values were obtained considering the main values of absolute wind direction reaching this wind park. A wind rose was created to understand where the main wind direction for this park. The wind rose that characterizes this park is represented in Figure 13.

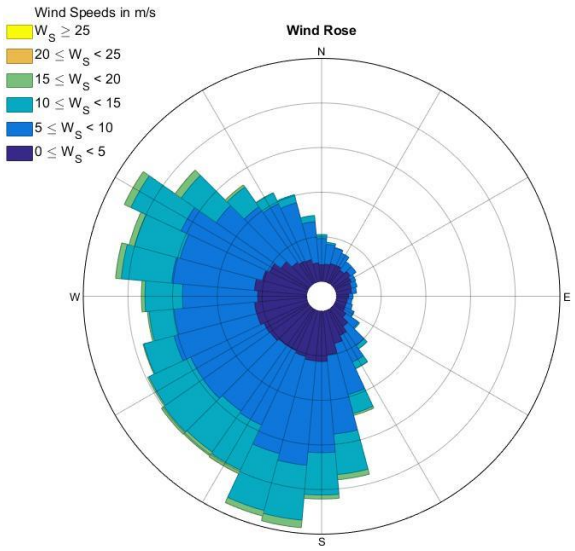


Figure 13 – Visual representation of the wind rose for the wind park

In this figure, the South represents 180°, the West represents 270°, and the North 0 and 360°. After analysing this figure, one can conclude that most of the wind that reaches the park comes from the sector [180-360] °, but both sectors were used in this analysis. The wind speed intervals in this figure are merely representative of this figure and are not related to the intervals considered for this statistical analysis.

So, the correlation values were obtained and represent in a way that would be easy to understand, creating a graph. In Figure 14, it is possible to see the graph that represents the correlation values for wind speed between turbines, for the last wind speed interval considered ([11-13] m/s) and wind direction sector [180-360]°. The graphs for the remaining interval values are located in Annex B.

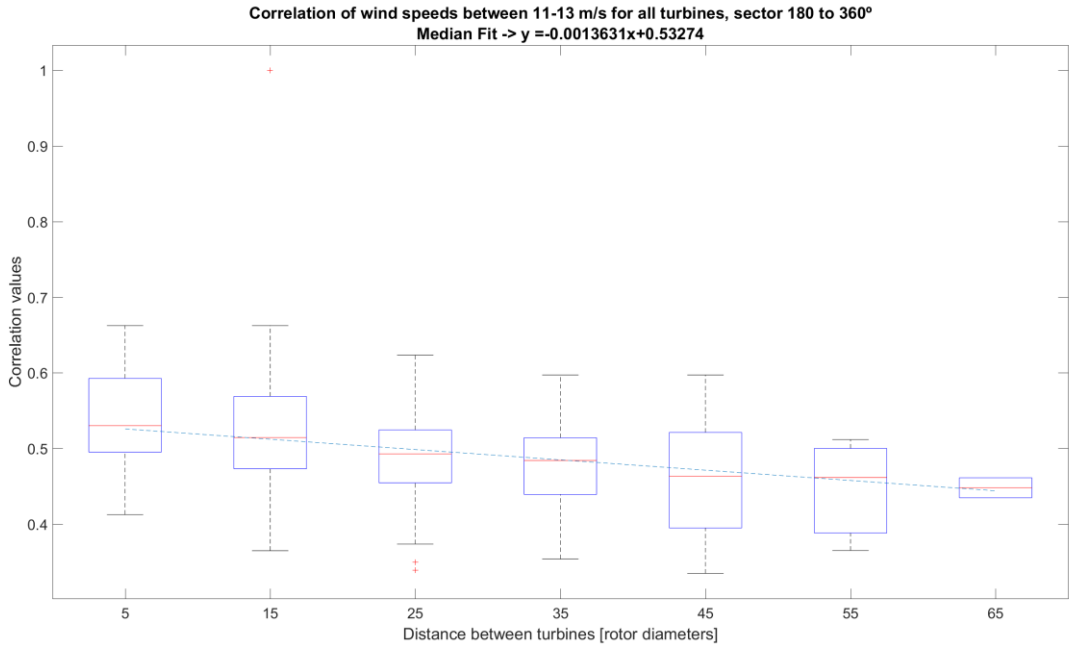


Figure 14 – Graphic representation of the correlation values between wind speed and distances between turbines for a wind speed interval of 11 to 13 m/s and wind direction sector of 180 to 360°

With this analysis and interpretation of these graphs, it was possible to conclude that there was a slight correlation between measured values of wind speed in every turbine and their respective distances. It is possible to see from the graph since the equation fitting is decreasing, which means higher distances have lower correlation values. So even though the slope of the median fit between the two variables is not as high as expected, it still has a negative slope, proving an inverse relation between the correlation values and the distance between turbines. This means that, for some cases, the assumption that distance affects the correlation between two wind turbines is correct. However, this does prove that closer turbines will not always be the highest correlated turbines because of the shortened distance.

Delays between turbines

In a wind farm, considering the layout of the park and different values of absolute wind direction and wind speed, the wind will reach the turbines at different times, and, sometimes, an inevitable delay between the information reaching two turbines can be found. For this park, an analysis on Matlab was made to understand if any delays between the turbines existed.

First, an analysis for 10-minutes (0.00167 Hz) was proceeded, the regular frequency for most models. However, it was impossible to see any delays since, within a range of 10 minutes, the wind had already passed through the whole park. Because of this, the analysis was repeated for a frequency of 0.0167 Hz (1-minute period data). In a wind park, no matter the frequency of the data, there will be periods with different wind speed intervals and wind direction sectors since the characteristics and conditions of the wind fluctuate over time. So, for this analysis, the same wind speed intervals as the previous analysis were considered, and for wind direction, it was considered 90-180° and 270-360°, excluding the sector from which most of the wind in the park comes from, which is 180-270°.

For the entire park, eight pairs of turbines with a delay were found for the first wind speed interval considered for the wind direction sector 90-180°, and nine pairs of turbines for sector 270-360°. Five pairs of turbines for the second wind speed interval were found for sector 90-180° and six for sector 270-360°. Lastly, for the third wind speed interval, for both wind direction sectors considered, no pairs of turbines with delays were found, concluding that for higher wind speed values, with this frequency, the wind had already swept the whole park.

As mentioned before, pre-processing data is an essential step in preparing the existing data and better understanding the information given. With this analysis, it was not possible to conclude if there could be any relation between the pair of turbines that had a delay related to the information of the wind passing through both of them.

Feature selection analysis

Another analysis studied was the feature selection analysis to understand further the relations between turbines and values of correlation between some of the signals such as wind speed and absolute wind direction. Feature selection is used in machine learning and data analysis to reduce the number of variables studied. This type of process is advantageous to decrease the computational cost of modelling and, in some cases, improve the model's performance. This method was chosen to better understand

the relations between turbines and find the pairs of turbines with a higher correlation. The traditional approach to this analysis was not used since the number of variables remained constant. Instead, it was used to rank the signals in order of relevance and to understand some of the information in these signals (wind speed and absolute wind direction).

Feature selection is a process of data analysis and there are various methods that can be used for this analysis. The method chosen for this dissertation was a mutual information function. Mutual information is a measure between two variables that quantifies the amount of information obtained about one random variable through the other. The mutual information function is given by equation (13) .

$$I(X;Y) = \int_X^1 \int_Y^1 p(x,y)\, log \frac{p(x,y)}{p(x)p(y)} dy\, dx \tag{13}$$

Where $p(x,y)$ is the joint probability function of variables X and Y and where $p(x)$ and $p(y)$ are the marginal density functions. The mutual information will determine the relation between the two variables. If X and Y are entirely unrelated and independent of each other, then the integral of this function will be zero. [19]

Mutual information was used to quantify how much information it was possible to retrieve from every pair of turbines in the park. In Figure 15, it is possible to see a representation of this analysis for turbine 1. This turbine is singled out in the colour pink, and it is possible to see how correlated it is to the others, analysing the colour bar on the right. For this turbine, it is conceivable to conclude that turbines closer to the turbine analysed, such as turbine 2, turbine 19, turbine 20, and turbine 4, have higher correlation values. Nevertheless, there are a few outliers, cases of turbines that are further away from turbine 1, with higher correlation values than closer ones. It is possible to see in Table 1 that, for example, turbine 10, which is in the middle of the park, has a higher correlation value than turbine 3, the closest turbine. This example could lead to the conclusion that distance does play a role in quantifying how much they are related to each other for most of the turbines. However, there are specific situations, such as the one mentioned, where this is not entirely correct and could be because of the park's layout and direction of the wind. The values present in the graph from Figure 15 and Table 1 are weights from the mutual information equation.

Table 1 – Correlation values of feature selection analysis for turbine 1

Turbine	Correlation values
4	0.8606
19	0.8283
2	0.8271
20	0.8119
5	0.7930
10	0.7819
9	0.7710
11	0.7596
18	0.7241
3	0.7193
7	0.7101
12	0.6989
17	0.6944
6	0.6856
13	0.6832
15	0.6728
8	0.6634
14	0.6302
16	0.6179

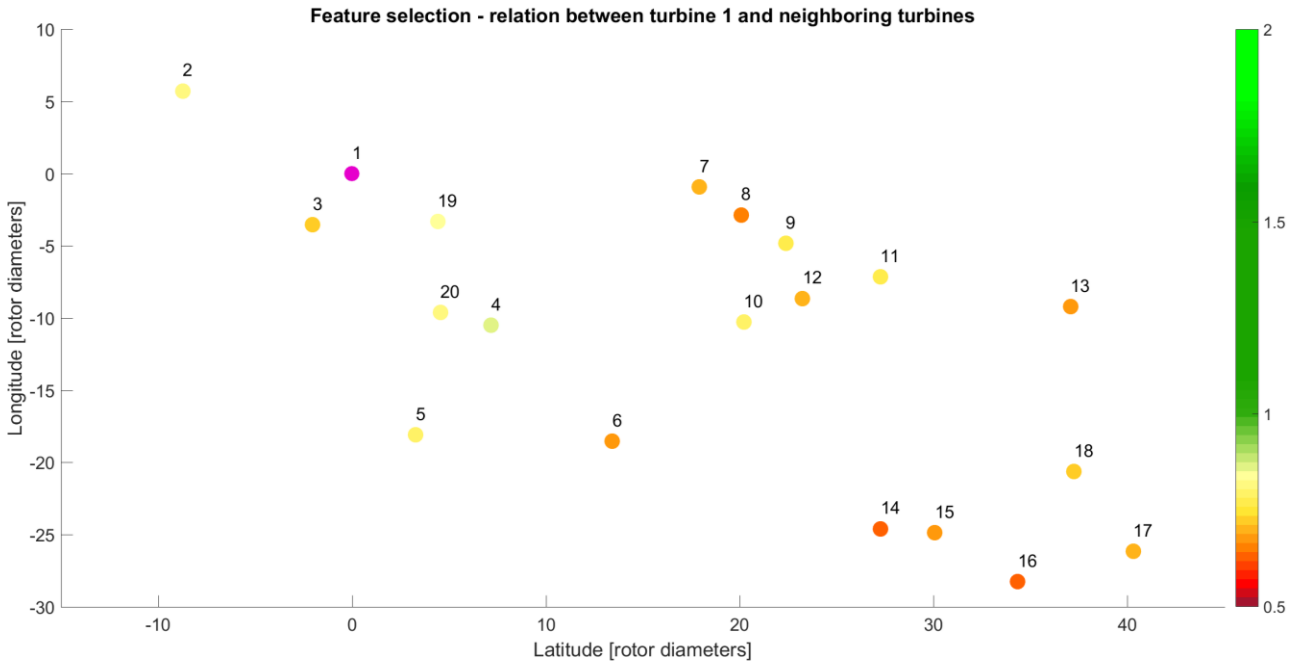


Figure 15 – Graphic representation of the feature selection analysis for turbine 1

The same analysis and same graph were obtained for the other remaining turbines. However, the conclusions that were possible with those graphs were the same as the example represented in Figure 15 – Graphic representation of the feature selection analysis for turbine 1 and Table 1. These results are represented in Annex C.

4. MODELS

4.1. Weighted models

4.1.1 Linear regression model

One of the objectives of this dissertation is to estimate and reconstruct wind variables such as wind speed and absolute wind direction values, and to achieve this, a linear model was chosen as a first approach. It was decided to start with this type of model to be more familiar with some of the parameters necessary in wind flow reconstruction models and better understand the first results.

For the regression model, the goal was to use the linear equations that describe the slope of wind speed and wind direction of every turbine to estimate the values of wind speed measured at a random turbine in the park. The linear equation used for this model is represented in equation (14), where $v_{est_{reg}}$ represents the values of wind speed that will be estimated for every turbine with the regression model, m represents the slope of the equation and it is a linear coefficient obtained by binning, v_i represents an average of the measured wind speed values at very turbine except the turbine that is being estimated, and b represents the coordinate with y = 0, and it is also a linear coefficient obtained by the software.

$$v_{est_{reg}} = mv_i + b \tag{14}$$

In Figure 16, it is possible to see the slope of the equation that describes the relation between the wind speed values of turbine 1 and turbine 3.

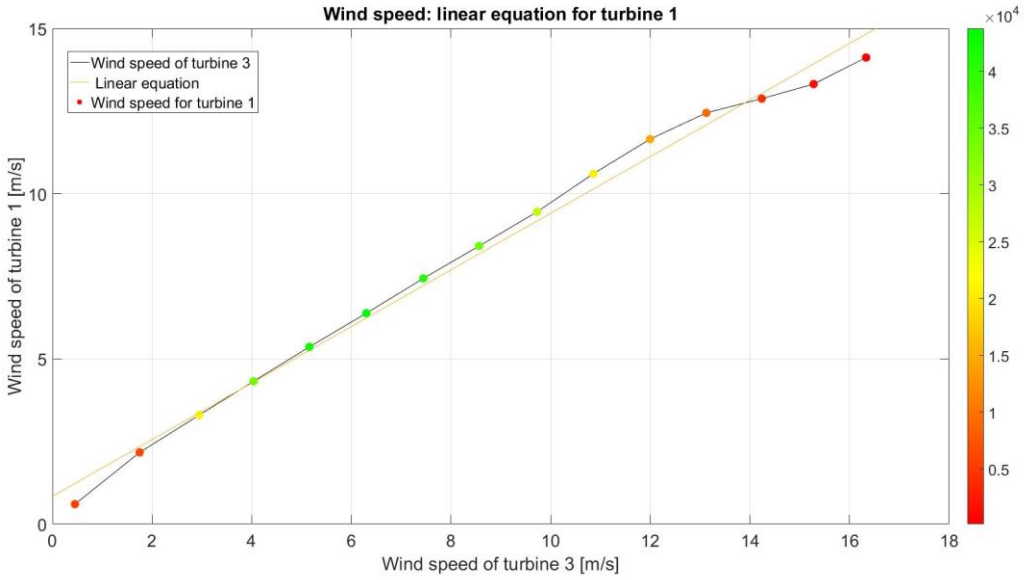


Figure 16 – Representation of the linear equation obtained after binning wind speed of turbine 1 and turbine 3

After obtaining the estimated wind speed values for every turbine through equation (14), it was essential to understand how the model performed. For this, it was necessary to calculate the error to see the differences between the real measured values and the estimated values: the root mean square error and standard deviation were estimated for every turbine (equations (15) and (16)) to conclude the accuracy of this model and its performance. This was made for all models, and standard deviation was only calculated for wind speed. In these equations, n represents the number of samples, y_i represents the values of the estimated variables and $\hat{y}_i$ represents the real measured values of the turbine.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{15}$$

These results are shown and discussed in chapter 5.

$$SD = \sqrt{\frac{\sum|y_i - \hat{y}_i|^2}{n}} \tag{16}$$

4.1.2 Distance weighted model

In order to start developing a new model, it was important to understand how to improve the results and decrease the error that existed with the first one. For this, it was understood that it was necessary to consider possible connections between the measurements of every turbine and apply different weights to the information coming from different turbines. These weights were chosen due to the correlation analysis in chapter 3.3.1, which showed a slight truth when assuming that closer distances mean a higher correlation between measured values and the ones estimated. It was desirable to prove this assumption correctly and understand that applying this logic to a model would improve the results. In Figure 17, it is possible to see that, for turbine 1, the information that can be extracted from turbine 19 will have a higher weight and a more significant impact than, for example, information from turbine 9, which will have a lower weight.

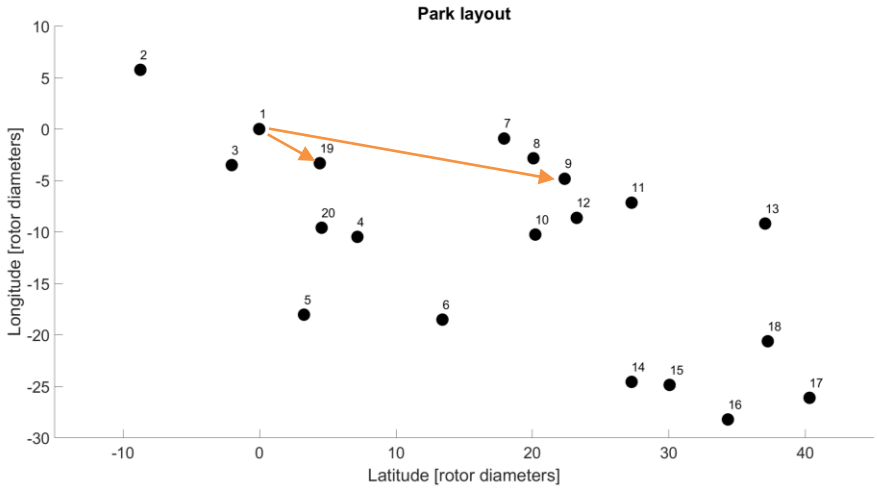


Figure 17 – Representation of the distances between turbines and the difference of the weights applied to their information

The equations to calculate these weights are shown in equations (17) to (21).

$$RD = \frac{d}{0.110} \tag{17}$$

$$weights_{dist} = \frac{1}{RD} \tag{18}$$

$$weights_{dist_{norm}} = \frac{weights_{dist} - weights_{dist_{min}}}{weights_{dist_{max}} - weights_{dist_{min}}} \tag{19}$$

$$v_{est_{dist_i}} = \sum_{i=1} v_{real_i} \times \overline{weights_{dist_{norm_i}}} \tag{20}$$

$$\varphi_{est_{dist_i}} = \sum_{i=1} \varphi_{real_i} \times \overline{weights_{dist_{norm_i}}} \quad (21)$$

Where d represents the distances between turbines in kilometres and these distances are converted to rotor diameters in equation (17) since 0.110 corresponds to the rotor diameter of the turbines in kilometres. RD then represents the final table of distances between every turbine, $weights_{dist}$ represents the weights calculated for the distance weighted model, $v_{est_{dist}}$ represents the values of wind speed being estimated for the distance weighted model and v_{real} are the real-time measurements, and $\varphi_{est_{dist}}$ represents the values for the variable absolute wind direction.

The distance weighted model represents a weighted average of real-time measurements of every turbine to determine the estimated wind speed and wind direction at a specific turbine in the park.

4.1.3 Correlation weighted model

After developing the distance weighted model, it was decided to experiment a different approach by using the same methods as the previous one but applying different weights to the real-time measurements. As elucidated in the correlation analysis from chapter 3.3.1, even though it proved that closer turbines will have a higher correlation for most cases, it is not 100% accurate to assume this will happen for every turbine. Because of this, the correlation weighted model, instead of using weights merely related to distances, uses the values of the correlation obtained for every turbine in this analysis. To retrieve the weights from this analysis, the equation representing the median fit of this correlation analysis for all wind speed values was used . The values for every turbine were calculated through equations (22) and (23).

$$weights_{corr} = -0.00034 \times d + 0.87 \quad (22)$$

$$weights_{corr_{norm}} = \frac{weights_{corr}}{\sum_{i=1} weights_{corr}} \quad (23)$$

$$v_{est_{corr_i}} = \sum_{i=1} v_{real_i} \times \overline{weights_{corr_{norm_i}}} \quad (24)$$

The first two equations represent the calculation of weights through the equation retrieved from the correlation analysis ($weights_{corr}$), and specifically equation (22) is a equation that represents the median fit achieved with the correlation analysis in chapter 3.3.1, when considering all values of wind speed. After that, it was necessary to normalize these weights to guarantee that the final sum was one (equation (23)). Subsequently, the next step was to estimate the values of wind speed for the correlation weighted model ($v_{est_{corr}}$), through equation (27). For this model, only wind speed values were estimated, since the correlation analysis was only tested for this variable.

4.1.4 Wake influenced model

For the next model developed, it was decided to continue the work developed in the distance weighted model. However, for this case, the distances between the turbines were considered a weighting factor and the angles between them. In the park layout represented in Figure 8, it is possible to see that the turbines have an associated distance between them and a respective angle (they are not all aligned with one another). As mentioned in [20], for a distance of less than three rotor diameters (3 RD) between two turbines, the probability of a wake being induced in one of the turbines is much higher. Because of this, for a distance between turbines lower than 3 RD, a different weighting process was considered. This weighting process considered the angles between pairs of turbines that were less than three rotor diameters away from each other. For an interval of measured absolute wind direction equal or similar to those angle values, the weight was calculated through a bell-shaped function. Thus, for values of wind direction where a pair of turbines could influence each other's measurements (possible wake caused by an upstream turbine), their information was weighted based on this function and tested two different scenarios.

These scenarios consisted of giving more weight to the information of these specific turbines. The remaining turbines had a constant weight calculated by the distance weighted model or giving less weight to the information of these turbines. For the remaining turbines, the same logic was applied.

The bell-shaped function describes a mathematical function with a characteristic bell-shaped curve. A bell curve is a common type of distribution for a variable, also known as the normal distribution. The equation for this function is represented below:

$$\begin{cases} weights_{wake} = \dfrac{h}{1 + \left|\dfrac{x - C}{a}\right|^{2b}} \\ weights_{wake} = \dfrac{1}{RD} \end{cases} \quad (25)$$

$$weights_{wake_{norm}} = \frac{weights_{wake}}{\sum_{i=1} weights_{wake}} \quad (26)$$

$$v_{est_{wake_i}} = \sum_{i=1} v_{real_i} \times weights_{wake_{norm_i}} \quad (27)$$

$$\varphi_{est_{wake_i}} = \sum_{i=1} \varphi_{real_i} \times weights_{wake_{norm_i}} \quad (28)$$

For this model, equation (25) represents the two different ways weights can be calculated with this model, depending if a specific pair of turbines can affect each other for specific values of the absolute wind direction of the park. The first equation represents the bell-shaped equation and the x represents the wind variables measurements, C determines the center of the bell-shape, and for this case, it will be the values of absolute wind direction; h determines the height of the curve and a and b represent two parameter functions: a represents the half-width and b (together with a) controls the slopes of the crossover points. The variables $v_{est_{wake}}$ and $\varphi_{est_{wake}}$ represent the estimated values of wind speed and wind direction, respectively, for the wake influenced model.

Since there were three parameters that changed the aspect and dimensions of the function (a, b and h), but it was not known how they would impact the results, it was necessary to proceed with extensive research on what types of values these parameters could range. In chapter 5.1.4, these values are shown and the results of the model discussed. After calculating the weights for all turbines through equations (25) and (18) (this equation is for the remaining turbines in the park that are not affecting each other), these weights are normalized (equation (26)) and applied to the measurements of wind speed (equation (20)) and absolute wind direction (equation (21)). A theoretical representation of the bell-shaped function is represented in Figure 18. [21]

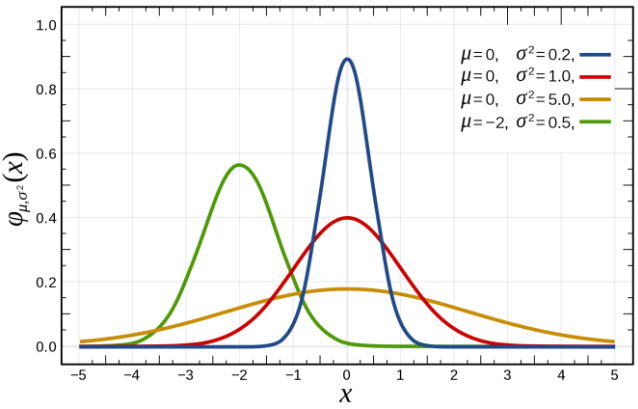


Figure 18 – Examples of a theoretical bell-shaped function

The weights are then multiplied by the measured values of wind speed and absolute wind direction of every turbine to obtain the estimated values for this model through equations (20) and (21).

It was decided to proceed with a correlation analysis between some of the initial sensor signals and the values of error obtained with this model. With this analysis, the goal was to understand if any signal had a higher correlation with the error obtained for this model. That would mean that that signal was negatively affecting the model, causing an increase in the error and the worst performance. With this information, it was possible to change the model or add a few parameters that would not allow for the signal to impact the model.

However, no valuable information was found to proceed with this hypothesis when concluding this analysis. The signals with the highest correlation with the error were signals dependent on the variable chosen, which was wind speed. These signals were, for example, wind speed estimator, rotor and generator torque speed, grid power, and bending moments from the blades. For this analysis, it was necessary the information of signals independent of wind speed and that could also be helpful, such as temperature and air density. However, none of these signals proved to have a high correlation with the error. In Figure 19, it is possible to see a representation of the highest to the lowest correlation values. One concluded that the signals with the highest were somewhat dependent to the values of wind speed.

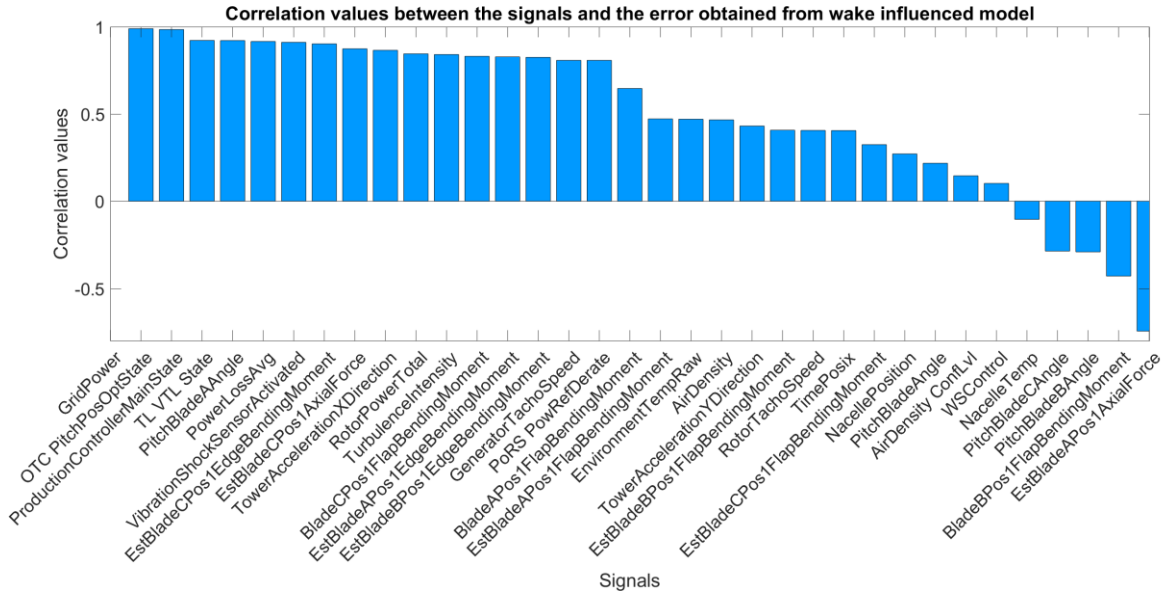


Figure 19 – Graphic representation of the correlation values between sensor signals and error obtained from wake influenced model

There is a clear transition in the graph from high to low correlation values, and these lower correlation values corresponded to the independent variables that were sought.

4.1.5 Mutual information weighted model

The fourth model developed was a mutual information weighted model, and it is an extension of what was developed in the wake-influenced model. However, different weights are applied. For this model, the same approach was considered related to the angles between turbines. So, in cases of turbines that could be causing a wake on the turbine estimated, the weight applied to their information will be calculated in the same way: through the bell-shaped function (equation (29)).

$$weights_{mi} = \frac{h}{1 + \left|\frac{x - C}{a}\right|^{2b}} \tag{29}$$

The weights applied to the information of the remaining turbines are calculated through the values obtained from the feature selection analysis (Chapter 3.3.1) and are represented in equation (30) and (31).

(30)

$$weights_{mi} = \frac{weights_{feature\ selection}}{\sum_{i=1} weights_{feature\ selection}}$$

$$weights_{mi_{norm}} = \frac{weights_{mi}}{\sum_{i=1} weights_{mi}} \quad (31)$$

These equations (29) and (30), represent the two ways weights can be calculated in this model. It has the same line of thought as the wake influenced model, except that for this model, the weights applied to the information of turbines who are not being influenced by any other turbine in the park, are obtained from the feature selection analysis (equation (30)).

This analysis made it possible to retrieve the values of the relation between every turbine through the mutual information function, regardless of the distances between them. As it was possible to conclude with the correlation analysis between wind speed values and distance between turbines, the difference between the correlation values of closer turbines and turbines further away is not as substantial as expected. Because of this, the weights retrieved from the mutual selection function are more accurate and will, hopefully, correspond to better results for the estimated values. These weights are represented in Table 1 for turbine 1, but they were retrieved for every one of the turbines.

$$v_{est_{mi_i}} = \sum_{i=1} v_{real_i} \times weights_{mi_{norm_i}} \quad (32)$$

$$\varphi_{est_{mi_i}} = \sum_{i=1} \varphi_{real_i} \times weights_{mi_{norm_i}} \quad (33)$$

The weights are then normalized (equation (31)) and it is possible to obtain the estimated values of wind speed ($v_{est_{mi}}$) and wind direction ($\varphi_{est_{mi}}$) for the mutual information weighted model, represented in equations (32) and (33).

4.2. Particle model

After developing these last five models, it was noticed that their results were converging, and the models were only incrementing in performance, but there was nothing entirely new to it. Because of this, one decided to test a completely new approach found when researching a few articles and papers related to this line of work (wind flow reconstruction), for literature review. This new wind model was found in [12], and it showed promising results in the paper. However, it was decided to create a version based on this model for this dissertation, with slight differences in parameters and a few improvements.

The main objective of this model was to create particles of wind that originate at every turbine in the wind park, with a specific value of velocity in the *x* and *y*-axis, a value of latitude and longitude, and age. These particles are created when they originate at every turbine, and in every second (the data frequency used for this model is 1 Hz which translates to a period of 1 second), the particles are updated. New particles are formed with the same origin (the position of every turbine). In Figure 20, an example retrieved from the article [12], a single turbine is represented in red. It is possible see that the blue particles are originated at the turbine. After that, they will be updated and gain a different velocity, position, and age since their age increases linearly with every second.

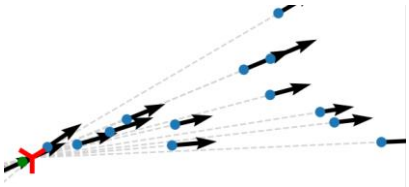


Figure 20 – Representation of a particle leaving a single turbine

As it is possible to understand, the particles only move on a 2D surface, and a value in the *z*-axis is not considered.

4.2.1 Turbine's area of influence

After having this concept aligned, it was essential to understand how it would be possible to retrieve the wind speed and wind direction of every turbine with this model. For this, a simulation of 1 hour was tested, where there were particles being created and others being updated for 3600 seconds, and a condition was created for every turbine: if there were particles in the area defined by equation (34), that represents an area of a circle around every turbine, the value of velocity in *x* and *y*-axis of those turbines would be used and those quantities transformed from the Cartesian representation to the polar

representation, attaining values of wind speed and absolute wind direction. The information of the particles surrounding that turbine was used, but it was necessary to exclude the information of the own turbine.

$$(L - x)^2 + (\theta - y)^2 = r^2 \tag{34}$$

In this equation, L represents the values of longitude of the particles analysed, x represents the coordinate in x axis that characterizes the center of the circle, θ represents the values of latitude of the particle, y represents the coordinate in y axis that, alongside with x, represent the coordinates of the center of the circle and r is the radius of the circle. After an adequate number of particles proves to be inside the area defined for a turbine, it is possible to estimate the values of wind speed and absolute wind direction in a turbine by transforming the cartesian values to a polar representation, represented in equations (35) and (36).

$$v_{est_p} = \|(u_x, u_y)\| = \sqrt{{u_x}^2 + {u_y}^2} \tag{35}$$

$$\varphi_{est_p} = atan2(u_x, u_y) \tag{36}$$

Where u_x and u_y represent the velocity components of the particle in the cartesian representation and v_{est_p} and φ_{est_p} represent the values of wind speed and absolute wind direction in the polar representation, respectively.

4.2.2 Inclusion of fluid dynamics

After having a basis of the particle model and testing its results (chapter 5), it was decided to test this model considering that the movement described by every wind particle is not independent and that, eventually, the particles are going to lose or gain energy which will translate to velocity. This way, it was no longer possible to assume that the fluid is incompressible and inviscid, which means that now, it was assumed that the wind has resistance to shear stress, and the fluid can be considered viscous. Viscosity is a measure of a fluid's resistance, in this case, air, to deform under shear stress and is commonly perceived as flow behaviour. Viscosity describes a fluid's internal resistance to flow. It may be thought of as a measure of fluid friction, and it also makes possible the acceleration of one relative to the other, and this was the necessary point for this model. [22]

For gases, viscosity is highly affected by the collisions that happen between the molecules. An increase in temperature will increase these collisions, leading to a higher coefficient of viscosity. For the particle model, it was desirable to consider the case where particles interact in this sense. If two particles collide, one of them will remove part of the other's energy and increase its speed, and, consequently, that particle will have a lower speed by the end of this interaction. The coefficient that will translate how much energy is exchanged between particles that interact with each other could be provided by the kinematic viscosity since these are particles of wind with a specific velocity, and kinematic viscosity is a measure of velocity. In contrast, dynamic viscosity is a measure of force. In equation (37), it is represented the calculus of the kinematic viscosity (ν), which is a product of the dynamic viscosity (η) divided by the density (ρ) of the fluid.

$$\nu = \frac{\eta}{\rho} \tag{37}$$

The challenge with using this equation is that it was not possible to obtain any information regarding the values of pressure of the fluid. Thus, standing in a difficult position since that parameter is necessary to proceed with this calculation. Even trying to approach the problem from a different perspective by calculating the viscosity coefficient (equation (38)). [22]

$$\delta = \frac{F \times r}{A \times v_f} \tag{38}$$

In this equation, δ represents the coefficient of viscosity, F represents the tangential force of the fluid, r represents the distance between the layers, A represents the area of fluid considered and v_f represents the velocity of the fluid in that area. Since the analysis is based on wind passing through a park, changing velocity and direction at every second, trying to provide values to the area and the velocity gradient of the fluid would be an extensive and, with lack of information, difficult task. Because of this, it was necessary to approach the problem in a different way and consider that when the particles interact, a specific value of energy is exchanged between them, where the velocity of one of the particles will increase. In contrast, the other will decrease (equilibrium of a system).

4.2.3 Particle collision

To understand the number of particles interacting, area similar to the one calculated in equation (34) was defined. However, instead of analysing the number of particles near the turbine, now the analysis is focused on the number of particles near one another. With that information, it is possible to know the particles that will interact with each other and calculate the values of wind speed and direction after their interaction. After knowing the number of particles inside the area defined for each particle, it is possible calculate the velocity values in the *x* and *y*-axis after their interaction (equations (39) and (40)).

$$u_{x_f} = u_{x_i} - \left(\frac{\sum_{j=1} u_{x_i}}{u_{x_j}}\right) \times E \tag{39}$$

$$u_{y_f} = u_{y_i} - \left(\frac{\sum_{j=1} u_{y_i}}{u_{y_j}}\right) \times E \tag{40}$$

Where u_{x_f} and u_{y_f} represent the final values of velocity in *x* and *y*, respectively, u_{x_i} and u_{y_i} represent the initial values of velocity before the interaction with other particles, u_{x_j} and u_{y_f} represent the values of the velocity of the particles close to the particle that is estimated, and E represents the percentage of energy removed from the particles with a higher velocity.

4.3. Machine learning

4.3.1 Machine learning algorithms

The last model developed was a machine learning model. Machine learning is a branch of artificial intelligence (AI) and computer science that focuses on using data and algorithms to imitate the way humans learn, steadily improving its accuracy. [23] The basic concept of machine learning in data science involves using statistical learning and optimization methods that let computers analyse datasets and identify patterns. Machine learning techniques use data mining to identify historical trends. The typical supervised machine learning algorithm consists of roughly three components:

1. **A decision process:** A recipe of calculations or other steps that takes in the data and returns a supposition of the type of pattern in the data the algorithm is looking to find.
2. **An error function**: a method of measuring how good the supposition was by comparing it to known examples, when these are available.
3. **An updating or optimization process**: the algorithm looks at the data it missed and then updates how the decision process comes to the final decision so that the missed parts will not be as plentiful next time. [24]

The algorithm will proceed with these evaluations during the training process. This process is the period where the algorithm will go through all the steps mentioned above with the training dataset to discover the best fit or model that guarantees a better accuracy to the forecasting results. After this, the chosen model is tested for a different dataset to understand if the accuracy maintains, and this data is called the test data set. For this specific case, our train and test datasets were divided into different periods. For example, the train and test dataset corresponded to the same wind speed measurements but simply for different hours in time when testing the variable wind speed.

4.3.2 Gaussian process regression model

An initial input was required to start training the model to discover the best fit to the dataset provided: the values of wind speed of the turbine that was going to be estimated. After this, the software trains the models specified and finds the one that has the lower root mean square error (equation (15)), which will be the most accurate model for the given dataset. For the existing training dataset, the best model was a Gaussian process regression model for both wind speed and absolute wind direction variables.

Regression models based on Gaussian processes are simple to implement, flexible, fully probabilistic models, and thus a powerful tool in many application areas. [25] A Gaussian process regression (GPR) model is a probabilistic supervised machine learning framework that has been widely used for regression and classification tasks.

This model can make predictions incorporating prior knowledge and provide uncertainty measurements over predictions. [26] In Figure 21, it is possible to illustrate a Gaussian process regression for a target test function in one dimension. The training data corresponds to the red dots, and they are subject to a Gaussian observation noise with a standard deviation of 0.03. The blue line shows the Gaussian process regression's mean prediction and the corresponding uncertainty's shaded region.

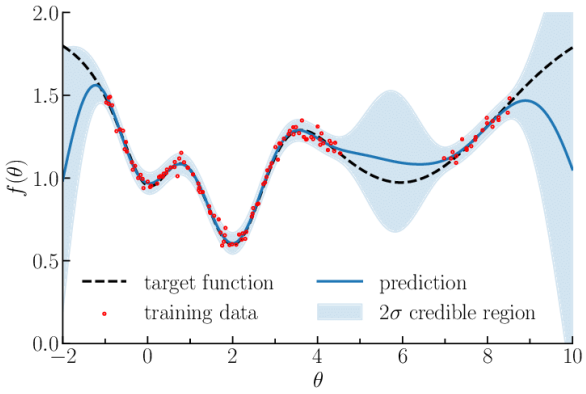


Figure 21 – Visual representation of a Gaussian process regression model

5. RESULTS

In this chapter, one will evaluate the quality of every model's prediction present in chapter 4 from the turbine measurements. For every model, the results were tested in every turbine and it was possible to understand probable causes for the differences between the results.

As for the flow reconstruction at the location of a specific turbine, the measurements of that same turbine are excluded and only the measurements of the remaining 19 turbines are used. Finally, the estimated values of wind speed and wind direction are compared with the measurements of the analysed turbine. For this chapter, it will only be shown a graphic representation of a time-series between the estimated values and real measurements for turbine 11, a turbine located in the center of the wind park. However, the time-series for the remaining turbines are presented in Annex D.

5.1. Weighted models

5.1.1 Regression model

Figure 22 shows a two-month time-series for turbine 11, between the real wind speed values and the estimated values obtained from the regression model.

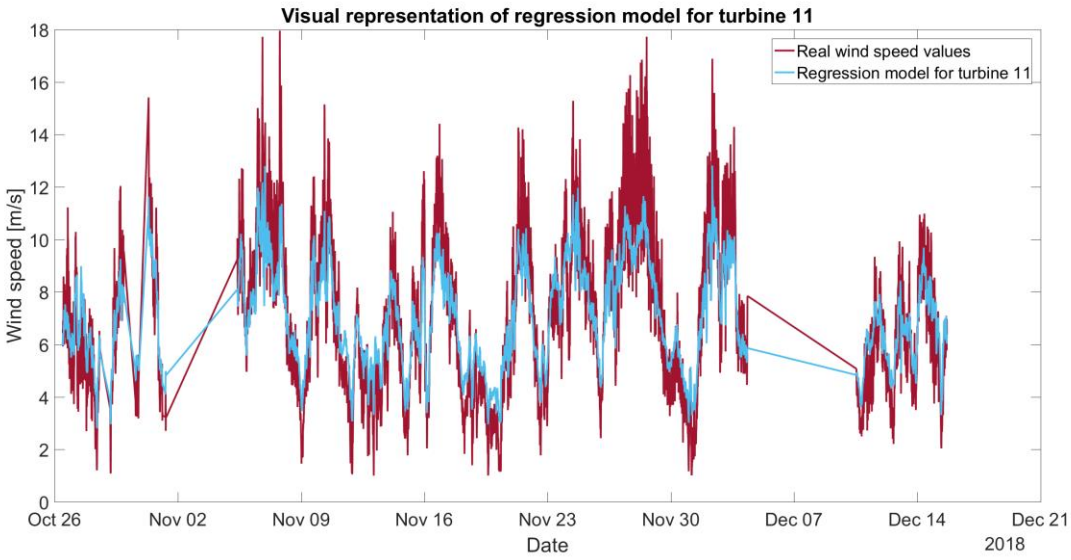


Figure 22 – Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for regression model

As it is possible to see in the graph, the regression model tends to follow an average of the real wind speed values. It fails to predict some of the significant variations in wind speed, for example in days 7th and 8th of November, where wind speed reaches values of approximately 18 m/s (values that do not

correspond to typical values of wind speed in this park), and 29th of November, where the model is not able to predict any values from 10 to 18 m/s, again, values considered as a higher dynamic of wind for the park. Since it is being considered periods where the turbines change from idle to production time, the same variation happens with lower values such as 5 to 0 m/s. For those values, the model cannot accurately predict the lower values of wind speed. It is possible to conclude that the model is underestimating the real measurements.

This model was tested for the two wind variables mentioned before in this dissertation, wind speed and absolute wind direction. For absolute wind direction, the results are presented in Figure 23.

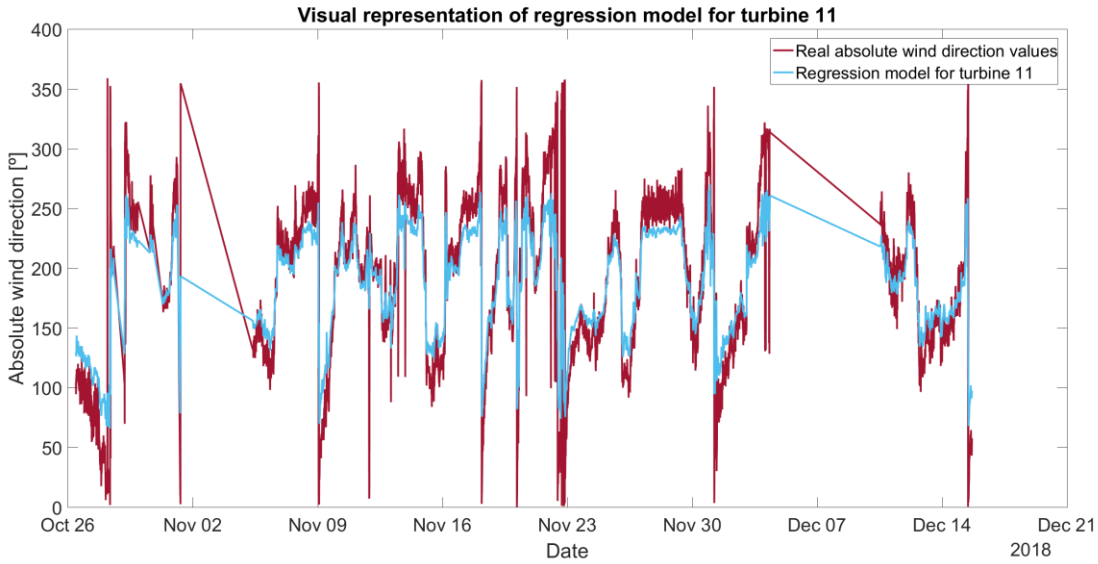


Figure 23 - Visual representation of a time-series for turbine 11 between real absolute wind direction values and estimated values for regression model

Visually, it is possible to understand that this model has a similar outcome for wind direction since it does not follow the variations accurately that this variable has throughout the days. However, the model can barely estimate the mean of absolute wind direction in some regions, for example, between the 23rd of November and the 30th of November. The estimated values are consistently higher or lower than the corresponding measurements.

As mentioned in chapter 4, after having the results for every turbine, the next step was to calculate the error associated with the estimated values of the model through equation (15). With this, the overall root mean square error (RMSE) and standard deviation (SD) was also calculated for the whole park by calculating a mean of the errors of every turbine, and the results are:

- RMSE: 1.65 m/s for wind speed and 37.87° for absolute wind direction

- SD: 0.18 m/s

Analysing this error, one could conclude that having an error below 2 m/s is a good result for a more straightforward approach since it is very close to the results obtained when putting a wind mast in front of a turbine and measuring their values of wind speed and direction. A wind met mast is a meteorological tower that can measure various wind characteristics such as wind speed and wind direction on a specific wind park. [27] For absolute wind direction, the error was calculated differently since it was not considered the entire interval of absolute wind direction values that range from 0 to 360°. This model, while estimating the values of absolute wind direction, is not able to capture such opposite values as 0 and 360°, and there is a jump that happens when the wind direction goes from values around 0° to measure 360° suddenly, and models used to reconstruct a wind flow cannot capture. Because of this, and to minimize our error, it was decided to analyse only the values inside a specific interval, choosing values that one knew the model could estimate. A correct approach would be to use an initial absolute wind direction interval of $-180° \leq \varphi \leq 180°$, in order to skip the jump in values when using an interval of absolute wind direction of $0° \leq \varphi \leq 360°$. However, since this was noticed after, it was decided to calculate the error for a specific interval on each model. Therefore, for this model, the error was calculated for an interval of $70° \leq \varphi \leq 270°$, which was the only range of values that this model started to estimate. This is the only model that uses short and close intervals since it produced the worst results.

5.1.2 Distance weighted model

For the second model developed, distance-weighted model presented in the previous chapter, there was a distinct improvement in the two variables' results. Starting with wind speed, it is possible to see in the time-series of Figure 24 that visually, this model tends to follow an average of the measurements and the last. However, it can capture the variations of wind speed and some of the spikes to higher values such as 14 to 18 m/s, as to lower values of wind speed, located in the region I of a wind turbine's power curve (Figure 5), more accurately than the regression model.

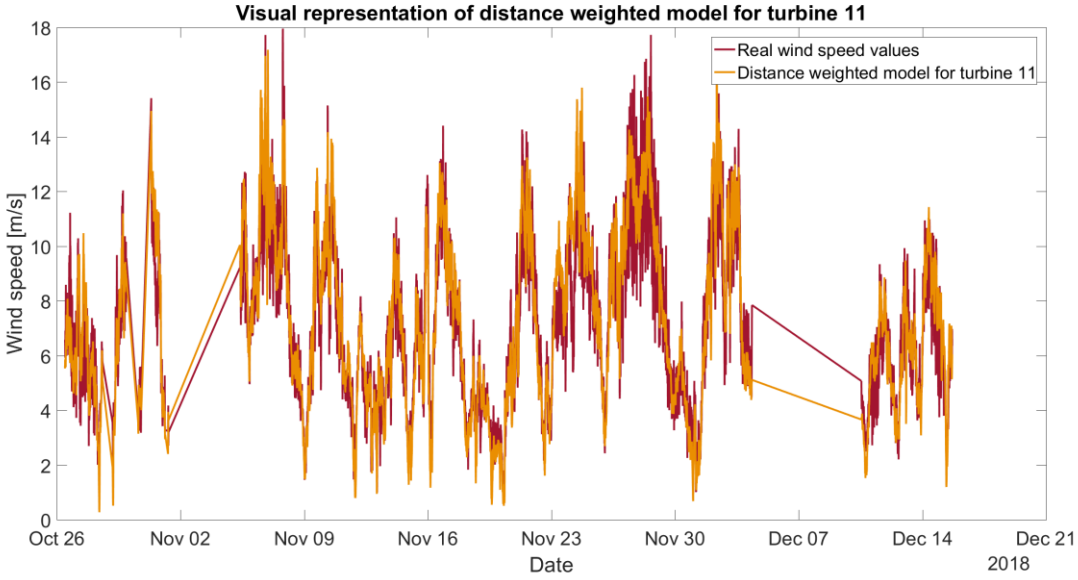


Figure 24 - Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for distance weighted model

For absolute wind direction, the results are presented in Figure 25.

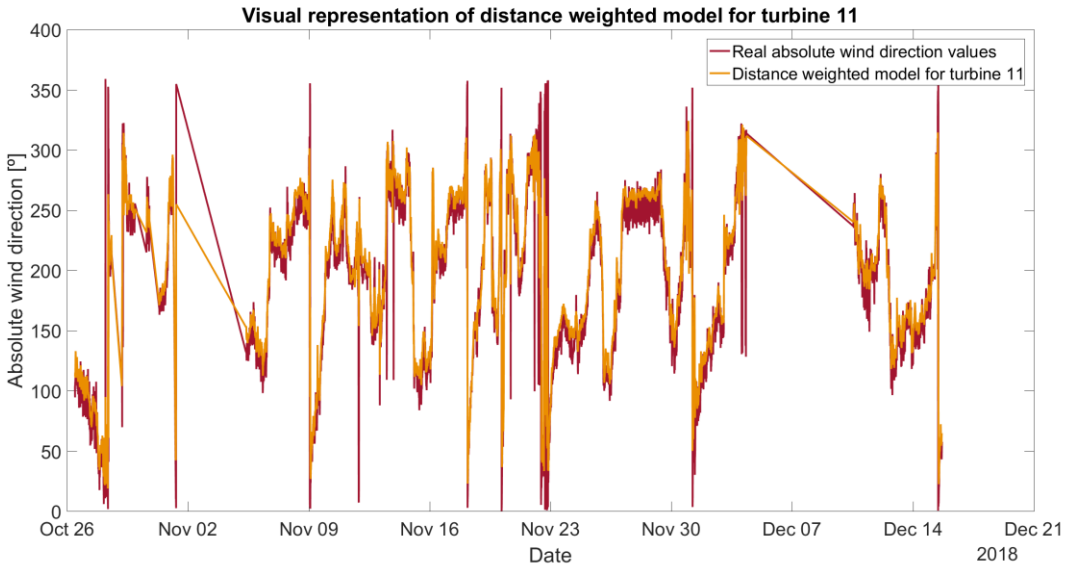


Figure 25 - Visual representation of a time-series for turbine 11 between real absolute wind direction values and estimated values for distance weighted model

Visually, it is possible to conclude that there are also better results for the time series of absolute wind direction for this model. It has a better performance and can accurately capture some of the spikes and variations in wind direction. It is also essential to understand that wind direction in a park will range from 0 and 360° when considering this circular scale. That is a much wider interval than wind speed, which will vary approximately from 0 to 18 m/s. Because of this, the values of absolute wind direction (that already consider the changes in the angle of the nacelle, a control strategy mentioned in chapter 2.1.2) are much harder to obtain an accurate and reliable representation. This model can even capture some of the higher variation in wind direction for this turbine, for example, on days 2nd November, 9th of November, and 23rd of November.

It was also possible to conclude with the other model that every turbine has a distinct behaviour, but that will be explained more thoroughly in the next chapter. For the distance weighted model, the results are as follows:

- RMSE: 1.45 m/s for wind speed and 29.82° for absolute wind direction
- SD: 0.15 m/s

For this model, the interval of absolute wind direction used to calculate the error was $45° \leq \varphi \leq 315°$, a range that can capture more wind direction values. Using a more flexible interval was only possible since this model demonstrates better results and can estimate values within a larger interval.

Considering that this model presents promising results, it is possible to see an apparent decrease in the error for both variables. For wind speed, the error went from 1.64 m/s to 1.45 m/s, corresponding to a decrease of approximately 0.20 m/s. Since it was already established that an error in the order of 1 m/s is already a good value, one can conclude that this decrease is significant for the scale considered. As for the values of absolute wind direction, the error went from approximately 37° to 30°, and even though, for such a complex variable such as wind direction, a 7°difference is still significant, there is still an improvement between models, which was the primary purpose of trying different and improved models.

5.1.3 Correlation weighted model

For the correlation weighted model, since its approach is very similar to the distances weighted model mentioned in the previous chapter, the results for every turbine were also similar. However, there was, in fact, a decrease in the overall error of the park of the variable wind speed. For this model, since the weights used are from a correlation analysis made specifically with wind speed values (chapter 3.3.1), the model was only tested for this variable. In Figure 26, it is possible to see, for turbine 11, this model follows a mean of the wind speed values and is capable of capturing and estimating some of the higher variations that happen for values outside the most common interval of wind speed [5-11] m/s. It is the same for the distances weighted model, and visually, it is possible to conclude that these two models are very similar.

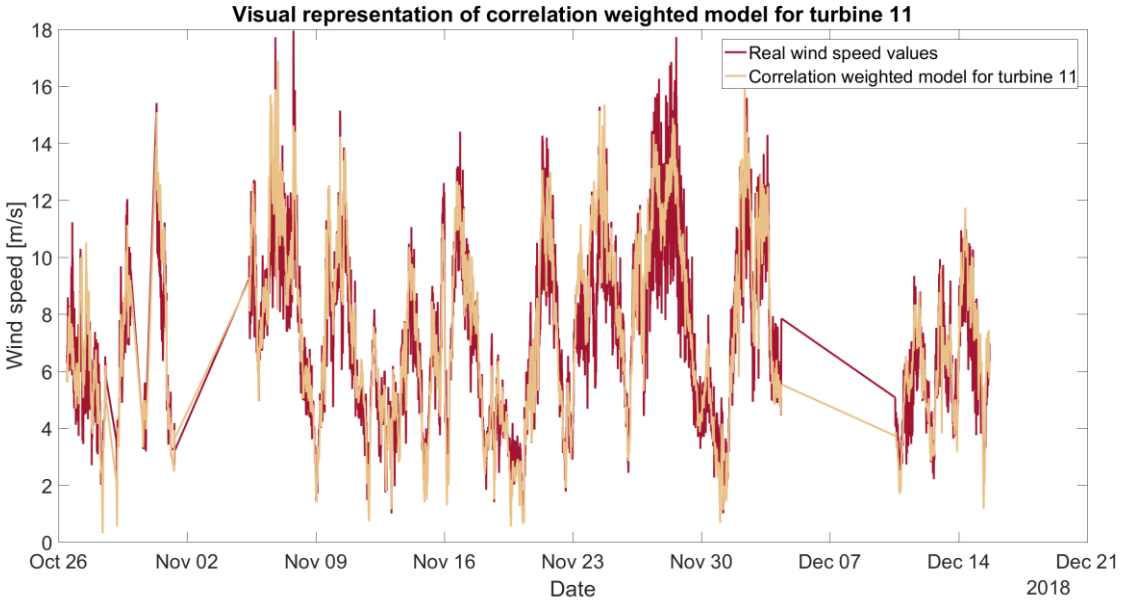


Figure 26 - Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for correlation weighted model

Despite the graphic similarities, the results of this model for wind speed are as follows:

- RMSE: 1.42 m/s for wind speed
- SD: 0.19 m/s

This error is a slight decrease from the error of the previous model (1.45 m/s). This means that trying to estimate wind speed values by attributing weights to real-time measurements on every turbine that came from correlation analysis that provides a specific value of correlation for every turbine will guarantee an improvement in the model's final results.

5.1.4 Wake influenced model

We expected some improvements compared to the distances weighted model with the wake influenced model since more information was added. However, for both variables, the improvements were not as significant as hoped. For this model, as mentioned in chapter 4.1.4, it was tested an interval of values for parameters a, b and h, to see what were the values that guaranteed better results. For a one tested value between 0.5 and 20, to ensure there was a wide variety of widths for the function and used 0.5 as the final value since it showed better results, for b one tested value between 0.5 and 4, since it was a parameter that affected more the slope of the function. A closer interval was desired, the final value was 4, and for h, one tested value between 2 and 10, where the final value was 2. Another test was checking if the model performed better considering more or less weight to the specific pairs of turbines (considering a spike or a depression in the bell-shaped curve). The best results were giving more weight to the information coming from these turbines.

For the variable wind speed, the time-series representing turbine 11 (Figure 27) is very similar to the distance weighted model's time-series, especially when compared together, as will be shown further in this dissertation.

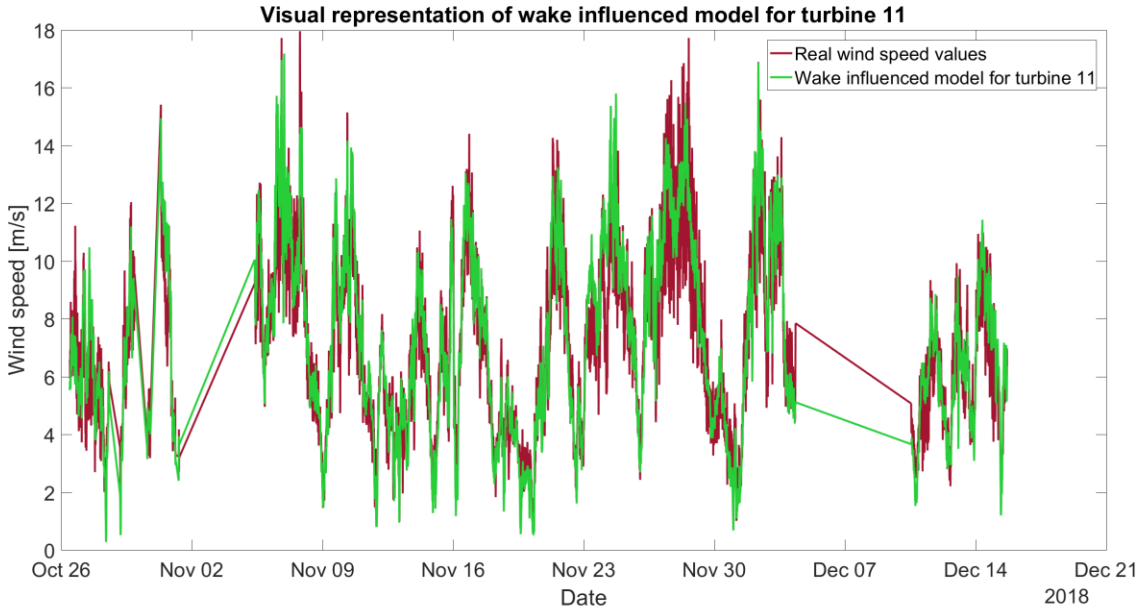


Figure 27 - Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for wake influenced model

For absolute wind direction, the behaviour of the time-series is also similar to the distances weighted model (Figure 28), as it follows the same patterns and the model mentioned throughout the entire dataset.

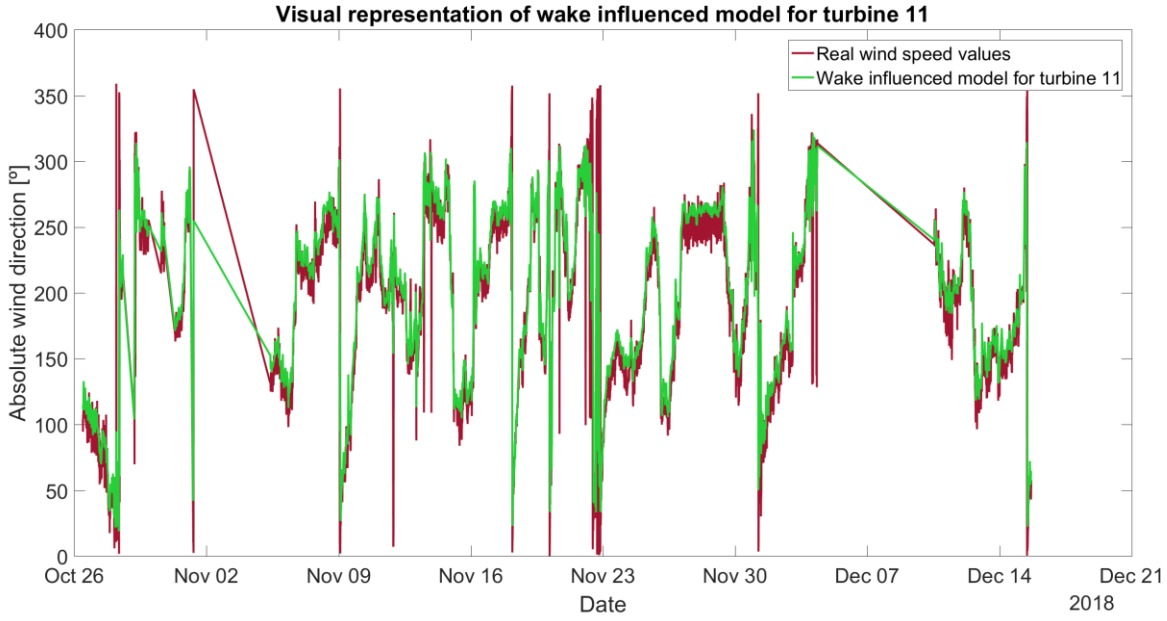


Figure 28 - Visual representation of a time-series for turbine 11 between real absolute wind direction values and estimated values for wake influenced model

For this model, the results are as follows:

- RMSE: 1.45 m/s for wind speed and 29.83° for absolute wind direction
- SD: 0.15 m/s

Compared to the last model, there is only a difference of 0.0002 m/s, which is an insignificant difference and that is the reason it was not shown. For the values of absolute wind direction, the error was calculated for an interval of 45 to 315°, as mentioned before. For this variable and interval of values for the error, there is a very slight error increase. However, since this dissertation is dealing with high variations between measurements for wind direction values, a difference of 0.0126 is minimal. It does not imply that this model has the worst results for this variable compared to the distances weighted model.

5.1.5 Mutual information weighted model

For the fifth model developed, the mutual information weighted model, the results were optimal compared to the other models. Starting with wind speed, the results for this variable are:

- RMSE: 1.30 m/s
- SD: 0.13 m/s

There is a decrease of 0.16 m/s compared to the last model, which is a considerably lower value and improves the model results. Visually, the time-series of wind speed for turbine 11 (Figure 29) shows a very similar pattern to models such as the distances weighted model and wake influenced model, but this would be expected since it is harder to capture differences of 0.10 m/s in a graph.

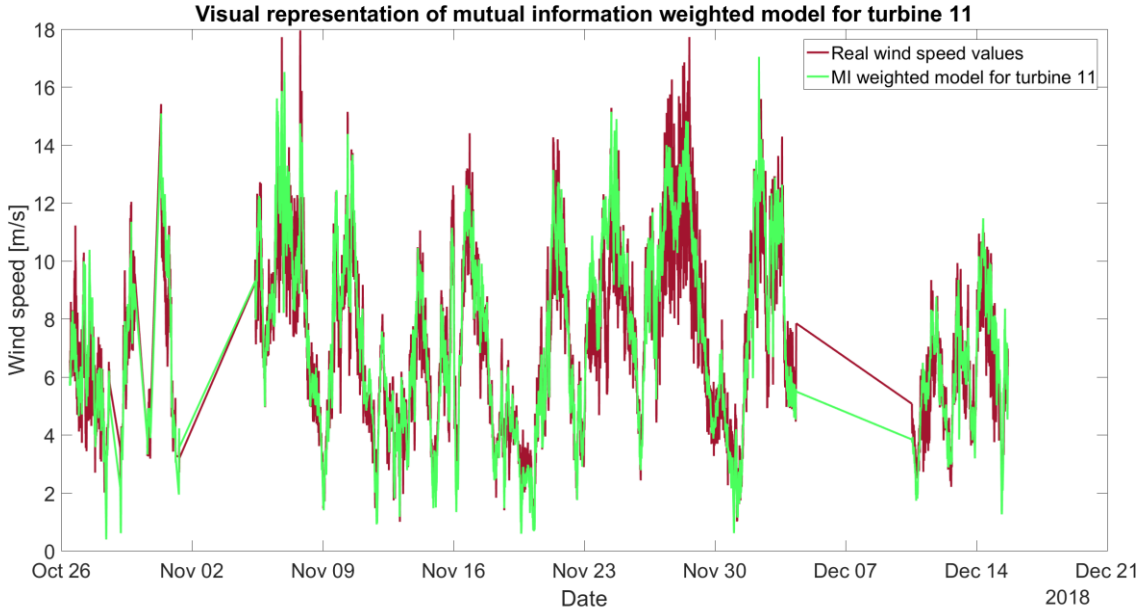


Figure 29 - Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for mutual information weighted model

For the values of absolute wind direction, it was noticed a decrease in the error:

- RMSE: 27.29°

This value differs 2.55° from the last model, which compared to the differences in absolute wind direction obtained so far, it is a very positive outcome. This error was also calculated for an interval between 45 to 315°. Visually, as it happens with wind speed, this difference is not perceptible in a time

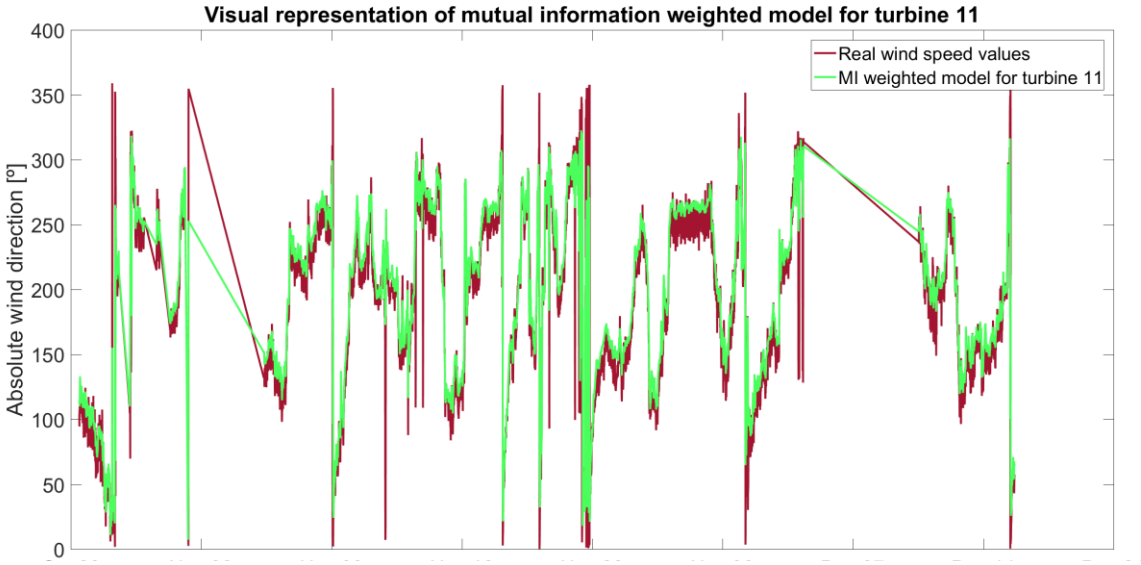


Figure 30 - Visual representation of a time-series for turbine 11 between real absolute wind direction values and estimated values for mutual information weighted model

series. However, one can conclude that for this model, the variations of the values of absolute wind direction for turbine 11 are eminently identical to the time-series of other models (Figure 30).

5.2. Particle model

5.2.1 Particles with free trajectory

For the particle model, the results are divided between two simulation times. The first simulation was done for 1 hour. Since this is a shorter period, it was possible to test different parameters and specific parts of the code since, with 1-hour simulation, one had less computational cost. The other simulation period tested was two months to have the same amount of data as the other models and compare results. However, since working with data with 1 second period, a simulation of 2 months corresponds to 4 406 400 seconds, leading to a high computational cost. Because of this, this simulation period was only tested after finding the optimal values of every parameter tested.

The first results are obtained considering that the wind particles are not interacting with each other, which means it was considered a fluid with no viscosity. In this fluid, the wind particles will have a one-dimensional trajectory, increasing velocity since no energy will be removed.

Since dealing with the condition that discloses that if no particles surround a turbine, it is not possible to calculate the values of wind speed and wind direction that the turbine would be measuring. With 1 hour simulation, some turbines had no particles near them, which meant that the program could not calculate any values regarding that specific turbine, which translated to a NaN (not a number) on Matlab. Another situation was turbines having no available particles for a specific period, which meant it was only possible to estimate wind speed and wind direction values for those limited samples.

For this first situation, where it was not considered viscosity in the fluid, one of the turbines with the best performance was turbine 3. In Figure 31, it is possible to see a time-series of this turbine for wind speed with a 1-hour simulation of the particle model.

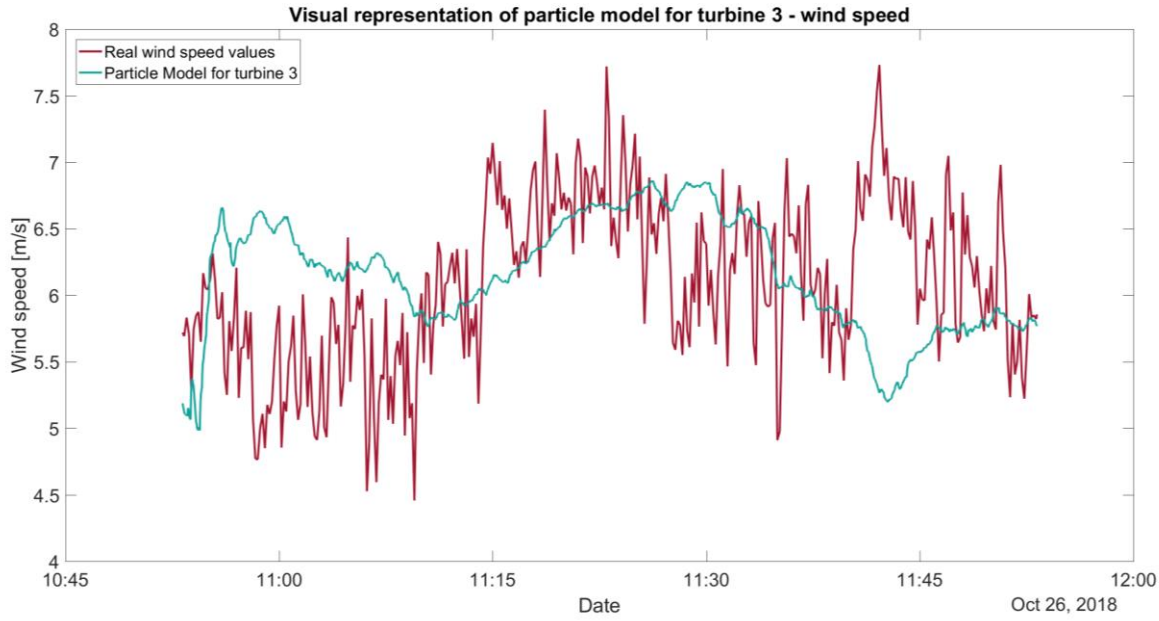


Figure 31 - Visual representation of a time-series for turbine 3 between real wind speed values and estimated values for particle model (1 hour simulation)

It is possible to conclude that the model tends to follow a mean of the measured values of wind speed. In some specific regions, such as the beginning and the end of the time-series, the model is estimating higher and lower, respectively, than the real wind speed values. The difference in values in this area is around 1 to 2 m/s. The results for wind speed are as follows:

- RMSE: 1.28 m/s
- SD: 0.64 m/s

This is a surprising low error, considering that even for turbines with better results, such as turbine 3, it is possible to see a time-series unable to accurately represent the variations of wind speed for the majority of the time. That can be explained by the fact that the model cannot estimate values in some turbines since it has no particles to calculate. Instead, it calculates only for the timestamps where there are particles, not considering the other times where NaN values exist, and that will lead to a low error in wind speed, meaning that the model is precise but not accurate. That can also be seen by the fact that the standard deviation (a measure of how dispersed the data is, compared to the mean) is relatively high for such a low error. That means that even though it is centred around 0 when analysing

the model signal (represented by the green line in Figure 31), it is possible to see that the signal follows the pattern of wind speed variation. However, it is much further away from the mean values of wind speed than the other models, which is why it has a higher deviation.

Another turbine that showed promising results was turbine 15, which time-series for wind speed is represented in Figure 32.

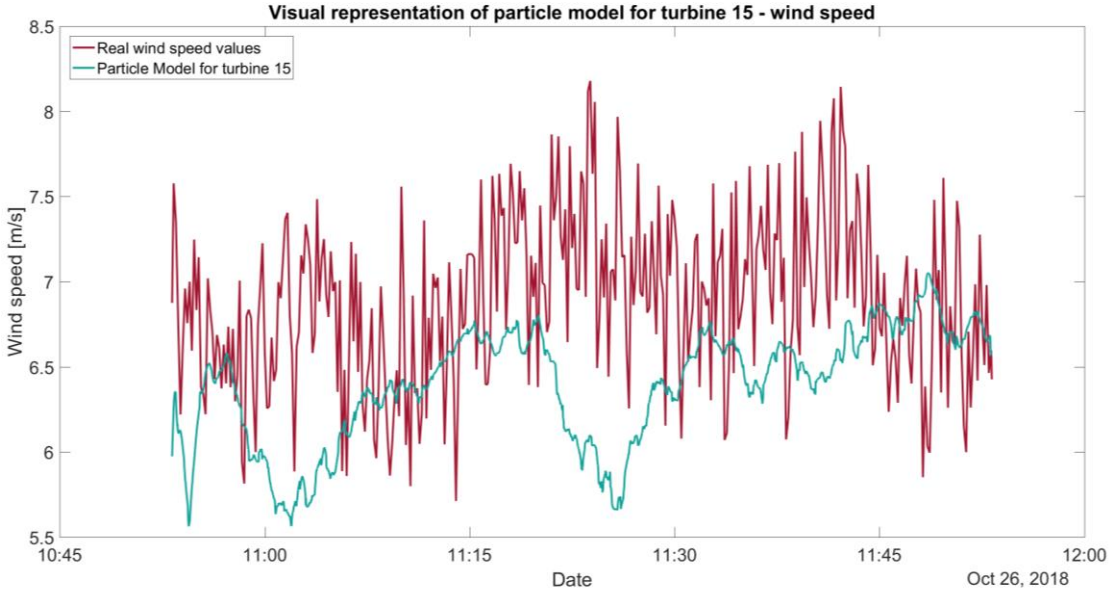


Figure 32- Visual representation of a time-series for turbine 15 between real wind speed values and estimated values for particle model (1 hour simulation)

For this turbine, even though the results are more promising than others in the park, they are not as accurate as turbine 3, as it is possible to by the graph. The estimated wind speed follows a below average pattern and is not able to predict difference of 2 m/s.

For the variable wind direction, unfortunately, the results are not as promising as wind speed. The error is similar to the other models for specific periods such as $45° \leq \varphi \leq 315°$. For that interval, the results are these:

- RMSE: 30.94°

Unfortunately, the time-series for this variable shows an estimated signal utterly divergent from the real wind speed values, with not expected results. Because it is being created a model that estimates the trajectory of a wind particle in space, it is easier to estimate more accurately its wind speed than wind direction. Because of this, wind direction had a poor estimated signal, and it was decided not to analyse this variable further in this model and conclude it needs changes for future work.

After experimenting with a simulation of one hour, one decided to test this model for a two-month-period simulation to have a basis where it would be possible to compare the results with the other models. That meant a high computational cost; however, it was essential to guarantee it would be possible to see the results of this model for periods equal to what was tested in the models before. In Figure 34, it is possible to see a representation of the first approach of this model for a simulation of twelve days. Initially, it was desired a simulation for two months, as the other models that were developed before. However, this proved to have an extremely high computational cost, and it was not known if the computer was even able to run such an extensive simulation. Because of this, one decided to experiment this model for a week (slightly longer than a week since the time considered is 12 days) and take conclusions from these results. For this simulation, in order to guarantee a lower computational cost, it was only used the data from a cluster of four turbines in the park. These turbines are turbine 9, 10, 11 and 12, they are located in the middle of the park and they were also used for an analysis on chapter 5.2.2. This cluster of turbines is represented in Figure 33.

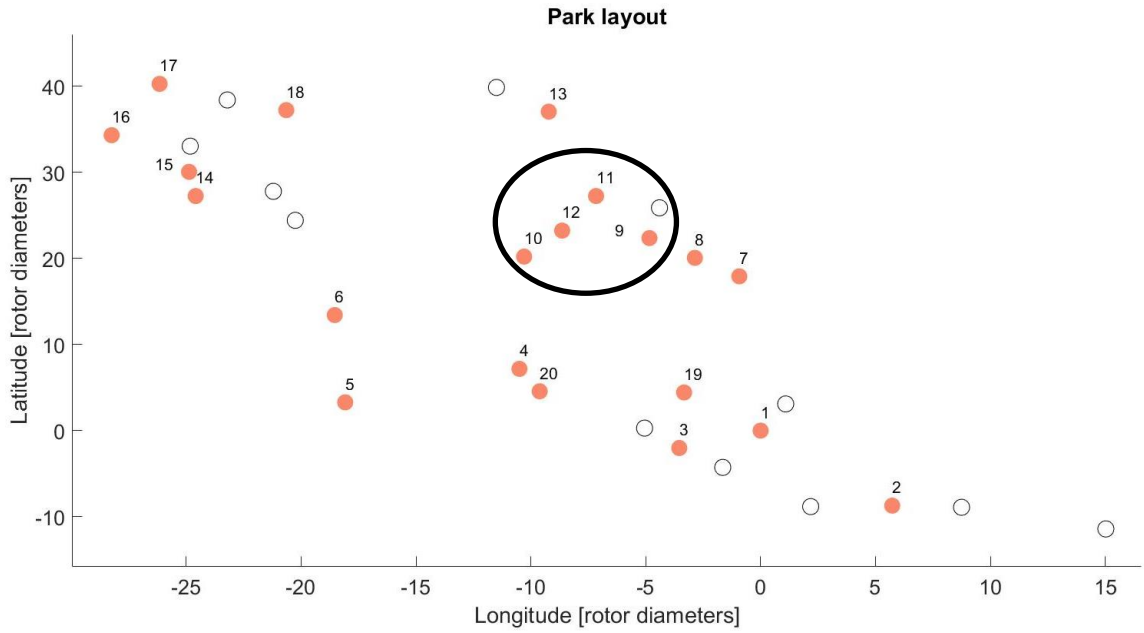


Figure 33 – Cluster of turbines in the middle of the park

For this simulation, the best results were found for turbine 11 and that is the turbine represented in Figure 34.

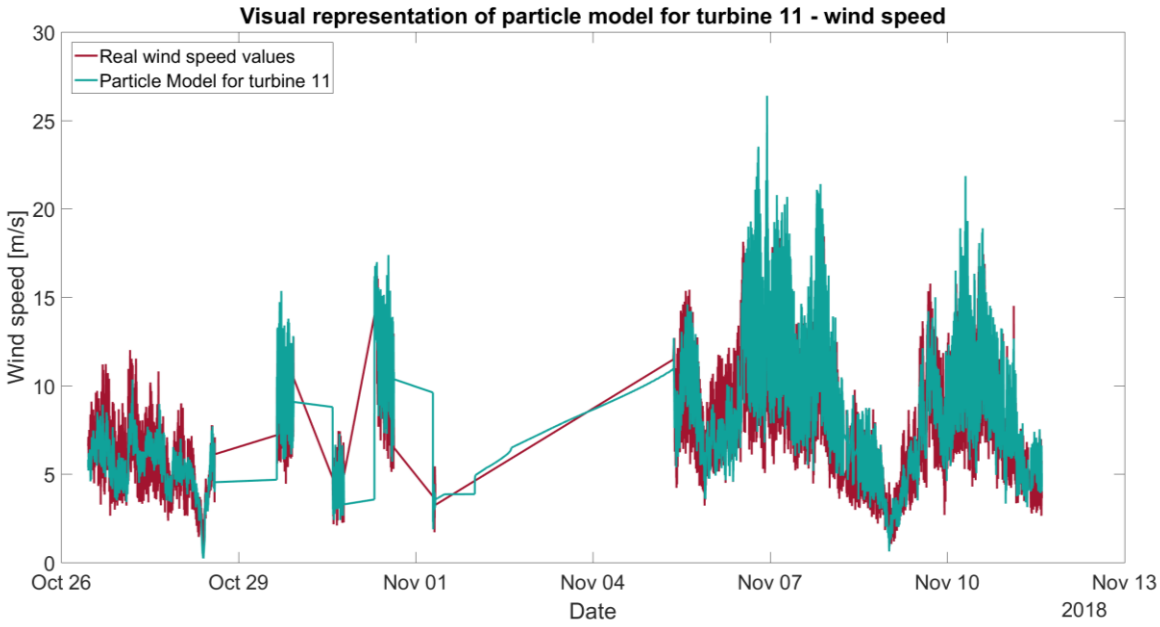


Figure 34 - Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for particle model (1 week simulation)

In this figure, it is possible to see that for a period of twelve days, the particle model is able to reconstruct the variations of wind speed in a much clear and accurate way than it does with one hour simulation. This could be explained since this a closed group of turbines, all close to one another, allowing an instant and more accurate flow of particles. Besides, these turbines are located in the middle of the park, which is an area usually known for better wind conditions, since the wind reaches this region with a more constant and less turbulent flow, with occasionally higher speeds as well.

For a twelve-days simulation, the results are as follows:

- RMSE: 1.65 m/s
- SD: 0.36 m/s

It is possible to see that even though the error has increased, compared to the one-hour simulation, the standard deviation has significantly decreased. That explains the way the estimated wind speed values of this time-series follow the variations of the velocity much more accurately than the one-hour simulation.

5.2.2 Interactions between particles considering viscosity

For the second result, viscosity was considered (equations (37) and (38)). It was decided to test different viscosity values, which translates to a percentage of energy calculated through the mean average of wind speed and wind direction of the particles analysed near the particle. Initially, the goal was to test this parameter for the whole park, considering different groups of particles that were near each other. However, to understand the nearest particles around every single particle created, for an entire park, where for 1-hour simulation there was over 40 000 particles, the computational cost and hours of the simulation were immensely high. It was not productive to spend days or weeks doing a simulation of 2 months. Because of this, one first tried to change the code by applying the percentage of viscosity for the whole park, but it was concluded this was a problem since it is being considered that all particles of the wind park affect each other. This statement is only correct if the particles are considerably near each other and not kilometres apart. To resolve this issue, one decided to test this for a specific number of turbines less than 500 meters from each other. This way a closed area of particles is being considered, as it was the intention in the first place, not the whole park. The chosen turbines were 9, 10, 11, and 12, a niche of turbines in the middle of the park. For 1 hour simulation, three different viscosity percentages were tested: 10%, 0.10%, and 0.010%. The best results were found for the percentage of 0.10:

- RMSE: 1.41 m/s
- SD: 0.48 m/s

The results with this viscosity were tested for simulations of 1 hour, and they are represented in Figure 35.

These results show that graphically, with this condition, the model tends to follow the mean of the real wind speed values slightly better. That is explained because even though the error has been increased, the standard deviation with this condition decreased, which means that this signal follows the mean of the values more precisely than the signal in Figure 31.

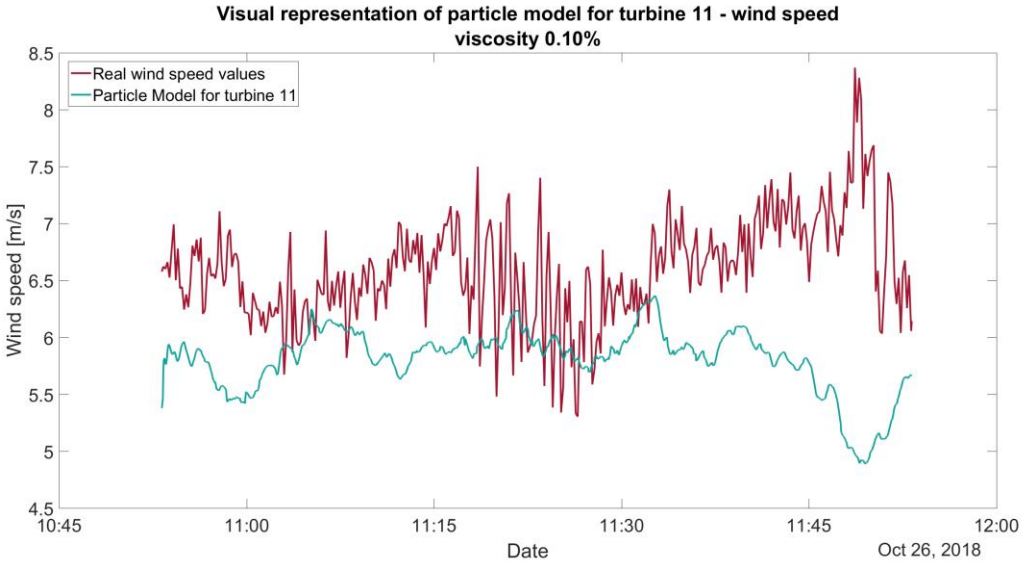


Figure 35 - Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for particle model (1 hour simulation) with a viscosity percentage of 0.10%

The graphs that represent the other values of viscosity that were tested for this simulation (10% and 0.010%), are represented in Figure 36 and Figure 37, respectively.

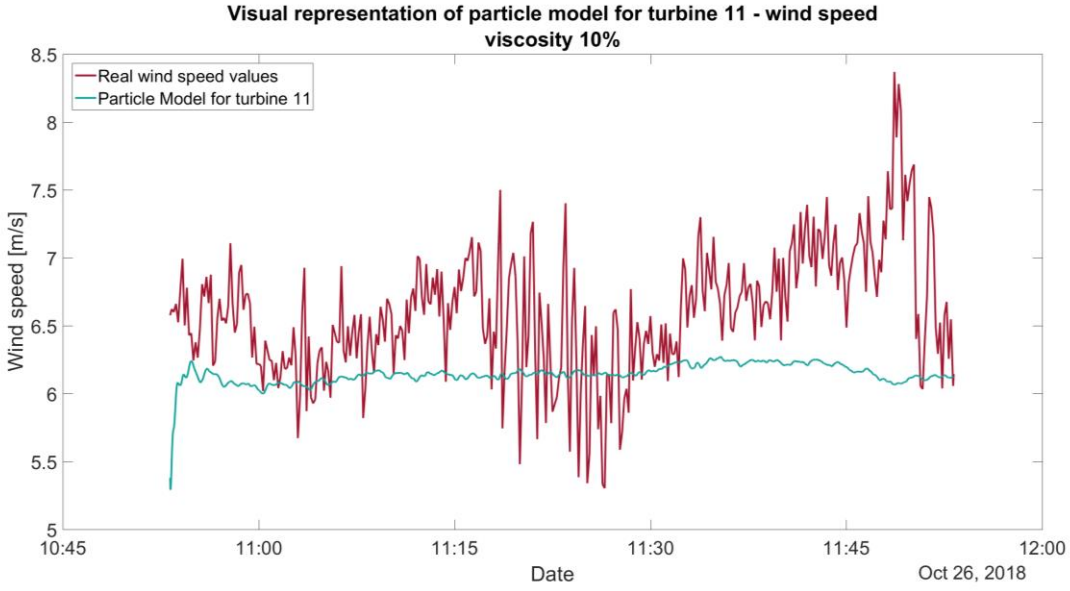


Figure 36 - Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for particle model (1 hour simulation) with a viscosity percentage of 10%

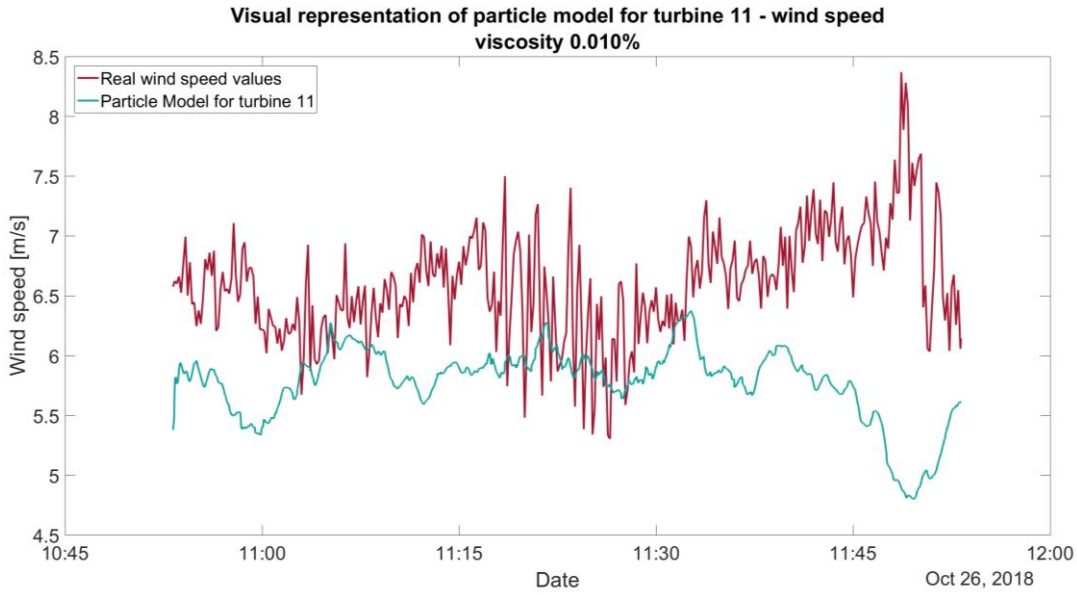


Figure 37 - Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for particle model (1 hour simulation) with a viscosity percentage of 0.010%

For the higher value of percentage tested (Figure 36), it is possible to conclude that this was a higher value than necessary. When 10% of the average of wind speed from all the particles is removed, their velocity is reduced to a point where it is practically impossible to follow any of the variations of the real wind speed from the turbine, as it is possible to observe in the figure.

For the value of 0.010% (Figure 37), since it is a much smaller percentage, when multiplied with the mean of wind speed, the final values will remain very similar to the initial values, and it is possible to see by the figure, that the variations of the estimated wind speed is very close to the results obtained from Figure 35. Because the results are so similar and this is such a small value, it was concluded that the best results were with a medium value of viscosity: 0.10%.

5.3. Machine learning model

5.3.1 Gaussian process regression model

The last model developed, the machine learning model also showed optimal results. For the values of wind speed, the following results for wind speed were obtained:

- RMSE: 1.20 m/s
- SD: 0.09 m/s

These results represent one of the lowest values of error and standard deviation for wind speed between all the models that have been developed. Besides, analysing the time-series of this model for the values of wind speed, it is possible to conclude that the estimated values of the model are very capable of capturing every variation, only failing to capture some of the higher values (18-22 m/s) in a few days in November.

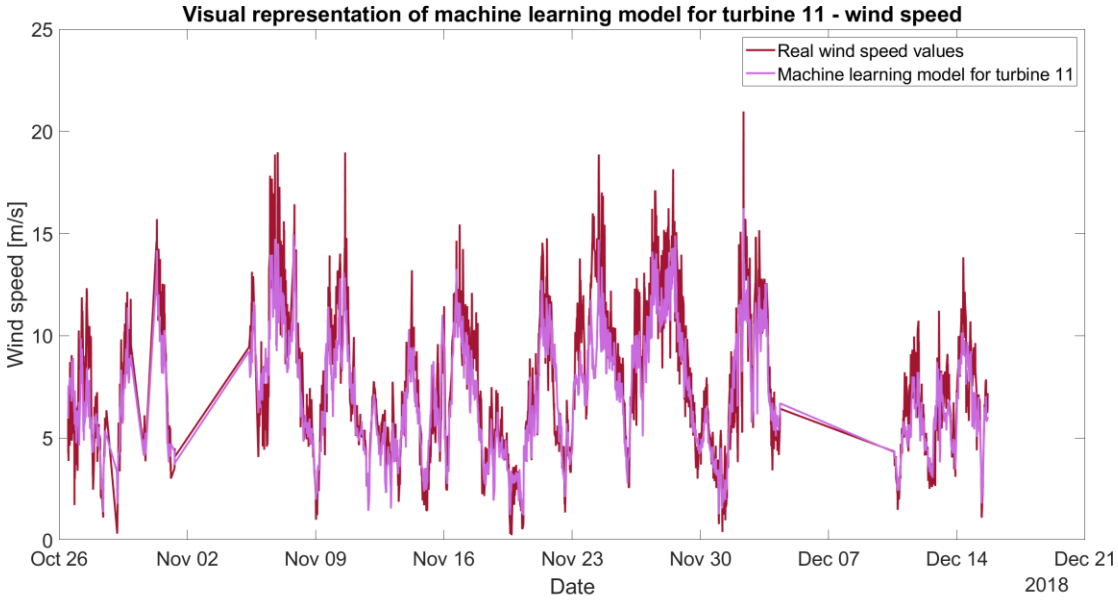


Figure 38 - Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for machine learning model

For the values of absolute wind direction, the results are as follows:

- RMSE: 24.83°

This value was calculated considering the entire interval of wind direction values, [0-360°] since this calculus is made automatically by the program where various machine learning models were tested,

the Regression learner app in Matlab. Even so, considering values such as 0° and 360°, which can cast uncertainty and inaccuracy to the estimated values, it was obtained the lowest error considering this variable with this model. Visually, the time series representing the estimated values of wind direction of turbine 11 for this model also shows that this model can estimate relatively well the endless variations and spikes of wind direction, similarly to the distances weighted and wake influenced models.

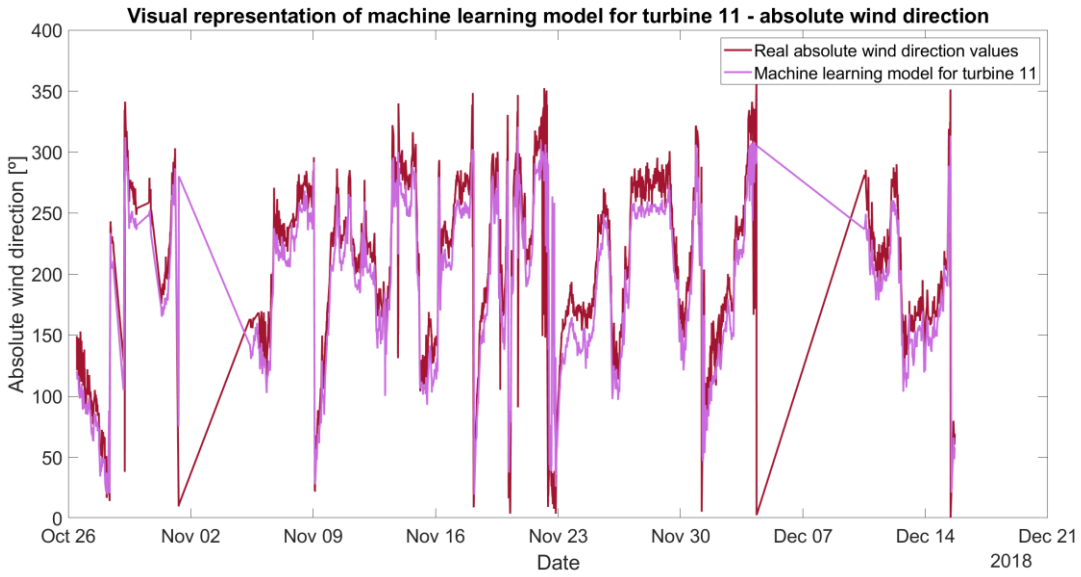


Figure 39 - Visual representation of a time-series for turbine 11 between real absolute wind direction values and estimated values for machine learning model

5.4. Comparison between models

After developing all the models, it was now essential to understand how they are connected and how the models differentiate, having models that will have a better performance than others. This analysis was made both for wind speed and absolute wind direction, and all the models were compared at the same time-series, a twelve days' time-series. Initially, the intention was to compare all the developed models for two months. However, because of the computational issue with the particle model mentioned in chapter 5.2, one decided to compare the models to a more sustainable time, which was twelve days (approximately one-week).

For wind speed, the time-series with all the models is represented in Figure 40. In the figure below (Figure 41), it is represented a zoomed version of this time-series, for a shorter amount of time, to understand with more clearance the differences between the models.

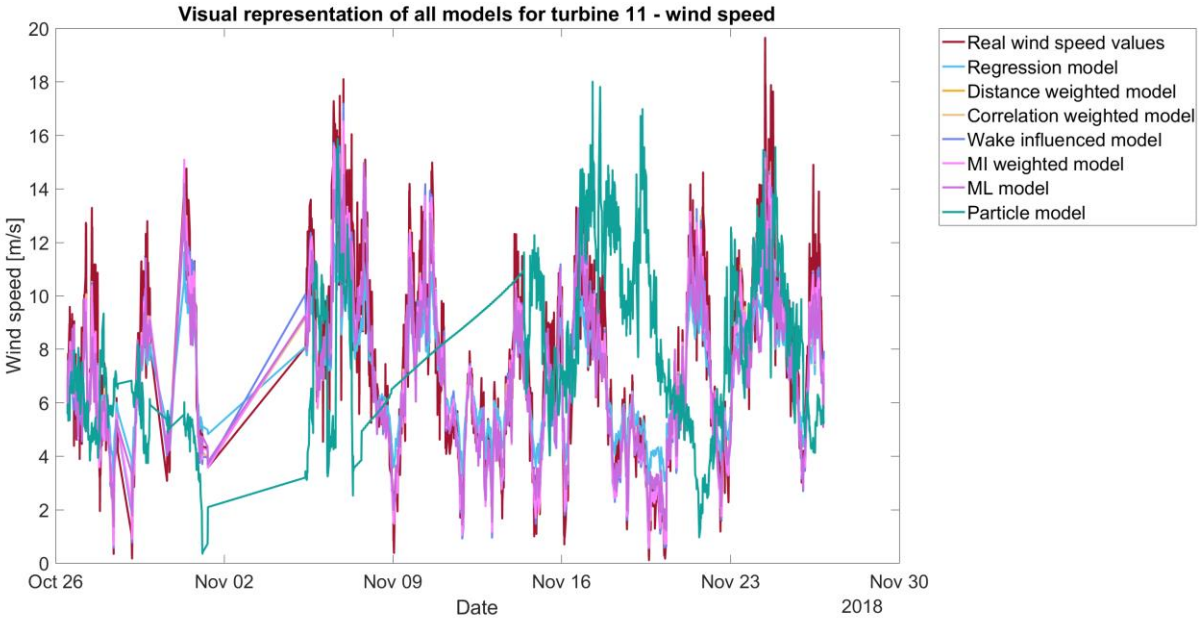


Figure 40 – Visual representation of a time-series for turbine 11 between real wind speed values and estimated values for all models developed

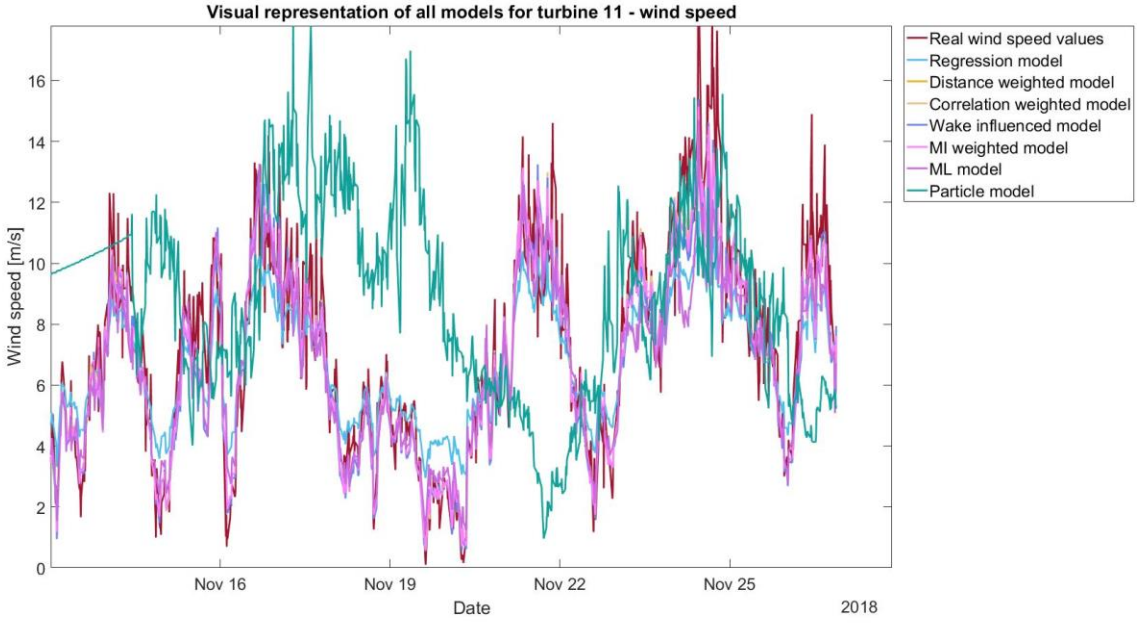


Figure 41 – Zoomed in version of the time-series for turbine 11 between real wind speed values and estimated values for all models developed

It is clear to see that the regression and particle model are the worst models in terms of performance since they are the ones who fail to achieve more significant variations in wind speed throughout the time. Even though the particle model stands out the most, this would be expected since this a model that is predicting the values of wind speed, based on the evolution of wind particles, while the regression model estimates values based on the information of other turbines, and this will necessarily lead this model closer to reality. However, there are a few periods where this model is capable of estimating with more accuracy some of the velocity variations, such as the last days of November represented in the time-series.

Even so, the remaining models can do a better job assuring that the estimated values are closer to the measured values at every turbine. There are a few regions, as it is possible to see in the graph, where none of the models can correctly calculate the real wind speed values, which can be seen by some of the higher speed winds, in the real wind speed values line. After analysing the direction from where the wind was coming at those periods of time, this lack of information in the model is not due to a specific value of wind direction, which means that no matter where the wind is coming from, there are still some values that cannot be precisely estimated. Some of the models are fully prepared and capable to reconstruct every single aspect and deviation of the wind, but there is not a 100% accuracy in any of them. The tables Table 2 and Table 3 show the difference between the errors of all models for every turbine for wind speed and the standard deviation values for all models. In Table 3, there is information about the extensive meaning of every acronym made for the models.

Table 2 – Representation of the root mean square values of all models for every turbine in the wind park (wind speed)

RMSE (m/s)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
RM	1.62	1.71	1.59	1.66	1.58	1.51	1.52	1.49	1.59	1.49	1.40	1.62	1.73	2.08	1.92	1.99	1.66	1.67	1.64	1.45
DWM	1.30	1.46	1.29	1.57	1.25	1.75	1.56	1.54	1.38	1.23	1.21	1.44	1.45	1.69	1.49	1.56	1.38	1.60	1.33	1.55
CWM	1.26	1.44	1.19	1.41	1.18	1.63	1.70	1.68	1.42	1.15	1.20	1.27	1.36	1.79	1.60	1.61	1.28	1.43	1.30	1.42
WIM	1.30	1.46	1.29	1.57	1.25	1.75	1.56	1.54	1.38	1.23	1.21	1.44	1.45	1.69	1.49	1.56	1.38	1.60	1.33	1.55
MIWM	1.18	1.29	1.19	1.31	1.07	1.46	1.38	1.44	1.33	1.14	1.11	1.30	1.24	1.52	1.43	1.42	1.17	1.30	1.17	1.44
MLM	1.16	1.22	1.15	1.10	1.13	1.16	1.19	1.20	1.20	1.15	1.03	1.18	1.28	1.39	1.23	1.37	1.25	1.31	1.11	1.21
PM									1.79	1.80	1.11	1.88								

Table 3 - Representation of the standard deviation of all models

SD (m/s)	All turbines
RM	0.18
DWM	0.15
CWM	0.19
WIM	0.15
MIWM	0.13
MLM	0.09
PM	0.36

RM	Regression model
DWM	Distance weighted model
CWM	Correlation weighted model
WIM	Wake influenced model
MIWM	Mutual information weighted model
MLM	Machine learning model
PM	Particle model

It is plausible to conclude, by analysing the error values that the mutual information weighted model and the machine learning model have a much clearer performance improvement than the other models since they have the slightest error for almost every one of the turbines. The model that has the worst results is confirmed to be the regression model and the particle model for turbines 9, 10 and 12. Analysing the standard deviation values, one can conclude that the particle model has the higher deviation. That explains the difference between the error and the graphic representation of this model: even though the error is not high, the time-series is much further from the real values of wind speed than the other models (Figure 40 Figure 41) because the standard deviation between the estimated values and the mean of wind speed is much higher, leading to a time-series further from the mean values. The machine learning model still presents the best results, even for standard deviation, and this is possible to see in the way the time-series of this model follows the real wind speed values with clear precision.

For absolute wind direction, the time-series of all models are represented in Figure 42 and Figure 43. The model that stands out, compared to the others, because it is always a few values over or under the measured wind direction, is the regression model. This model is rarely fully capable of estimating the correct wind direction values. However, with the real-time models and, particularly, with the machine learning model, it is possible to see that they can deduce the majority of the variations in absolute wind direction that happened throughout the entire dataset in this wind park. It is also noticeable that, for all models, there are particular spikes and changes in the values of wind direction that they cannot quite estimate, one of the reasons as to why there is a big difference in the error between this variable and wind speed. Wind speed is a variable with a much more constant outlier and variation. Wind direction sometimes will have a jump from 0 to 360° or generally low to high values, the models cannot correctly estimate.

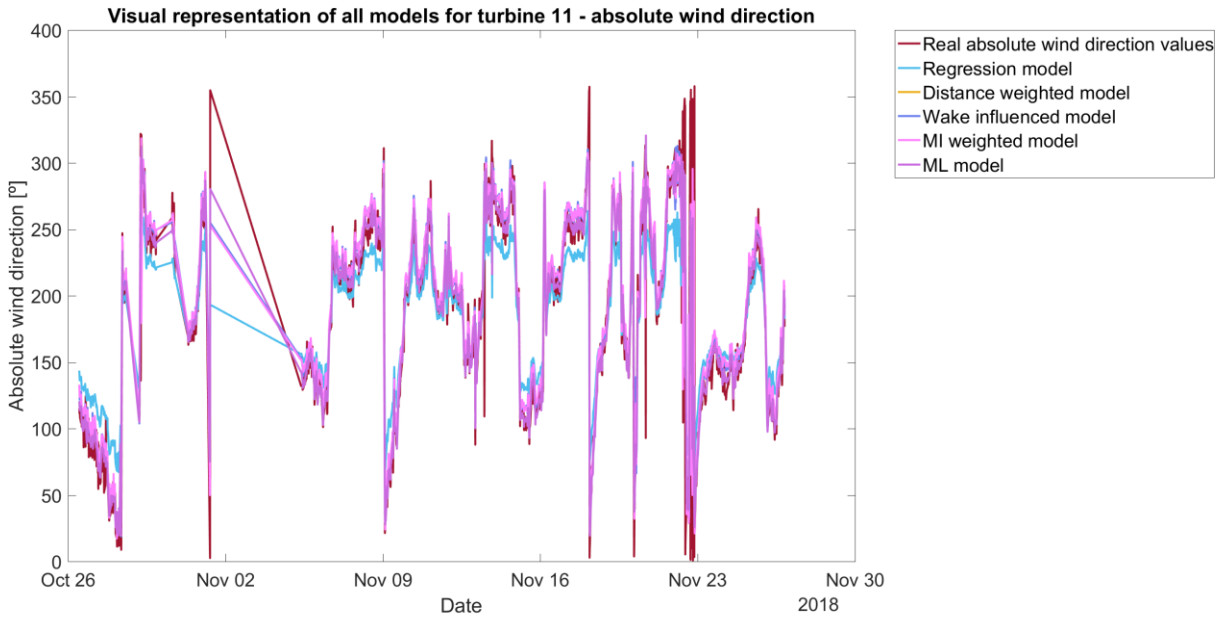


Figure 42 - Visual representation of a time-series for turbine 11 between real absolute wind direction values and estimated values for all models developed

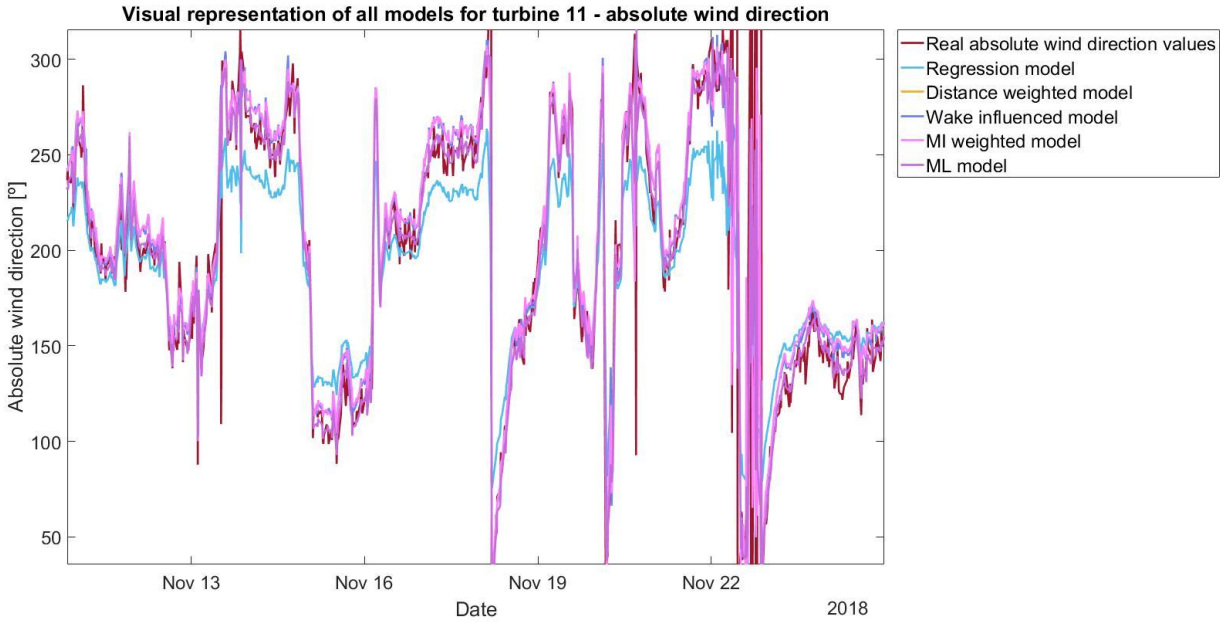


Figure 43 - Zoomed in version of the time-series for turbine 11 between real wind speed values and estimated values for all models developed

In Table 4, it is possible to see the variation in the error of every model for each of the turbines, but for values of absolute wind direction. Analysing these values, one can conclude that, as it was expected from the visual representation of the time-series, the mutual information weighted model and the machine learning model, are the most accurate models and obtain the best results. It is also possible to see from this table, that the particle model presents the worst results for absolute wind direction, and the difference between the other models is much higher for this variable. This is the reason it was decided to not represent the time-series of the particle model for absolute wind direction, since the results were, unfortunately, dissatisfactory.

Table 4 - Representation of the root mean square values of all models for every turbine in the wind park (absolute wind direction)

RMSE (°)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
RM	64.69	50.76	57.98	60.27	46.51	37.99	41.10	35.71	33.04	33.26	32.30	35.81	36.26	36.02	37.16	46.67	37.99	41.91	37.75	40.43
DWM	26.53	31.33	30.95	30.77	34.68	27.04	35.18	28.48	24.34	25.44	24.59	29.35	28.71	28.36	31.28	37.41	26.12	34.96	27.07	33.82
WIM	26.54	31.33	30.93	30.89	34.68	27.04	35.17	28.48	24.39	25.44	24.59	29.60	28.71	28.36	31.10	37.41	26.12	34.96	27.07	33.85
MIWM	25.06	27.17	26.95	29.26	30.29	28.16	28.36	29.50	27.18	24.41	23.56	28.01	25.55	26.33	24.99	32.64	24.13	29.54	23.53	31.15
MLM	22.89	24.82	28.20	23.95	24.34	25.67	26.36	27.18	21.36	21.04	21.33	27.33	26.02	20.78	28.55	26.94	21.82	27.89	17.69	32.37
PM									110.25	67.51	111.57	67.38								

Comparing every model, it is possible to conclude that for both wind speed and absolute wind direction, the mutual information weighted model and the machine learning model are the ones that have the lowest amount of error compared to the others. The regression model can be considered the most discardable model, not only because it presents worst results compared to most models, but also because it is a model with less chances to be developed and improve. On the contrary, even though the particle model presents the worst results in terms of accuracy for some turbines, this model has an exceptional opportunity to continue to be developed and become an exponentially better model.

It is also clear that there are differences in the performance of the models, depending on the turbine, since every turbine has a distinct location in the park and can be affected by different conditions (such as temperature, skewed wake or wind, and obstacles).

6. CONCLUSIONS AND FUTURE WORK

6.1. Conclusions

This dissertation presents seven distinct models that can reconstruct the wind field of a particular turbine at a park with 20 turbines. Every model has a specific behaviour and outcome. However, some models presented better and more realistic results, much more clearly than others.

The regression model can be considered the simplest model since it uses linear equations that describe wind speed in other turbines in the park. This approach can lead to situations where the model cannot estimate spikes when the wind is stronger or when the turbines are not producing (low wind speed values) since it only estimates a wind speed average. Even for different wind direction sectors, the model cannot predict significant changes. This model served as a test to better understand the present dataset and comprehend the basics of wind field reconstruction.

By using real-time measurements from the turbines and simply applying different weights to the information coming from them provided immediately better results than the first model, which assumed only an average from the linear equation of every turbine. Within the real-time models, it is possible to conclude that the one that presents the worst and best results is the distance weighted model and mutual information weighted model, respectively. The distance-weighted model presents a reasonable hypothesis. It considers that turbines further away will have a lower correlation value, meaning their information will be considered less relevant than the closer information from turbines. However, many factors such as temperature, wind conditions, obstacles in the park, and the remaining turbines that are not being evaluated, will affect how the wind crosses this park and how it reaches every turbine.

With the wake-influenced model, the conclusion was that considering turbines who may possibly be in wake does not change or affect the results the way it was expected. The difference between the error of both models (distance-weighted and wake-influence model) is so insignificant that it can almost be considered they have the same outputs. That leads to the conclusion that the number of turbines in wake, depending on the direction of the wind, is irrelevant. For most part of a normal production day, turbines are generating energy, and the altered and slower wind behind their rotor does not affect any of the closer turbines. This park was well designed and constructed in a place that allows a clear path for every turbine.

With some of the statistical analyses made during this dissertation, such as the correlation analysis and the feature selection analysis, one tested some of the real-time models using information from these sources, such as the correlation weighted model and the mutual information weighted model. The results with the correlation weighted model did not improve as expected. However, it is understandable since it was considered weights from the median fit applied to the correlation values and, when dealing with mean or median values and not exact values from a statistical analysis, the improvement in the model was not gratifying. However, with the mutual information weighted model, the used weights came from the mutual information function. This function can study the dataset and measure the amount of information one can obtain from one random variable given another. For this dissertation, this means one will have a final quantity that measures the amount of information it is possible to obtain from a random turbine, given the information of another. It will declare, for a specific turbine analysed, the turbines with more mutual information, which will be crucial while estimating wind speed and absolute wind direction values exclusively from those turbines. It is possible to conclude that this is one of the best approaches since the mutual information weighted model has the best results and better performance than the other real-time models.

As for the particle model, it is understandable that the results are not as good in reality as they are on paper since this dissertation is dealing with a particular technique: wind particles created at each turbine where their trajectory and speed is estimated at each second. Some turbines, which were more singled out in the park, such as turbine 13 (Figure 8), did not have the information of any particle surrounding them, which meant that it was not possible to estimate any value of wind speed and wind direction for this turbine. For other turbines, they would have particles where it was possible to retrieve information in some periods, and in other periods, there would be none. This instability in the estimation of wind variables would cause the absolute error of the model to be small (since calculating the error, the software only considers real values). However, when analysing a time-series, it is possible to conclude that the model cannot accurately estimate all the values.

After considering a viscosity condition for the fluid, however, it was concluded that this helped improve the results of the model. Even though this condition was not applied as initially planned due to a high computational cost, it was only possible to test it for a closed group of turbines. It was concluded that considering viscosity in the fluid provides a more realistic view of the wind particles near the turbines and a more accurate estimate of the values of wind variables.

It is also essential to highlight the importance of the particle model because this was a model created with a very different approach compared to the others for a reason. This model can provide predictions of wind characteristics such as wind speed and wind direction. In contrast, the remaining models can only estimate wind variables from real-time measurements of other turbines. This critical difference between the two techniques allows a particular advantage to the particle model since it can predict wind conditions shortly (seconds and minutes). With this, it could be possible to predict sudden changes in the wind direction of a park, wind gusts, and any atypical behaviour in the wind. With this information, it would be possible to prepare the turbines (control variables such as pitch angle of the blades) for the conditions they will face. That is merely a possibility since this type of dynamic was not implemented into this dissertation's work, but it is a positive outcome of this model that was worth mentioning.

With the machine learning model, since this was a model tested through an automatic program (regression app on Matlab), it was expected that the model with better results in this program would also be the best in terms of the overall developed models. It was decided to approach machine learning since this is a growing topic in every engineering field (28). It is a way to test how well the algorithm estimates the values of the two wind variables in question with the dataset provided. The best results were obtained by a Gaussian process regression (GPR) that infers a probability distribution over all possible values, guaranteeing improved and more accurate results than some of the other models. For our dataset, this was the model that outshined the others since it is also an algorithm that works well with small datasets , so it guaranteed one of the best results for wind speed and absolute wind direction.

One other conclusion that was possible to make with the results, was that every turbine had a specific behaviour and different outcome. Even when testing the same model, the error associated with every turbine was distinct from one another. That allowed us to conclude that the turbines had a specific error for both variables estimated. That is due to their location and different interactions with other turbines in the wind park. When examining Figure 8, it was possible to conclude that, for example, turbine 2 and turbine 9 can be affected by other turbines. That can lead to unpredictable values of wind in these turbines, which lead to a worst performance from the models. However, turbines that are further away from the others and in a more isolated space, such as turbine 5, turbine 6, and turbine 13, present better results since the wind possibly arrives at these turbines with no skewness, alteration, or lower speed. One pair of turbines in the park that are not isolated from the others but present the best results and lower errors for every model tested is turbine one and turbine 3. These two turbines are surrounded by three

turbines that were not analysed in this dissertation but are present in the park, and since the models did not use information from these turbines, they can be seen as partially isolated. However, these turbines are also in a position where they are upstream of the park. They will produce energy with a more stable wind and allow easier forecasting of wind variables for wind field reconstruction.

In this dissertation, one can conclude that statistical methods need to be involved in wind field reconstruction. However, since wind is such a complex field and depends on so many variables that will change daily, to estimate its values accurately and precisely, statistical parameters need to go hand-in-hand with fluid dynamics concepts to optimize the results.

6.2. Future work

A few topics could be approached in the future to improve and optimize the present work.

For future work, it would be wise to try and re-do the analysis on every model with a few changes in the dataset: develop the models only considering periods where it is known that the turbines are producing energy (only consider production time); develop the models dividing the dataset into two sections, considering a low speed and more stable wind for this first section and a high speed and more turbulent wind for the second section and experiment a dataset with a period of half an hour between the measurements, instead of a 10 min measurement, to see if with a half an hour time stamp, it would be easier to estimate the variations in the wind. With these alterations in the dataset, it will be possible to see the differences in the results of every model.

Statistically, high-pass filtering could be applied to all the data, allowing for high frequency and low-frequency data changes. That could also change the results of all models and perhaps improve them.

As for the models, the particle model, with a more flexible computational time, could be explored with different periods, apply the viscosity condition as initially intended, and possibly add or consider new parameters.

LIST OF REFERENCES

1. Shin JY, Ouarda TBMJ, Lee T. Heterogeneous mixture distributions for modeling wind speed, application to the UAE. Renew Energy [Internet]. 2016;91:40–52. Available from: http://dx.doi.org/10.1016/j.renene.2016.01.041
2. O.L. Hansen M. Aerodynamics of Wind Turbines. Second. London: Earthscan; 2008. 181 p.
3. Huleihil M, Mazor G. Wind Turbine Power: The Betz Limit and Beyond. Adv Wind Power. 2012;3–30.
4. Bell DA. Fundamentals of Wind Energy. Vol. 30, Physics Bulletin. 1979. 529–531 p.
5. Ranjbar MH, Zanganeh H, Gharali K, Nasrazadani SA, Kia HZ. Reaching the Betz Limit Experimentally and Numerically. Energy Equip Sys [Internet]. 2019;7(3):271–8. Available from: http://energyequipsys.ut.ac.irwww.energyequipsys.com
6. Wind Energy Technologies Office. How Do Wind Turbines Work? [Internet]. Available from: https://www.energy.gov/eere/wind/how-do-wind-turbines-work
7. Saint-Drenan YM, Besseau R, Jansen M, Staffell I, Troccoli A, Dubus L, et al. A parametric model for wind turbine power curves incorporating environmental conditions. Renew Energy. 2020;157(May):754–68.
8. Boersma S, Doekemeijer BM, Gebraad PMO, Fleming PA, Annoni J, Scholbrock AK, et al. A tutorial on control-oriented modeling and control of wind farms. Proc Am Control Conf. 2017;(May):1–18.
9. Akay B, Ragni D, Ferreira CS, Bussel GJW Van. Investigation of the root flow in a Horizontal Axis. Wind Energy. 2013;(May 2014):1–20.
10. Knudsen T, Bak T, Soltani M. Prediction models for wind speed at turbine locations in a wind farm. Wind Energy. 2011;14(7):877–94.
11. Guillemin F, Nguyen HN, Sabiron G, Di Domenico D, Boquet M. Real-time three dimensional wind field reconstruction from nacelle LiDAR measurements. J Phys Conf Ser. 2018;1037(3).
12. Borraccino A, Schlipf D, Haizmann F, Wagner R. Wind field reconstruction from nacelle-mounted lidar short-range measurements. Wind Energy Sci. 2017;2(1):269–83.
13. Theuer F, Floris Van Dooren M, Von Bremen L, Kühn M. Minute-scale power forecast of offshore wind turbines using long-range single-Doppler lidar measurements. Wind Energy Sci. 2020;5(4):1449–68.
14. Rott A, Petrović V, Kühn M. Wind farm flow reconstruction and prediction from high frequency

SCADA Data. J Phys Conf Ser. 2020;1618(6).

15. Rinnan Å, Nørgaard L, Berg F van den, Thygesen J, Bro R, Engelsen SB. Chapter 2 - Data Pre-processing. Infrared Spectrosc Food Qual Anal Control [Internet]. 2009;3:29–50. Available from: https://www.sciencedirect.com/science/article/pii/B978012374136300002X
16. Pandey P. Data Preprocessing: Concepts [Internet]. 2019. Available from: https://towardsdatascience.com/data-preprocessing-concepts-fa946d11c825
17. Upadhyay A. Formula to Find Bearing or Heading angle between two points: Latitude Longitude [Internet]. Available from: https://www.igismap.com/formula-to-find-bearing-or-heading-angle-between-two-points-latitude-longitude/
18. Vogt W, Johnson R. Correlation and Regression Analysis. Correl Regres Anal. 2015;
19. Huijskens T. Mutual information-based feature selection [Internet]. 2017. Available from: https://thuijskens.github.io/2017/10/07/feature-selection/
20. Valldecabres L, Nygaard NG, Vera-Tudela L, von Bremen L, Kühn M. On the use of dual-doppler radar measurements for very short-term wind power forecasts. Remote Sens. 2018;10(11):1–22.
21. Nonlinear System VS Linear System [Internet]. 2015. Available from: https://researchhubs.com/post/maths/fundamentals/bell-shaped-function.html
22. Maheshwar M. International Journal of Research in Pharmacy and Chemistry a Review Article on Measurement of Viscosity. Int J Res Pharm Chem [Internet]. 2018;8(1):69–77. Available from: www.ijrpc.com
23. IBM Cloud Education. Machine Learning [Internet]. 2020. Available from: https://www.ibm.com/cloud/learn/machine-learning
24. Tamir DM. What Is Machine Learning? [Internet]. 2020. Available from: https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/
25. Quinonero-Candela J, Edward Rasmussen C. A Unifying View of Sparse Approximate Gaussian Process Regression. 2005;6:1939–59.
26. Wang J. An Intuitive Tutorial to Gaussian Processes Regression. 2020; Available from: http://arxiv.org/abs/2009.10862
27. Lang S, McKeogh E. LIDAR and SODAR measurements of wind speed and direction in upland terrain for wind energy purposes. Remote Sens. 2011;3(9):1871–901.

ANNEX A

In annex A, it is presented the tables of distances and angles between every turbine in the park, mentioned in chapter 3.2.2.

Table 5 – Angle between every turbine in the park [°]

Turbines	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		181.49	176.91	1.55	7.40	0.80	0.01	0.01	359.98	0.03	359.87	359.94	359.69	359.57	359.43	359.16	359.04	359.31	0.95	2.65
2	1.06		2.44	1.08	2.68	0.63	0.06	0.02	359.96	0.03	359.83	359.92	359.60	359.60	359.45	359.19	359.04	359.29	0.87	1.45
3	357.19	183.15		0.80	3.71	0.56	359.97	360.00	359.99	0.02	359.94	359.97	359.82	359.66	359.54	359.31	359.21	359.46	359.96	1.16
4	182.35	182.32	181.33		177.36	0.75	359.79	359.96	0.04	360.00	0.08	0.02	0.05	359.66	359.56	359.33	359.30	359.58	183.31	180.42
5	188.79	184.51	184.82	357.94		0.03	359.72	359.94	0.07	359.97	0.22	0.08	0.32	359.87	359.82	359.67	359.68	359.91	344.17	351.74
6	182.22	182.49	181.71	181.36	180.06		359.08	359.83	0.16	359.93	0.39	0.17	0.48	359.79	359.74	359.53	359.58	359.89	182.14	181.27
7	180.08	180.57	179.77	179.06	178.41	177.73		0.06	359.91	0.24	359.68	359.75	359.47	358.78	358.63	358.30	358.33	358.73	179.77	179.18
8	180.23	180.68	179.95	179.37	178.77	178.63	180.21		359.91	3.10	359.71	359.69	359.54	358.55	358.47	358.18	358.30	358.71	179.96	179.45
9	180.35	180.77	180.09	179.61	179.06	179.11	180.21	180.06		179.85	359.77	359.25	359.63	358.06	358.19	358.00	358.24	358.68	180.11	179.66
10	180.82	181.26	180.54	179.98	179.38	179.30	180.96	183.67	0.27		0.21	0.09	0.08	359.03	358.97	358.70	358.83	359.24	180.56	180.05
11	180.42	180.82	180.22	179.82	179.38	179.52	180.16	180.04	179.95	179.97		180.06	359.74	270.52	355.62	356.96	357.85	358.32	180.21	179.86
12	180.60	181.02	180.36	179.88	179.36	179.42	180.34	180.13	179.54	179.97	0.18		359.95	358.10	358.35	358.19	358.48	358.93	180.36	179.93
13	180.40	180.74	180.26	179.95	179.64	179.77	180.10	180.03	179.97	180.00	179.90	179.99		180.75	181.55	187.05	352.28	305.49	180.23	179.98
14	181.45	181.92	181.27	180.74	180.37	180.25	180.60	180.21	179.58	180.12	91.86	179.32	1.93		359.93	359.47	359.82	0.49	181.18	180.83
15	181.33	181.80	181.17	180.66	180.34	180.22	180.46	180.16	179.73	180.09	176.98	179.59	2.75	179.95		359.19	359.81	0.73	181.07	180.75
16	181.32	181.80	181.20	180.69	180.44	180.27	180.39	180.13	179.79	180.08	178.57	179.70	8.51	179.75	179.45		0.52	3.24	181.06	180.79
17	181.04	181.48	180.94	180.50	180.30	180.16	180.27	180.08	179.87	180.05	179.30	179.82	173.58	179.94	179.91	180.36		177.75	180.81	180.58
18	180.89	181.31	180.77	180.36	180.10	180.05	180.24	180.07	179.89	180.04	179.35	179.85	126.37	180.19	180.41	182.66	357.33		180.67	180.42
19	181.21	181.57	179.94	2.76	163.04	0.98	359.96	360.00	359.99	0.03	359.92	359.95	359.78	359.55	359.42	359.15	359.06	359.34		46.30
20	183.38	182.63	181.62	0.36	171.09	0.58	359.85	359.97	0.03	0.00	0.05	0.01	0.01	359.68	359.58	359.36	359.32	359.58	226.79	

Table 6 – Distances between every turbine in the park [km]

Turbines	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		1.15	0.45	1.40	2.02	2.51	1.98	2.23	2.52	2.50	3.10	2.73	4.21	4.04	4.29	4.89	5.28	4.69	0.61	1.17
2	1.15		1.26	2.50	2.93	3.61	3.02	3.31	3.61	3.64	4.21	3.86	5.30	5.17	5.43	6.03	6.43	5.83	1.76	2.23
3	0.45	1.26		1.27	1.70	2.37	2.22	2.43	2.69	2.56	3.25	2.84	4.35	3.97	4.24	4.83	5.28	4.71	0.71	0.99
4	1.40	2.50	1.27		0.94	1.12	1.58	1.65	1.78	1.44	2.24	1.78	3.29	2.70	2.97	3.57	4.03	3.49	0.84	0.30
5	2.02	2.93	1.70	0.94		1.12	2.48	2.49	2.55	2.05	2.90	2.43	3.84	2.73	3.04	3.60	4.17	3.75	1.62	0.94
6	2.51	3.61	2.37	1.12	1.12		1.99	1.87	1.80	1.17	1.97	1.53	2.80	1.66	1.96	2.54	3.07	2.63	1.94	1.38
7	1.98	3.02	2.22	1.58	2.48	1.99		0.32	0.65	1.06	1.23	1.03	2.30	2.79	2.95	3.50	3.70	3.03	1.51	1.75
8	2.23	3.31	2.43	1.65	2.49	1.87	0.32		0.33	0.82	0.92	0.72	2.00	2.51	2.65	3.20	3.39	2.72	1.72	1.86
9	2.52	3.61	2.69	1.78	2.55	1.80	0.65	0.33		0.64	0.60	0.43	1.69	2.23	2.36	2.89	3.06	2.38	1.98	2.03
10	2.50	3.64	2.56	1.44	2.05	1.17	1.06	0.82	0.64		0.85	0.38	1.86	1.75	1.93	2.51	2.81	2.19	1.90	1.72
11	3.10	4.21	3.25	2.24	2.90	1.97	1.23	0.92	0.60	0.85		0.47	1.10	1.91	1.97	2.44	2.53	1.84	2.55	2.51
12	2.73	3.86	2.84	1.78	2.43	1.53	1.03	0.72	0.43	0.38	0.47		1.52	1.80	1.93	2.47	2.69	2.03	2.15	2.06
13	4.21	5.30	4.35	3.29	3.84	2.80	2.30	2.00	1.69	1.86	1.10	1.52		2.00	1.88	2.11	1.89	1.25	3.65	3.58
14	4.04	5.17	3.97	2.70	2.73	1.66	2.79	2.51	2.23	1.75	1.91	1.80	2.00		0.31	0.88	1.44	1.18	3.43	2.99
15	4.29	5.43	4.24	2.97	3.04	1.96	2.95	2.65	2.36	1.93	1.97	1.93	1.88	0.31		0.60	1.13	0.92	3.68	3.27
16	4.89	6.03	4.83	3.57	3.60	2.54	3.50	3.20	2.89	2.51	2.44	2.47	2.11	0.88	0.60		0.70	0.89	4.28	3.86
17	5.28	6.43	5.28	4.03	4.17	3.07	3.70	3.39	3.06	2.81	2.53	2.69	1.89	1.44	1.13	0.70		0.69	4.68	4.33
18	4.69	5.83	4.71	3.49	3.75	2.63	3.03	2.72	2.38	2.19	1.84	2.03	1.25	1.18	0.92	0.89	0.69		4.08	3.80
19	0.61	1.76	0.71	0.84	1.62	1.94	1.51	1.72	1.98	1.90	2.55	2.15	3.65	3.43	3.68	4.28	4.68	4.08		0.69
20	1.17	2.23	0.99	0.30	0.94	1.38	1.75	1.86	2.03	1.72	2.51	2.06	3.58	2.99	3.27	3.86	4.33	3.80	0.69	

ANNEX B

In annex B, it is presented the graphs for the correlation analysis proceeded in chapter 3.3.1, for the intervals [0-5] and [5-11] m/s and an absolute wind direction interval of [180-360°].

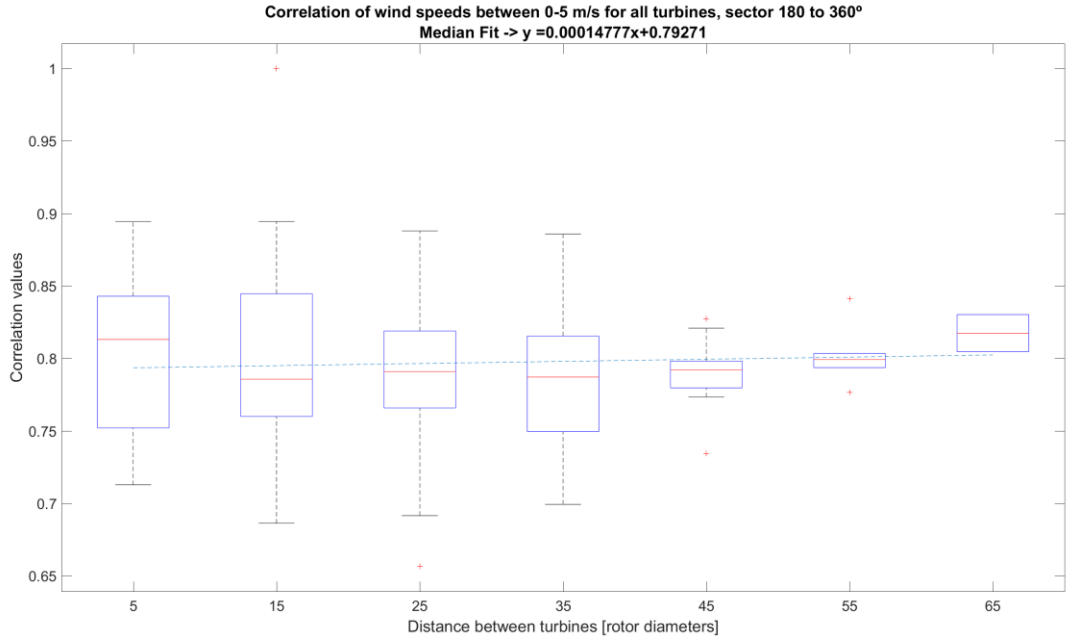


Figure 44 - Graphic representation of the correlation values between wind speed and distances between turbines for a wind speed interval of 0 to 5 m/s and wind direction sector of 180 to 360°

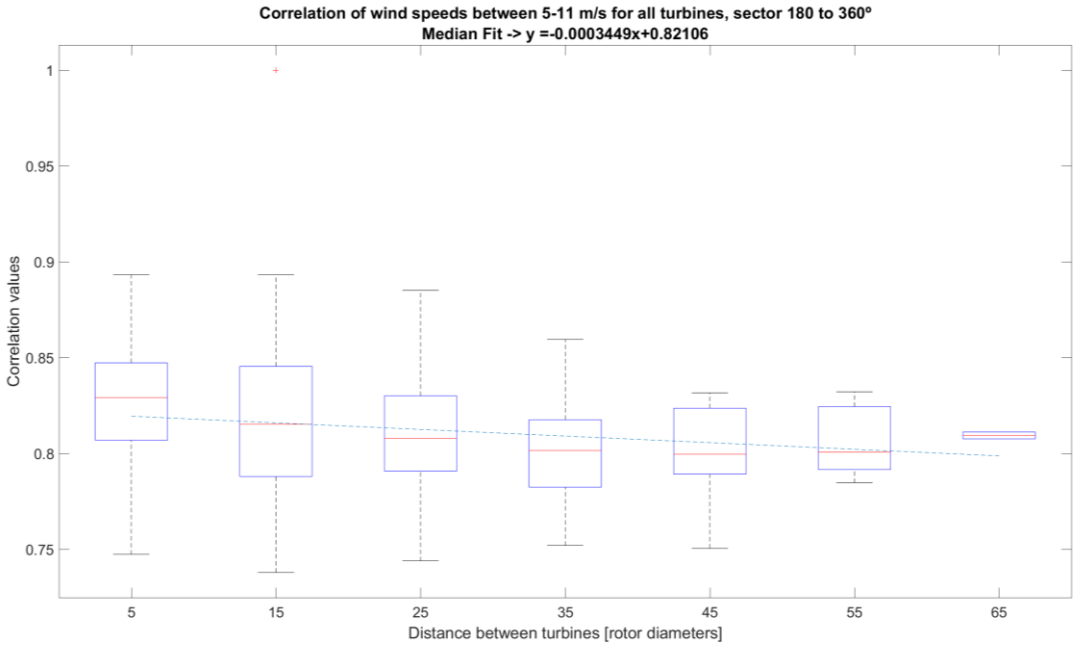


Figure 45 - Graphic representation of the correlation values between wind speed and distances between turbines for a wind speed interval of 5 to 11 m/s and wind direction sector of 180 to 360°

ANNEX C

In annex C, it is presented the graphs for feature selection analysis presented in chapter 3.3.1 of every remaining turbine in the wind park.

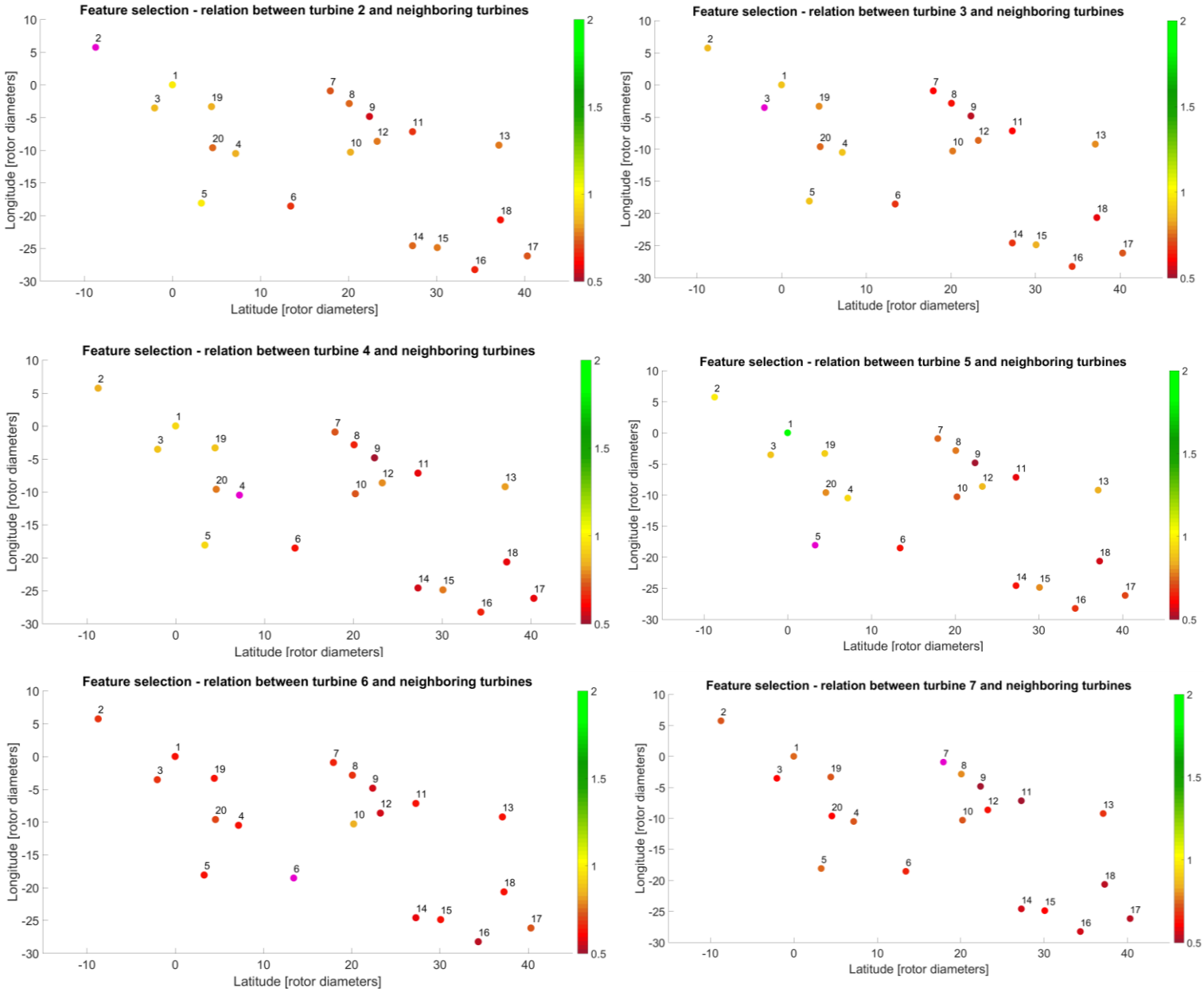


Figure 46 - Graphic representation of the feature selection analysis for turbines 3, 4, 5, 6 and 7

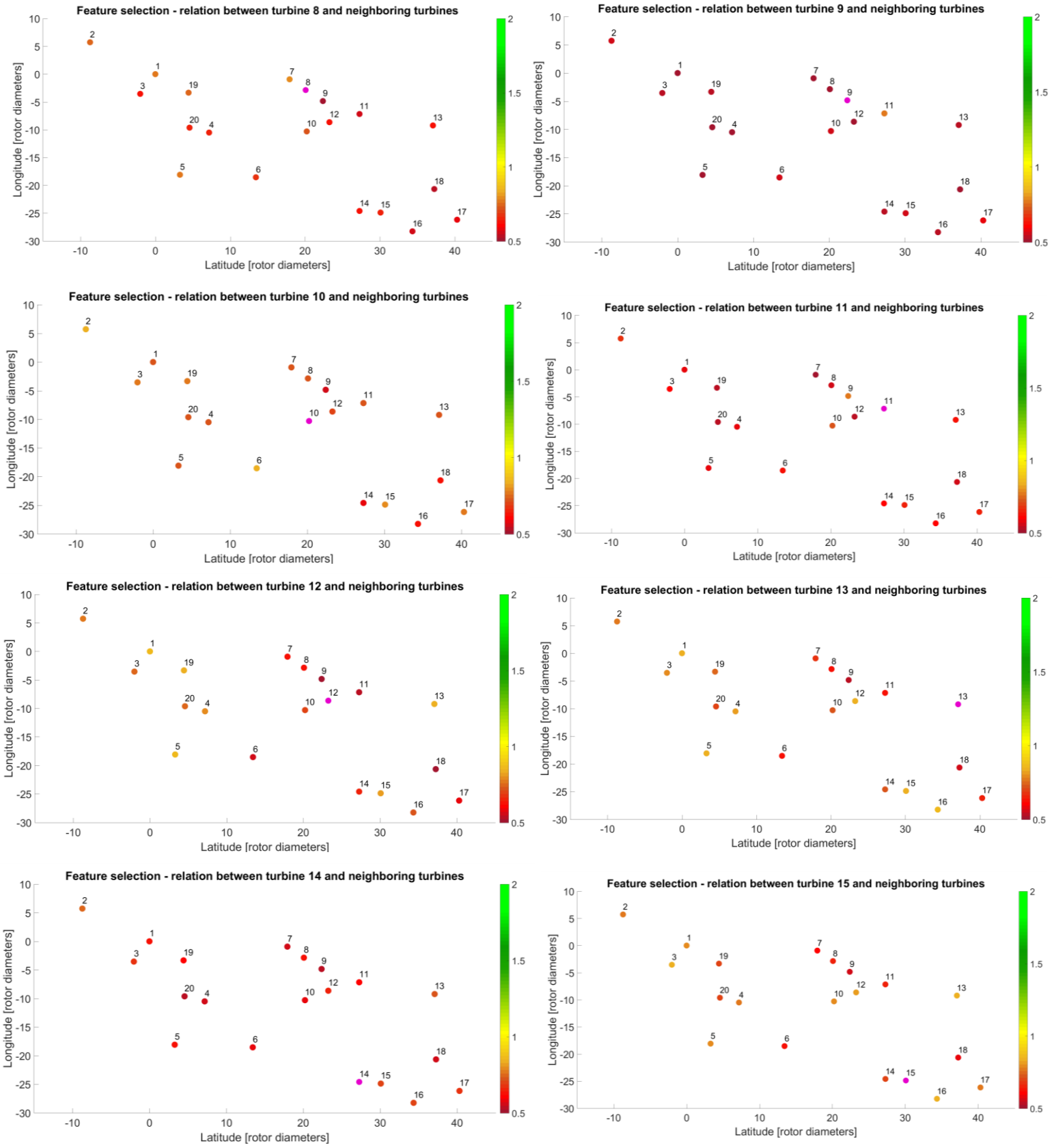
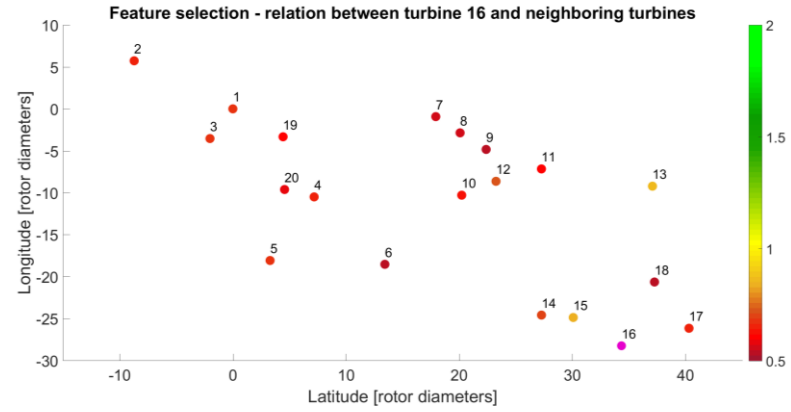
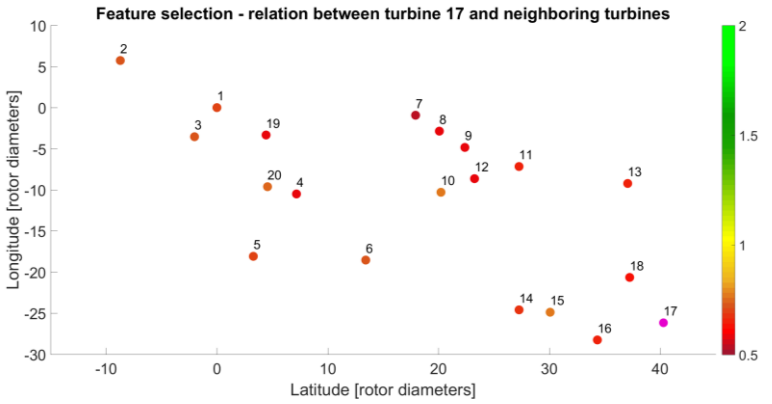
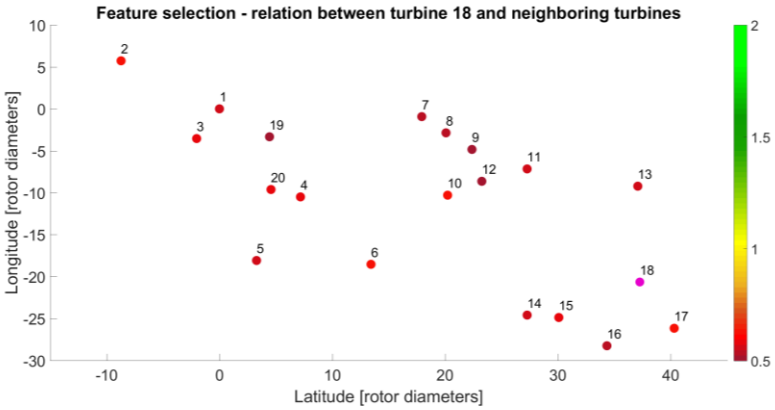
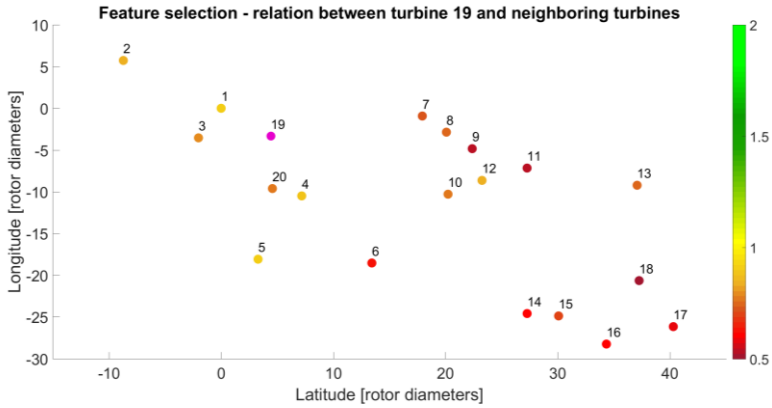


Figure 47 - Graphic representation of the feature selection analysis for turbines 8, 9, 10, 11, 12, 13, 14 and 15









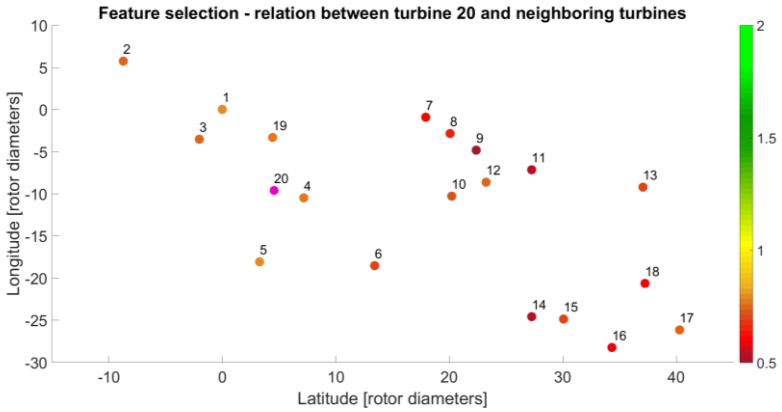
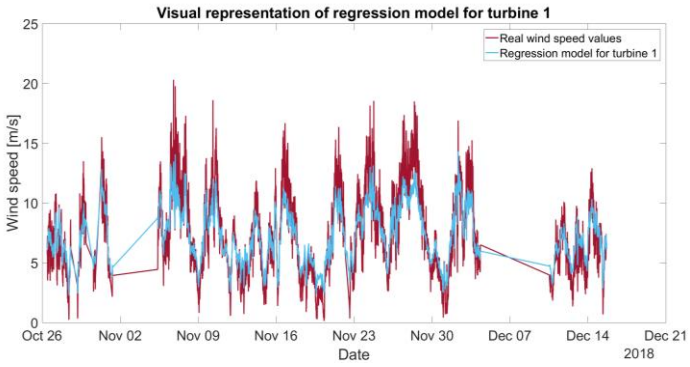


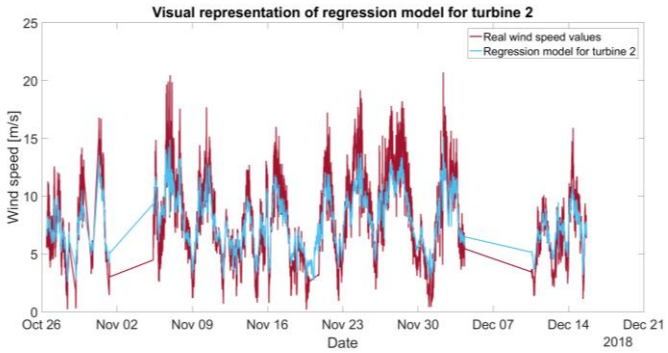
Figure 48 - Graphic representation of the feature selection analysis for turbines 16, 17, 18, 19 and 20

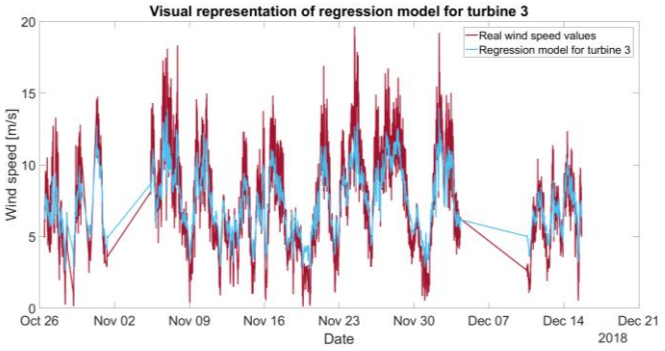
ANNEX D

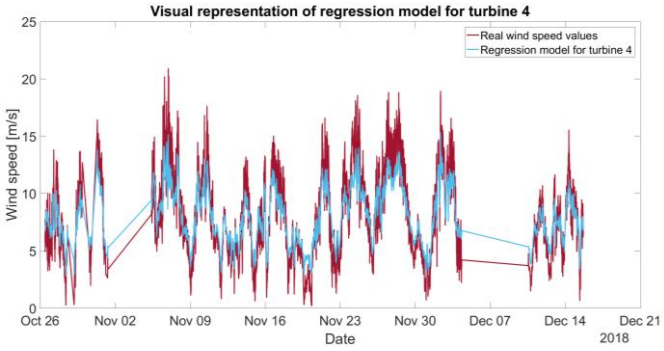
In annex D, it is presented the graphs that represent the time-series for the variable wind speed and absolute wind direction, presented in chapter 5 for the remaining turbines in the wind park.

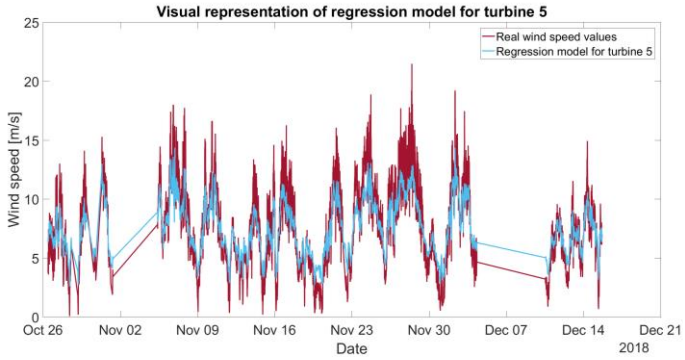
Wind speed: Regression model

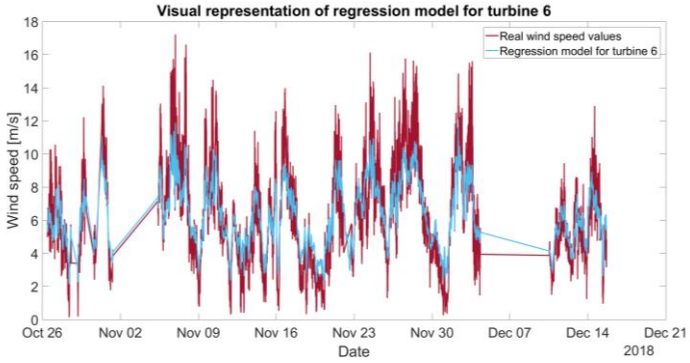


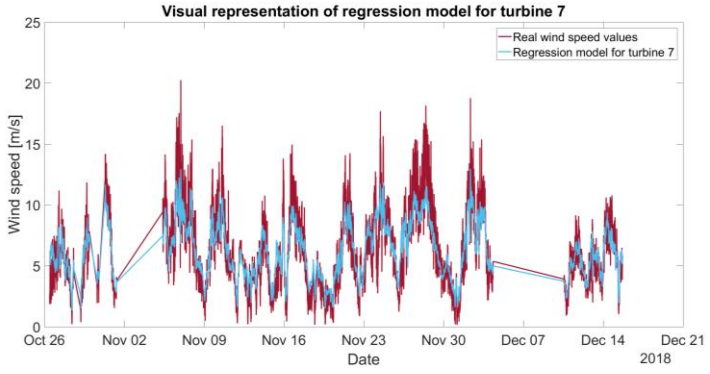


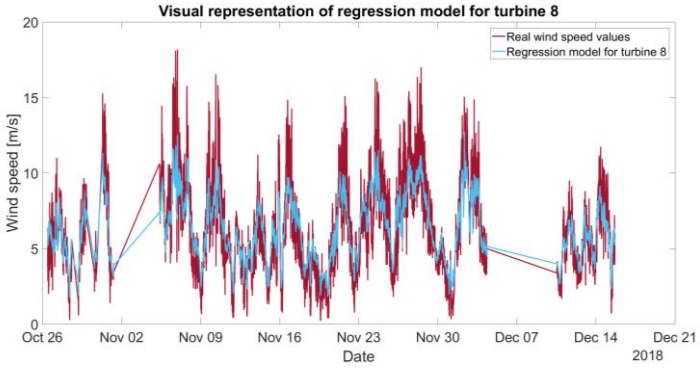


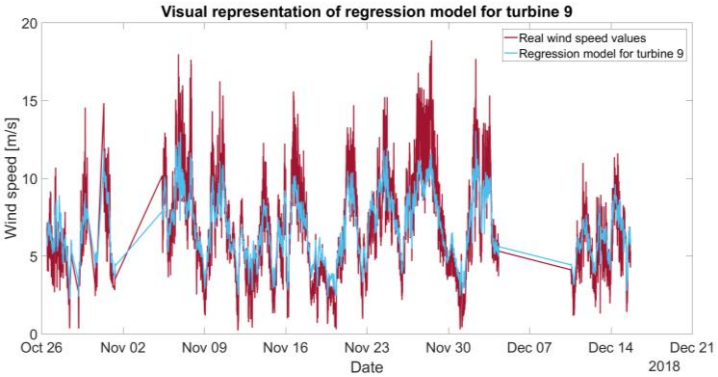


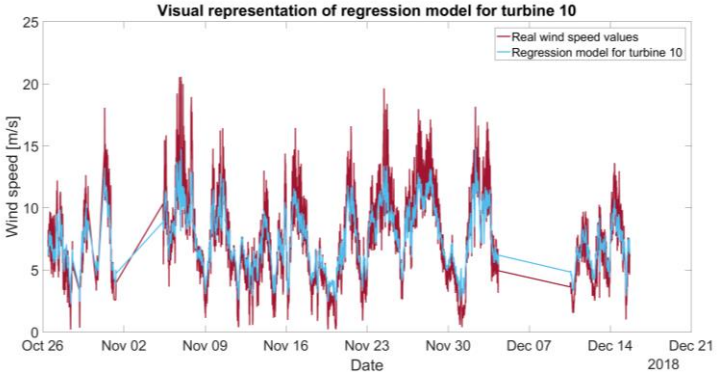


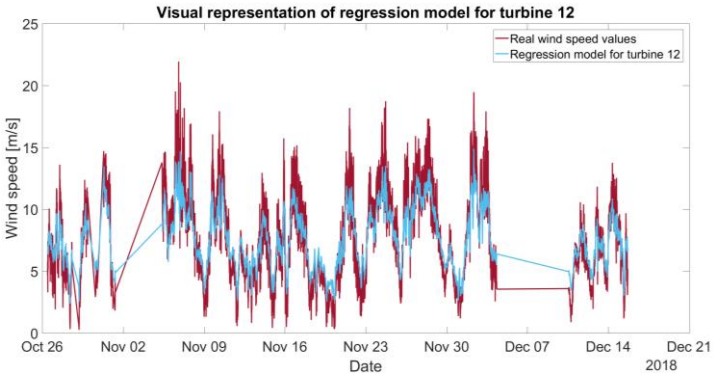


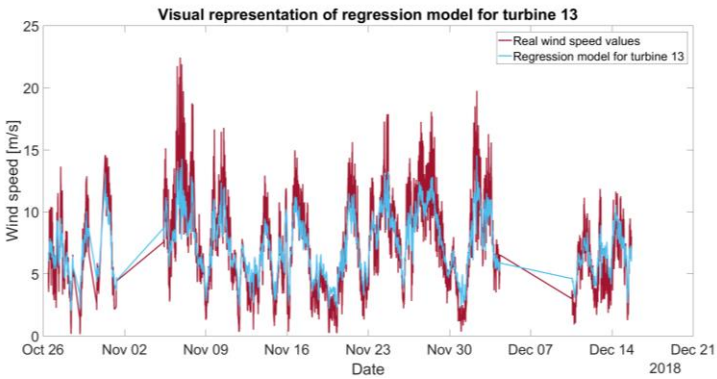


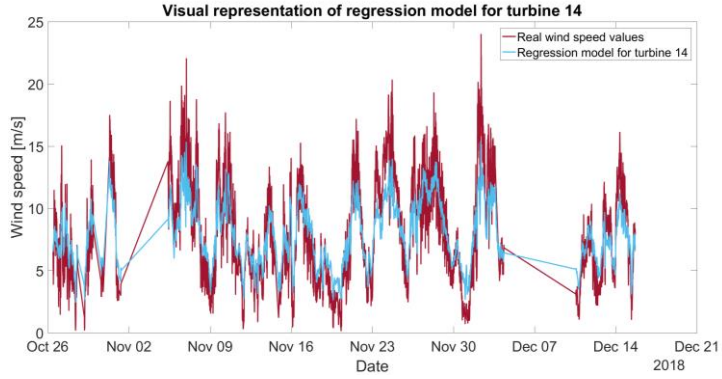


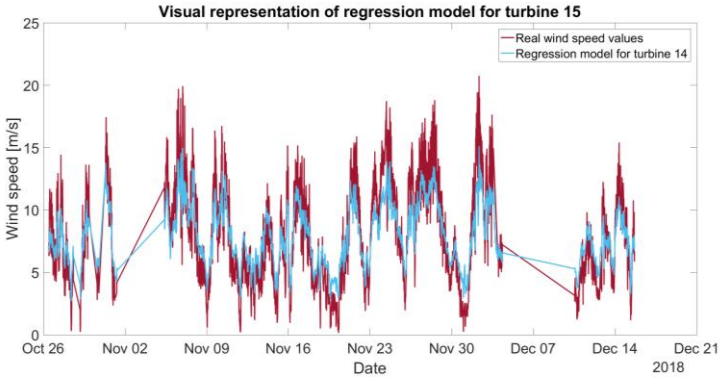


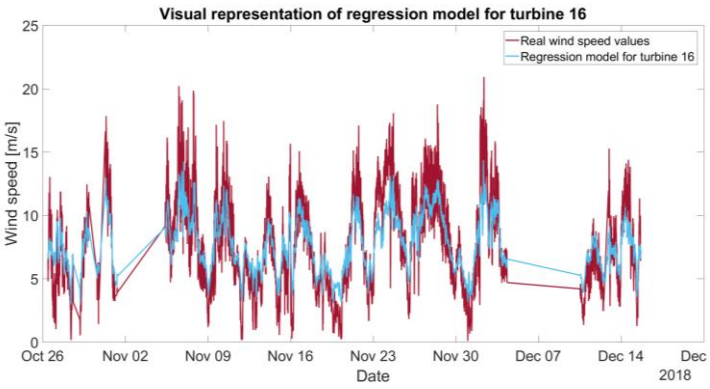


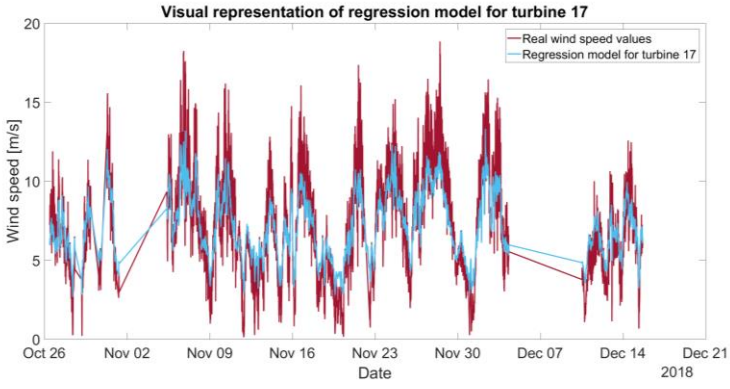


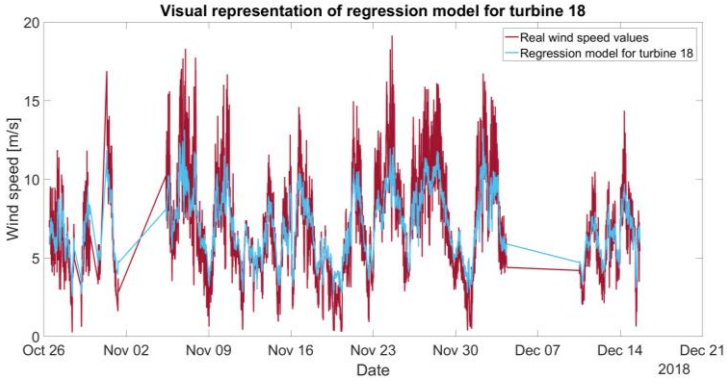


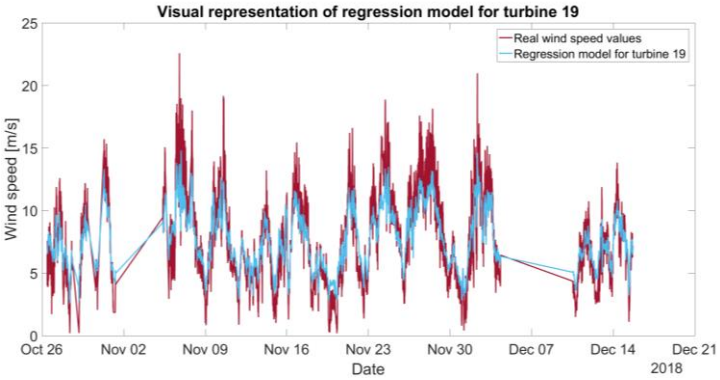


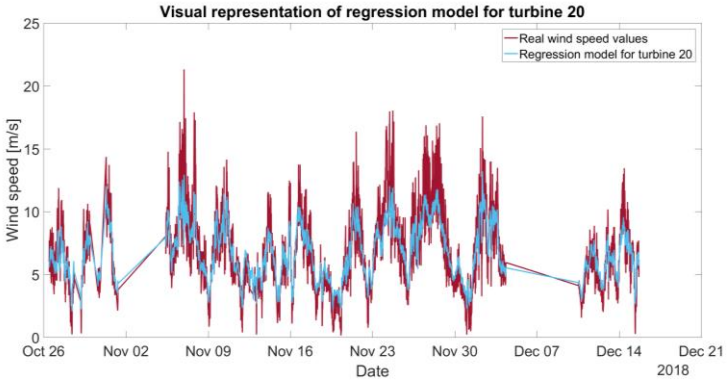




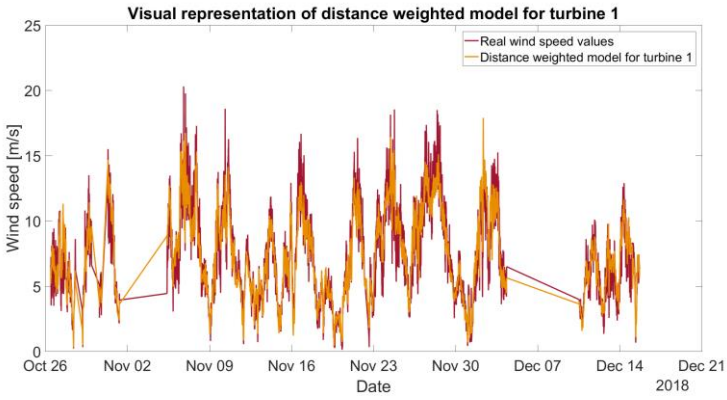


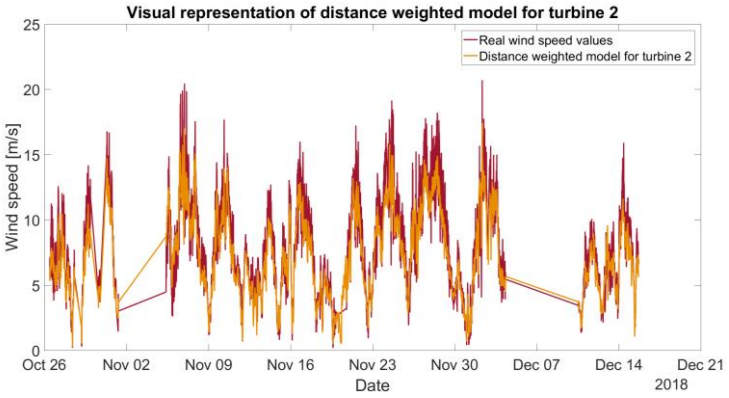


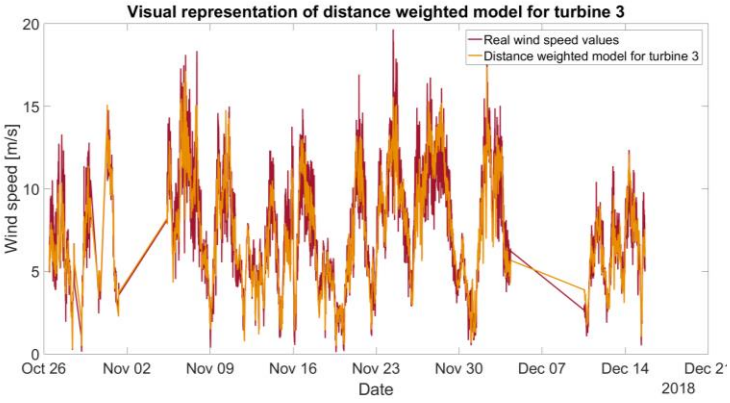


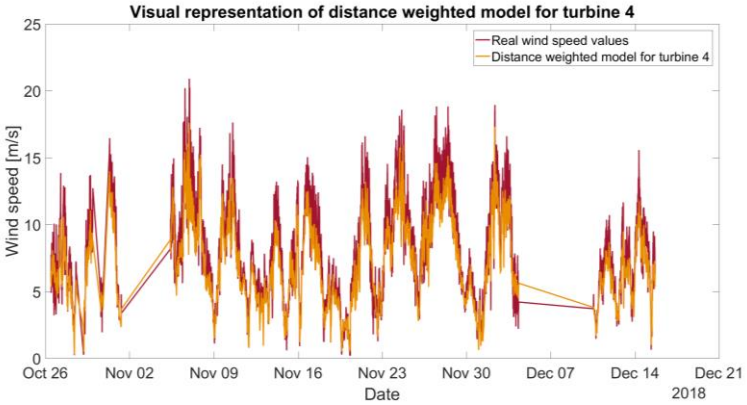


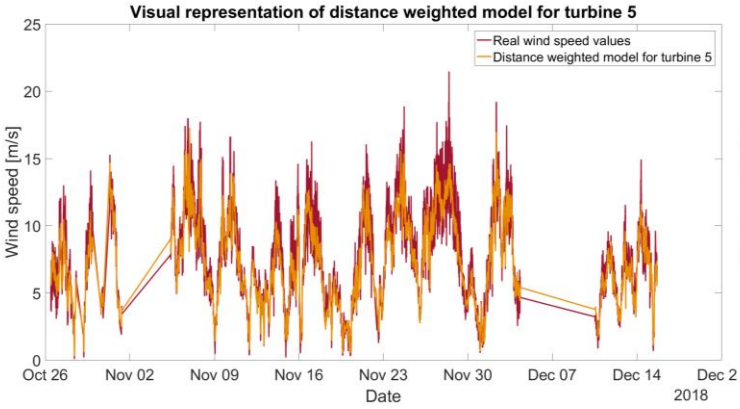
Distance weighted model

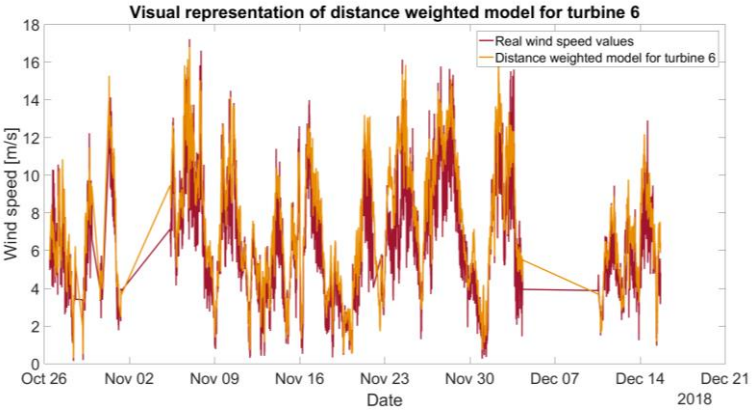


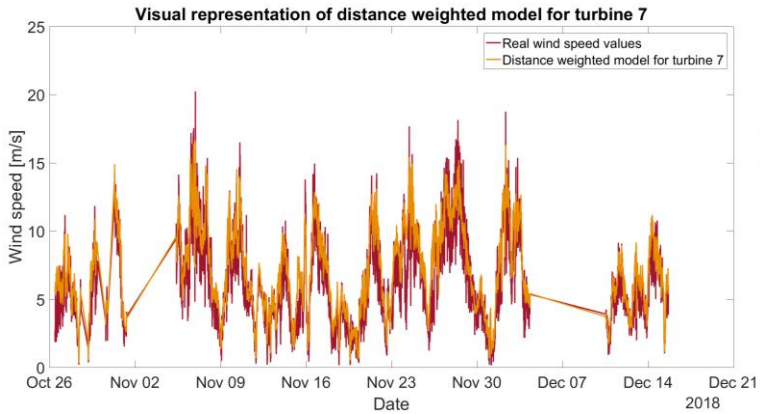


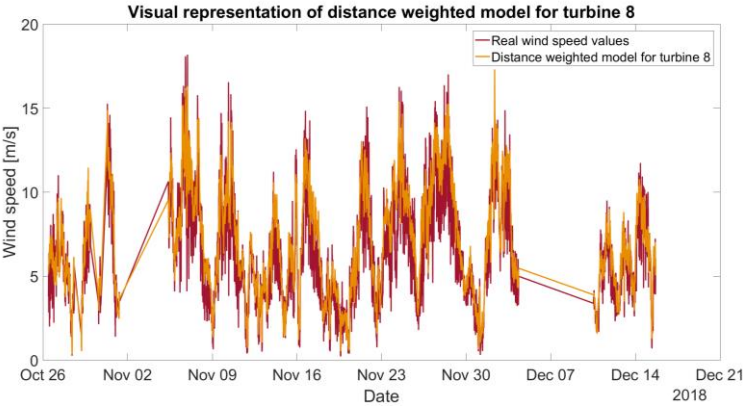


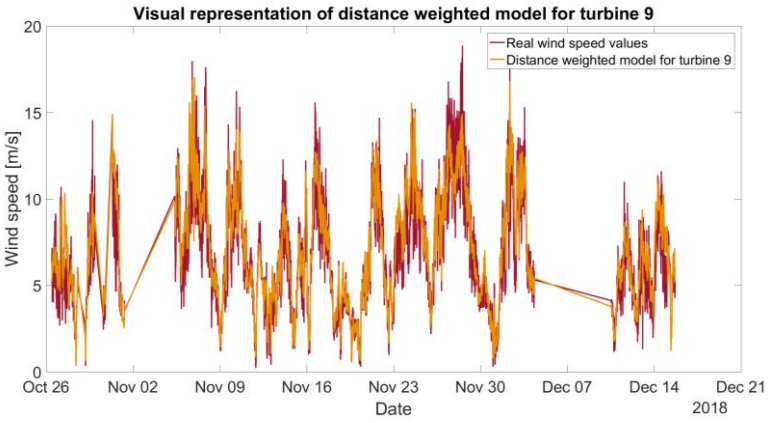


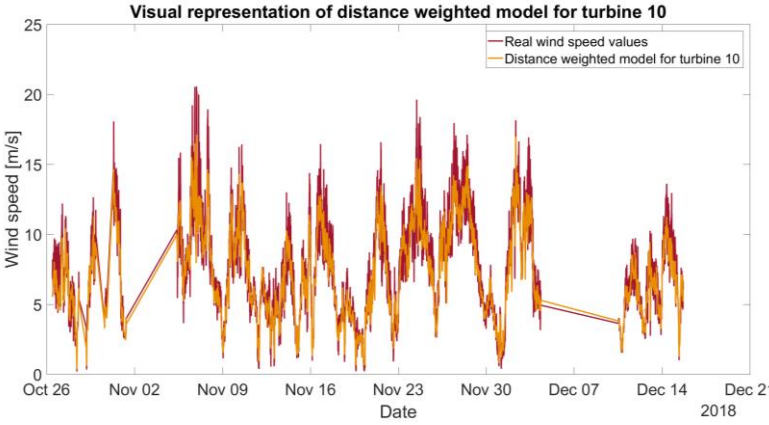


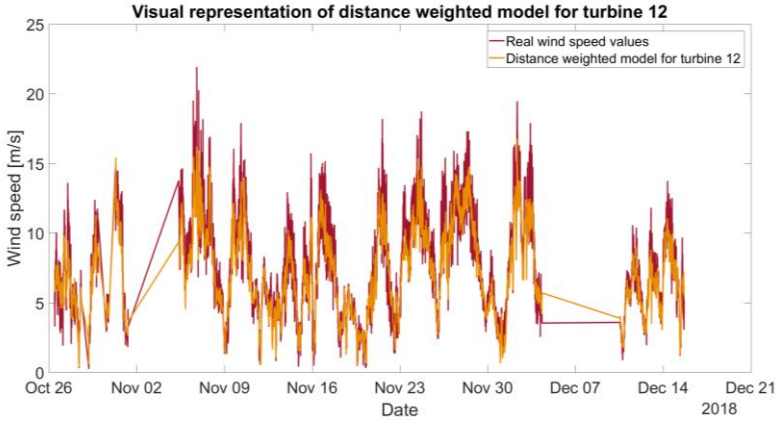


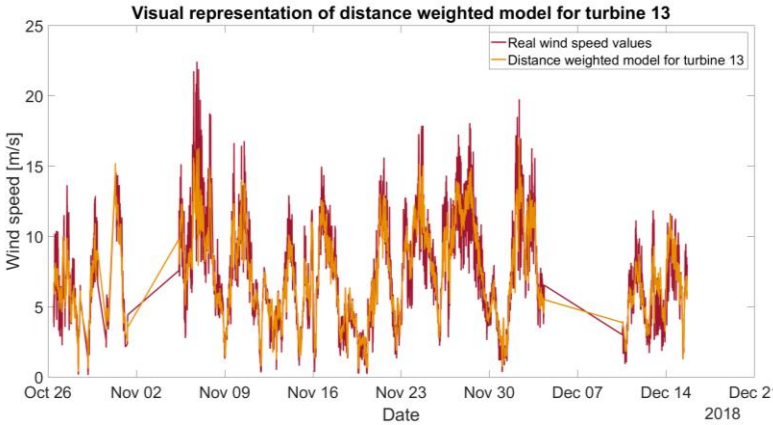


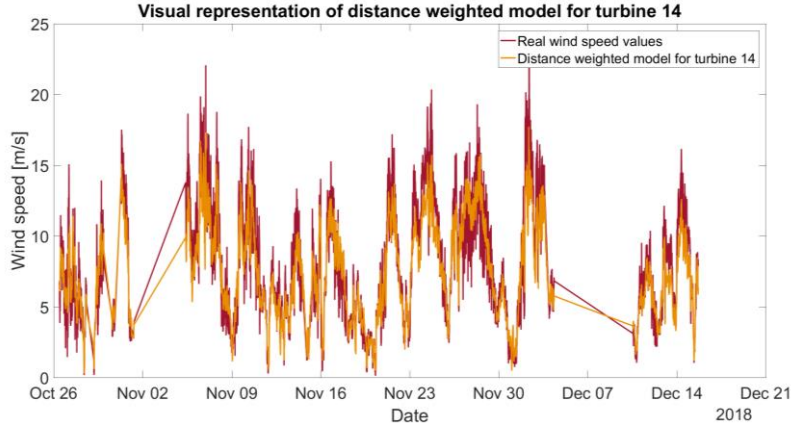


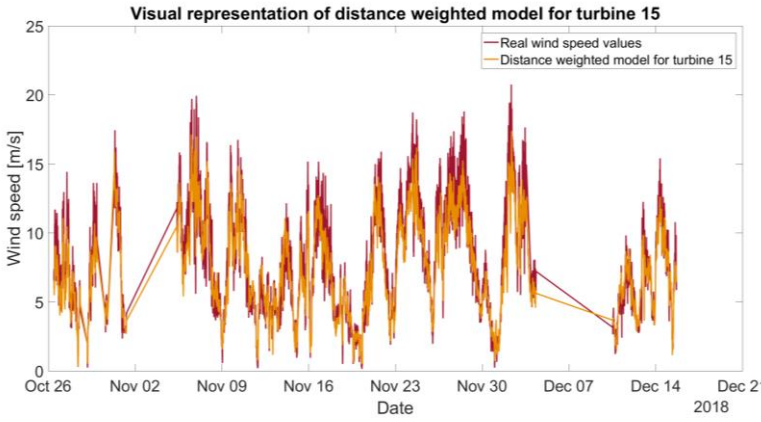


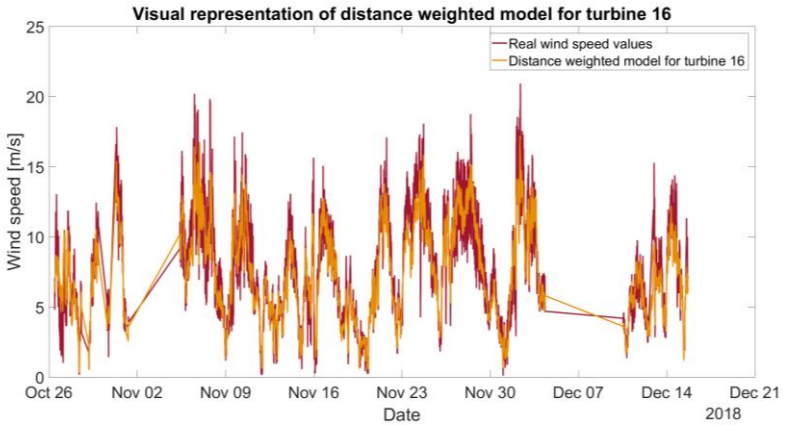


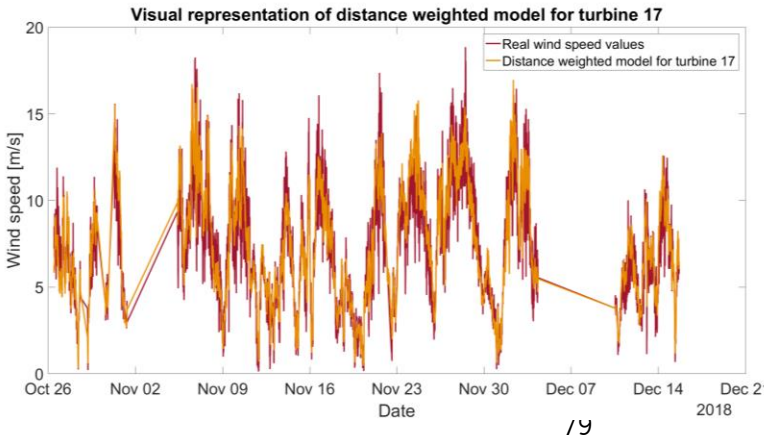


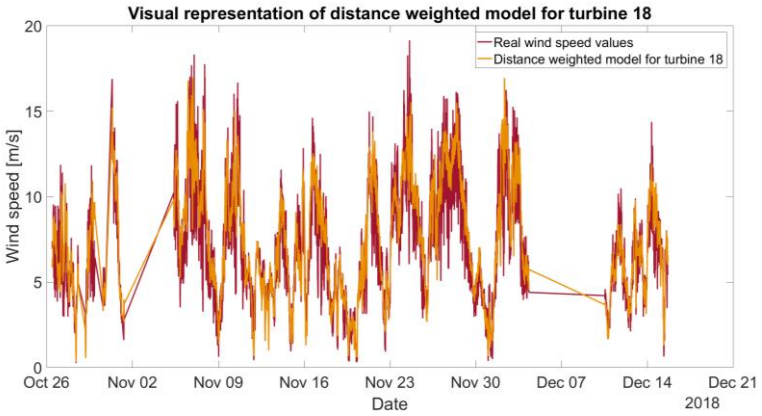


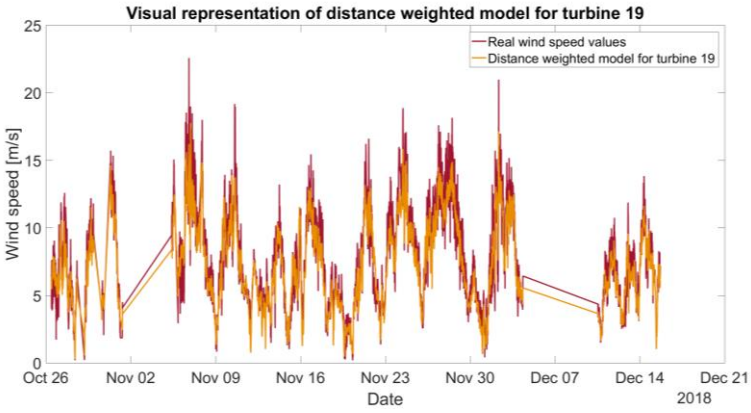


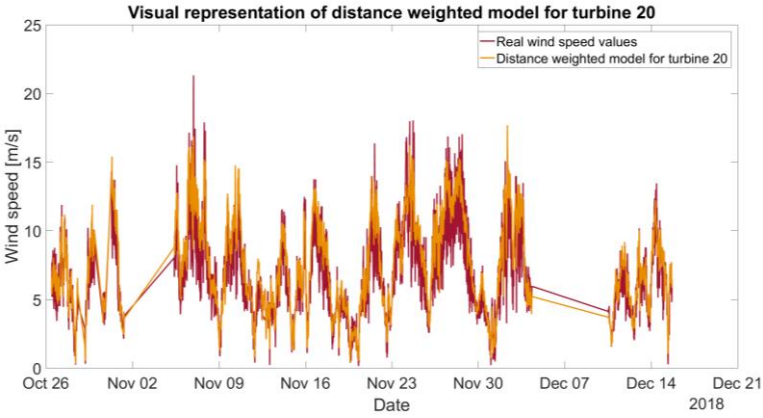




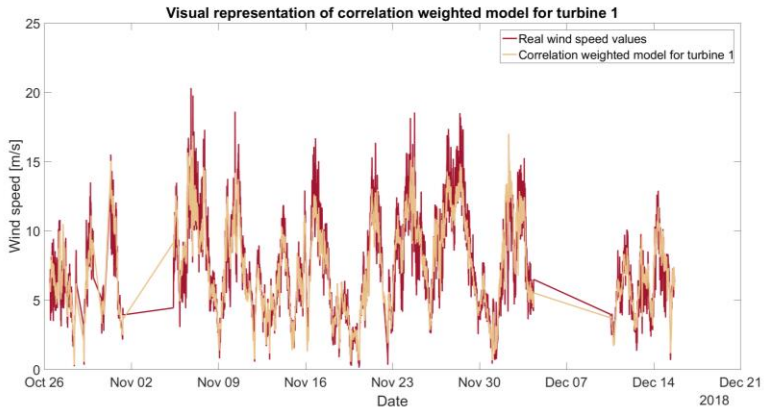


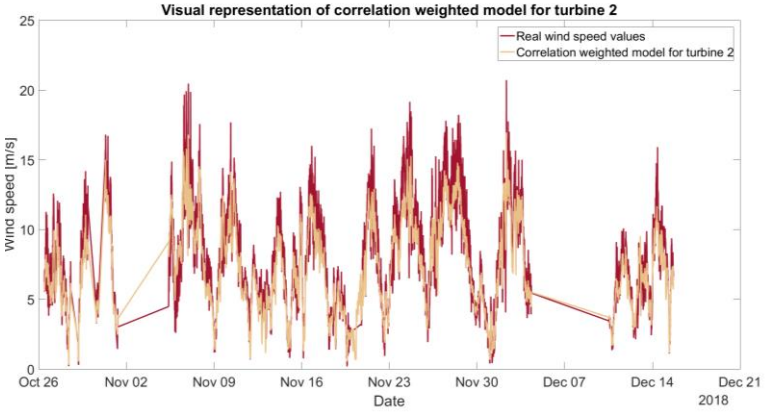


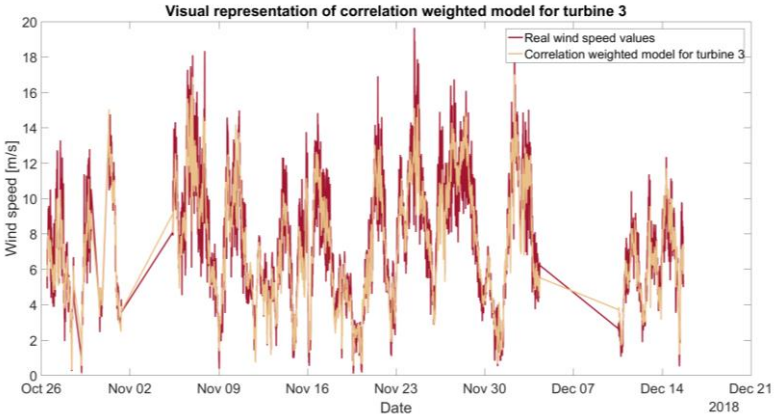


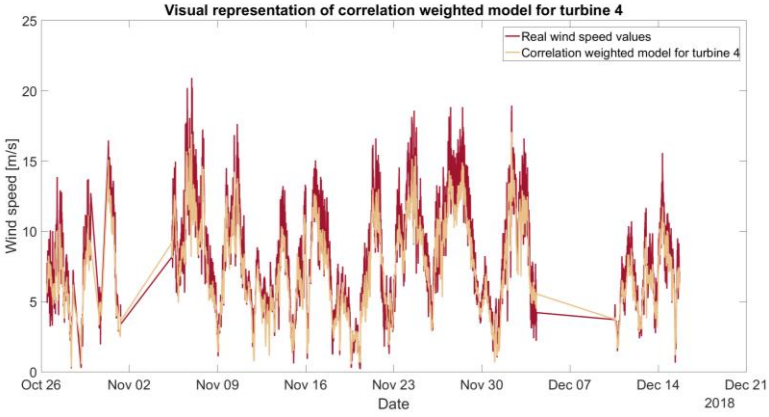


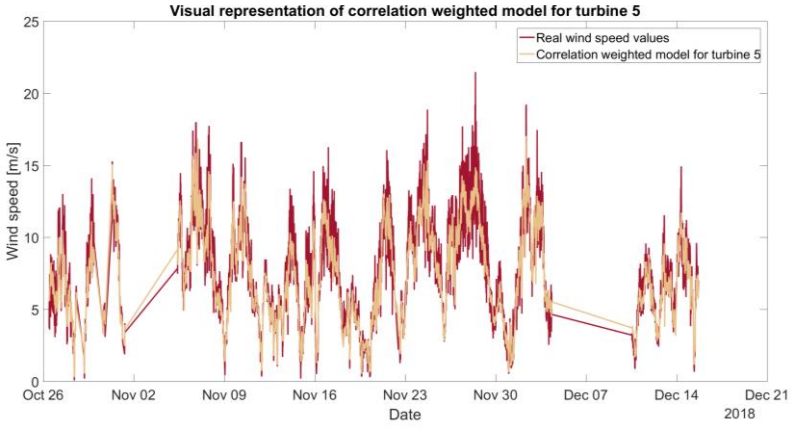
Correlation weighted model

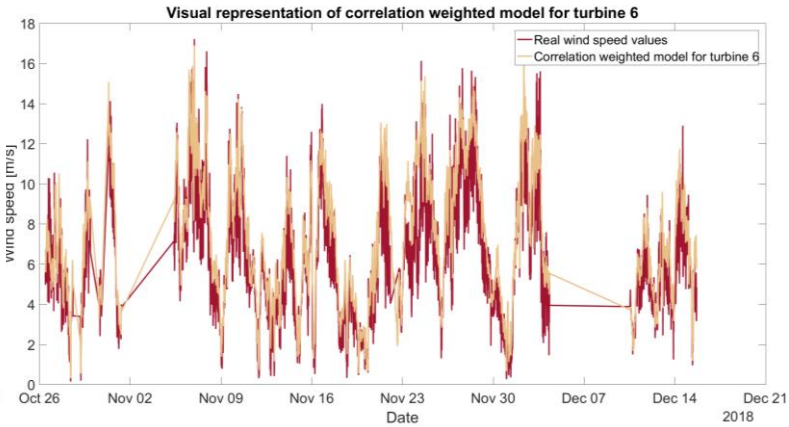


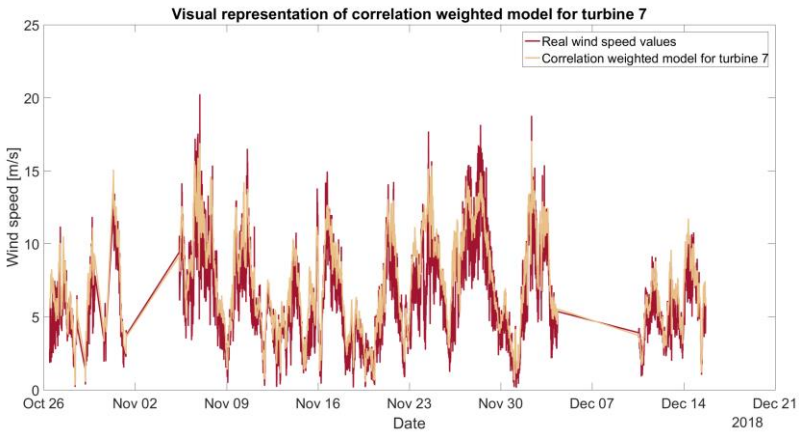


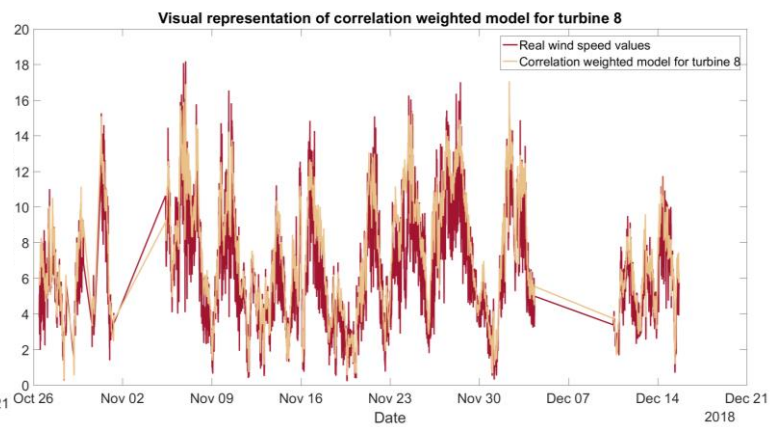


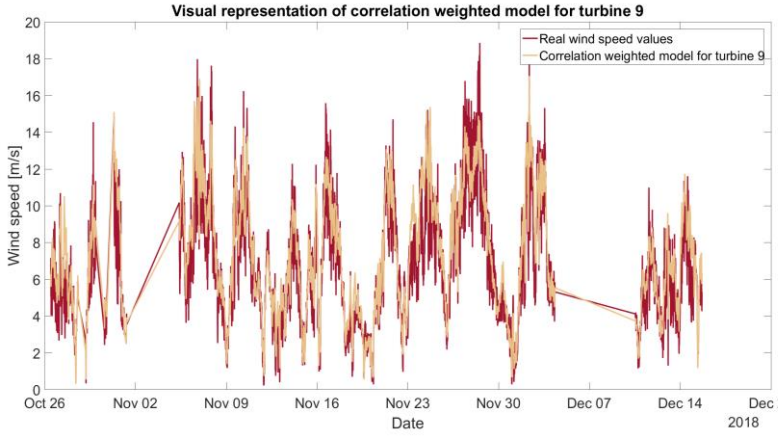


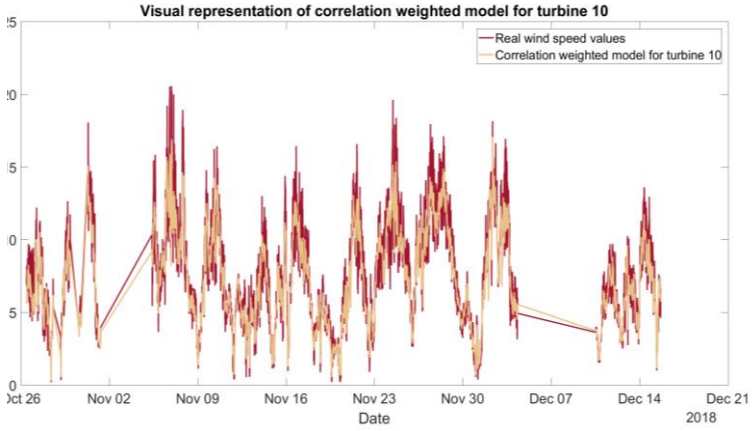


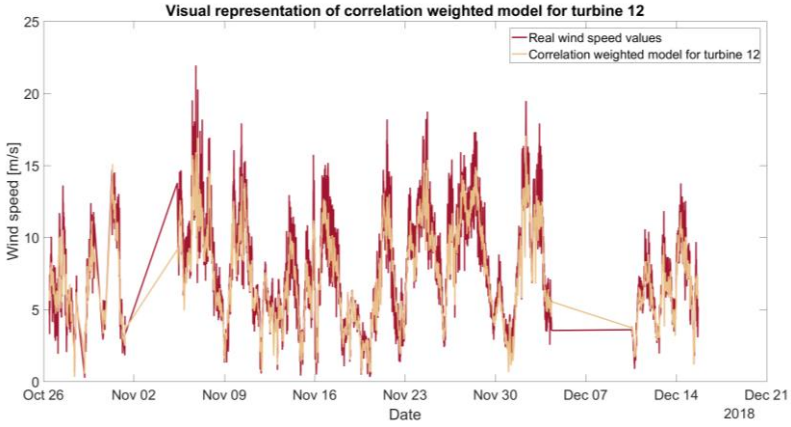


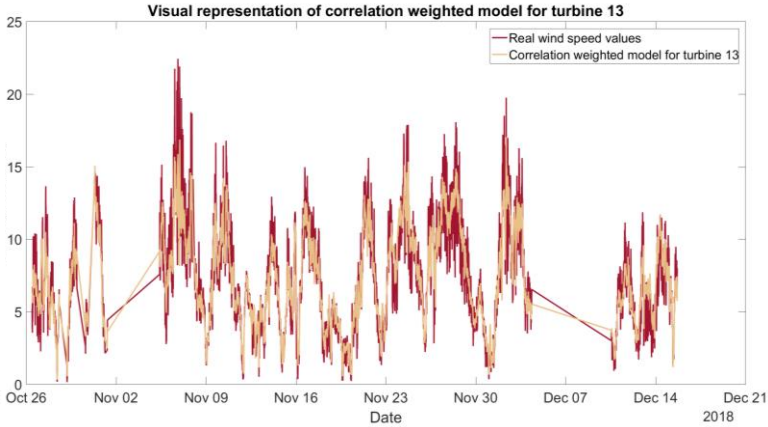


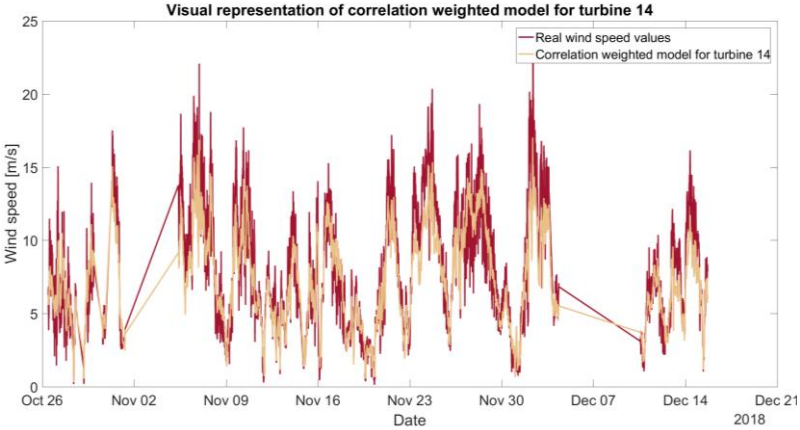


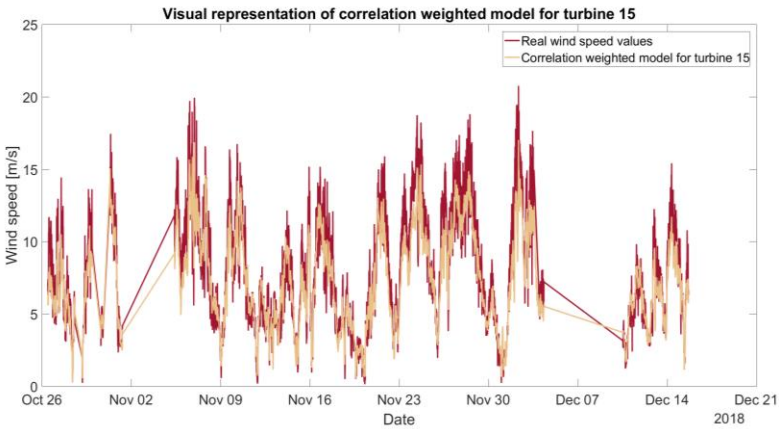


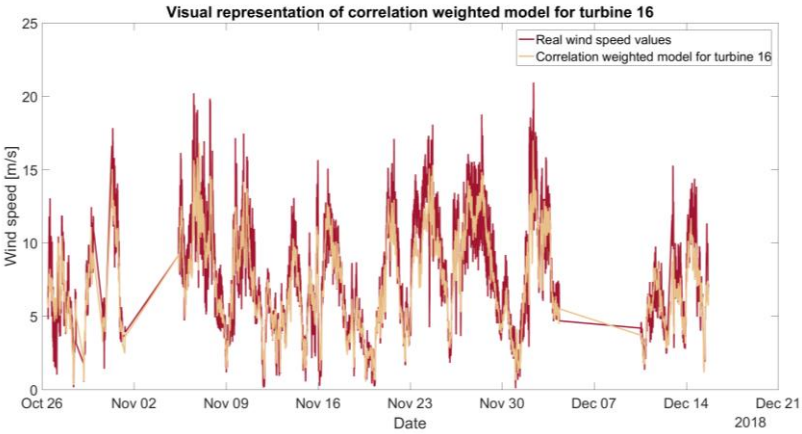


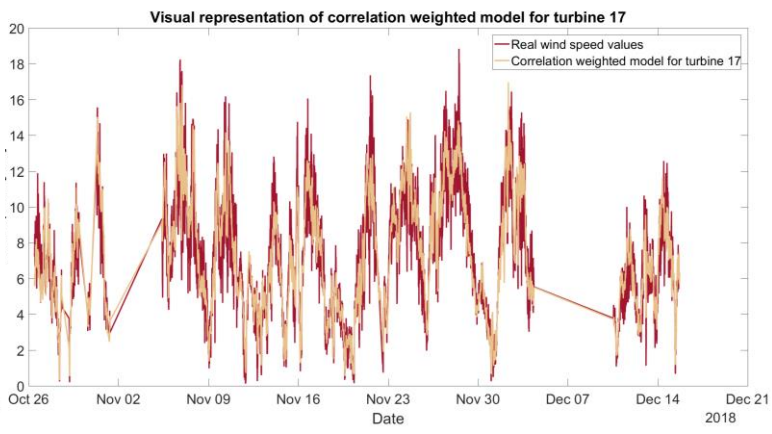


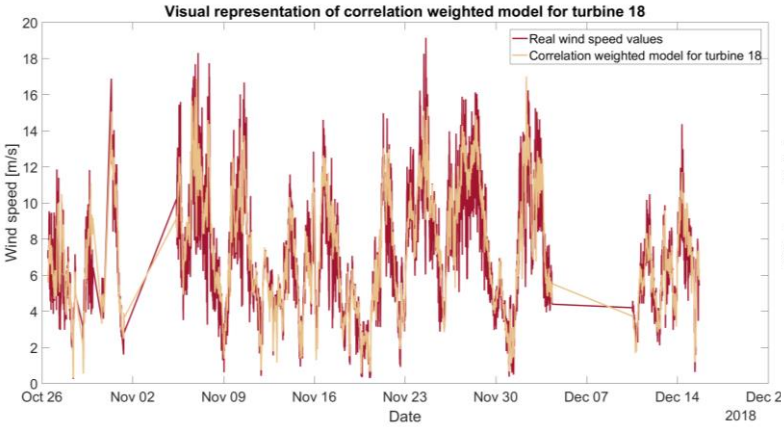


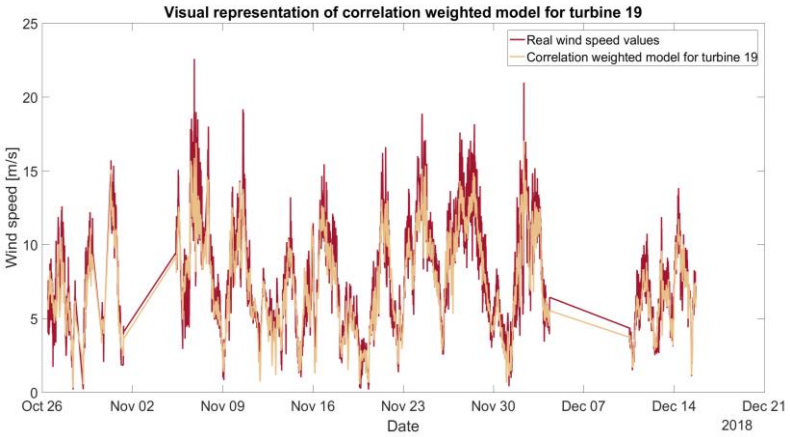


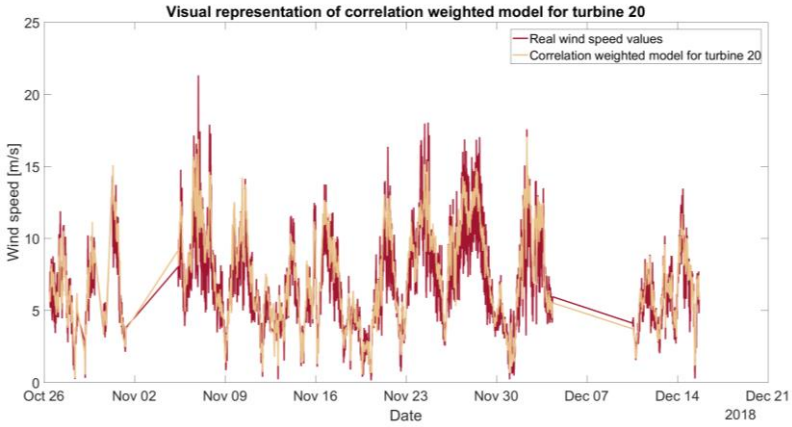




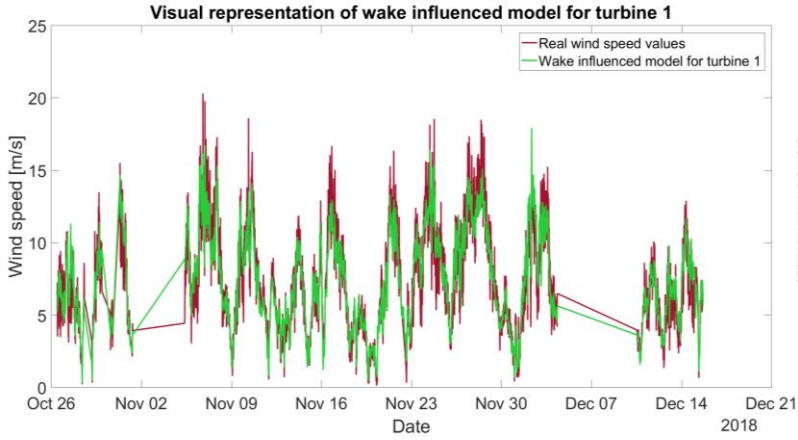


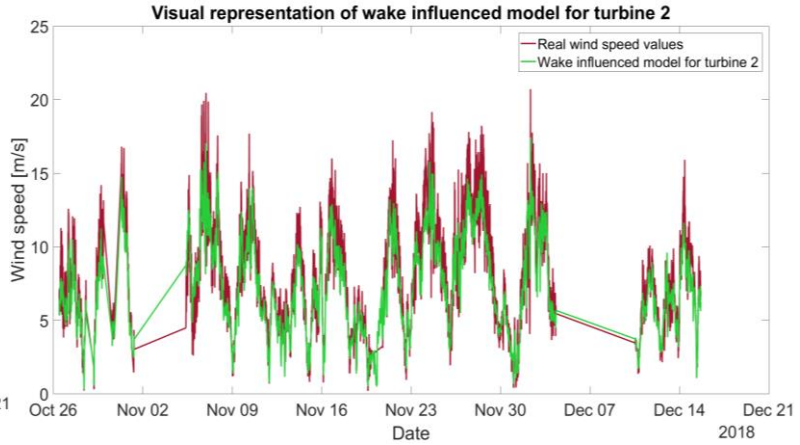


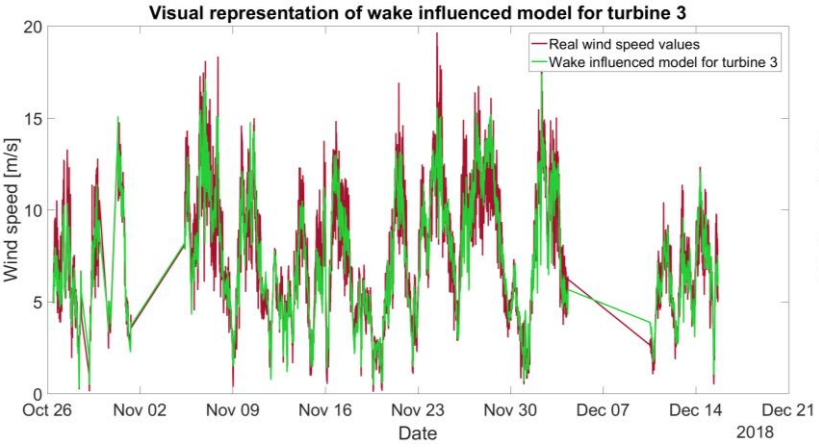


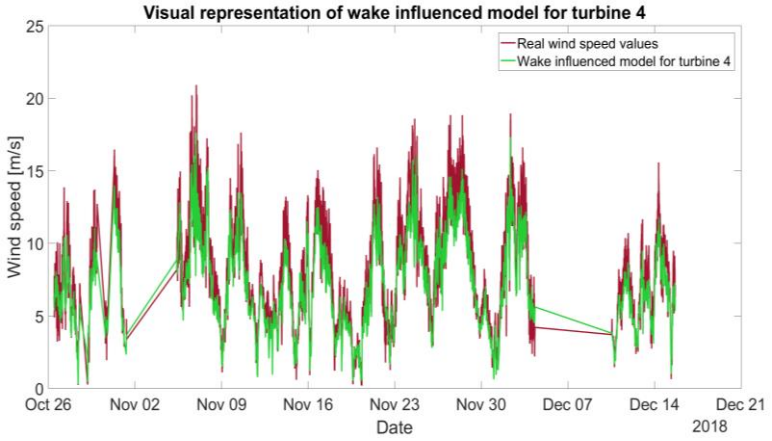


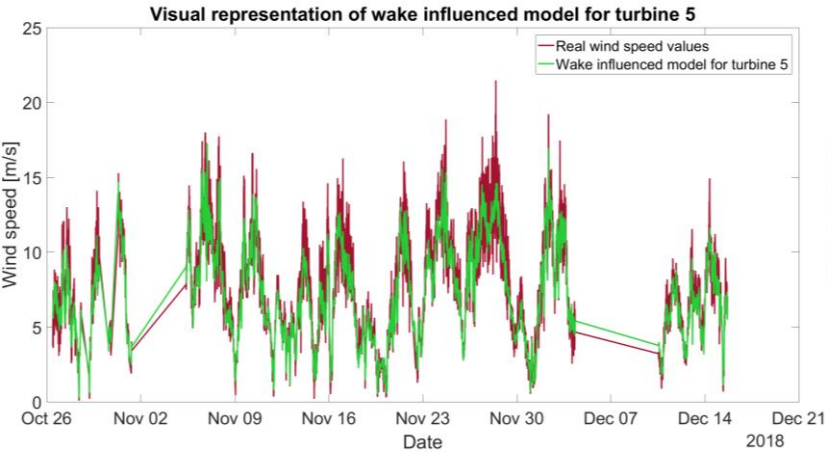
Wake influence model

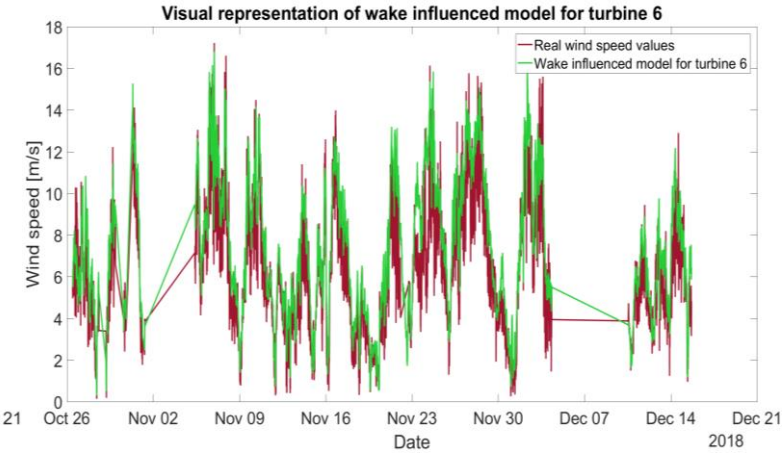


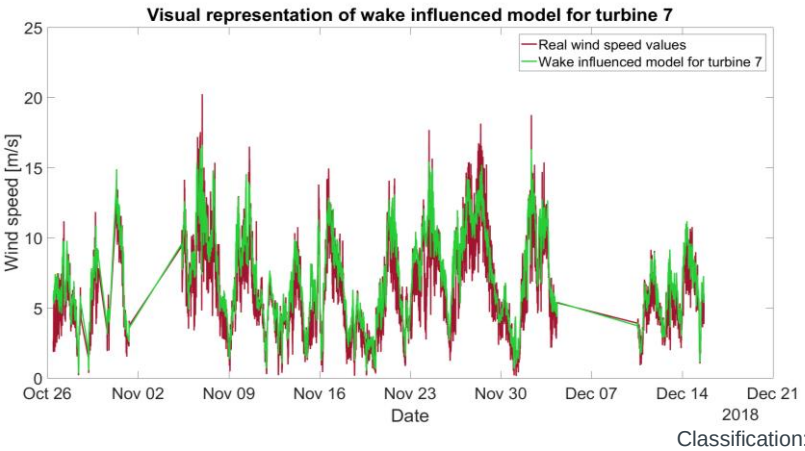


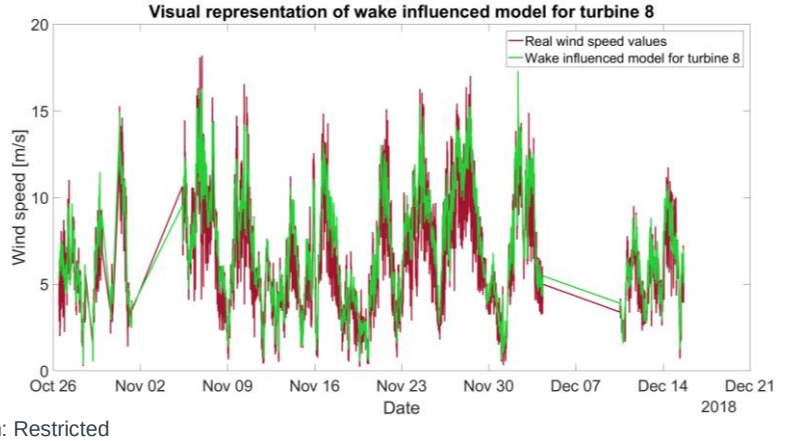


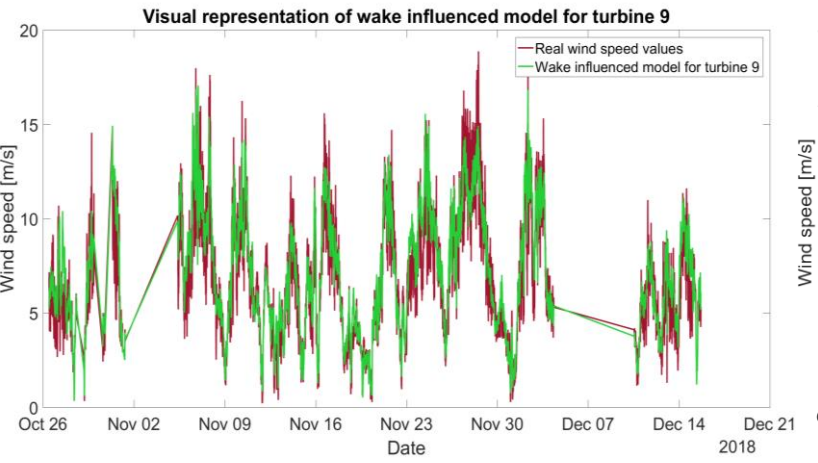


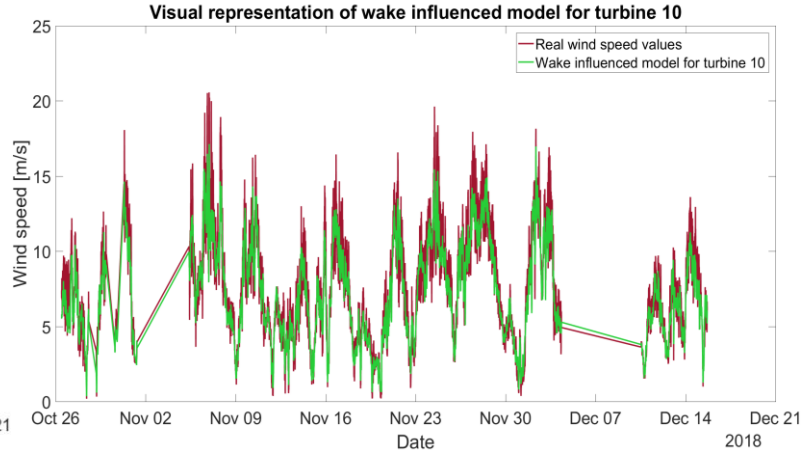


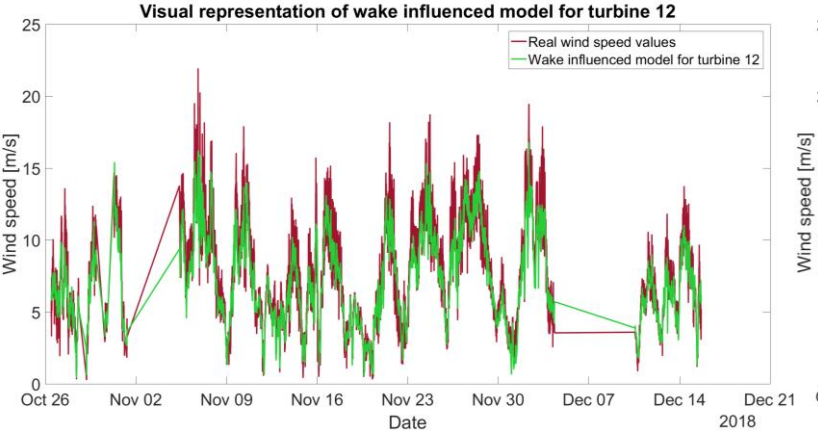


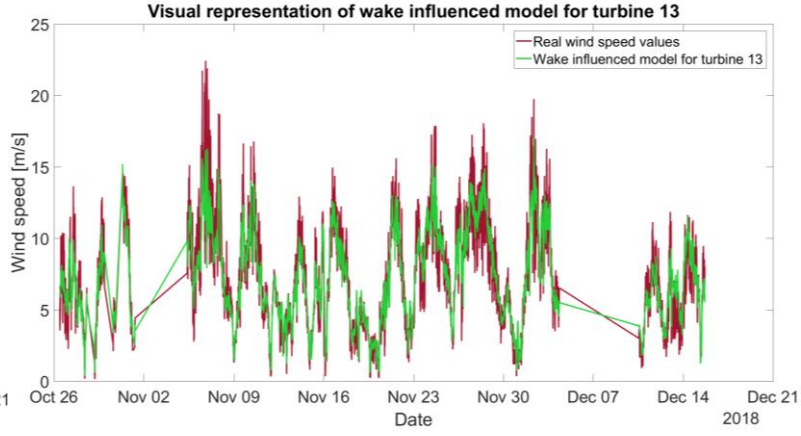


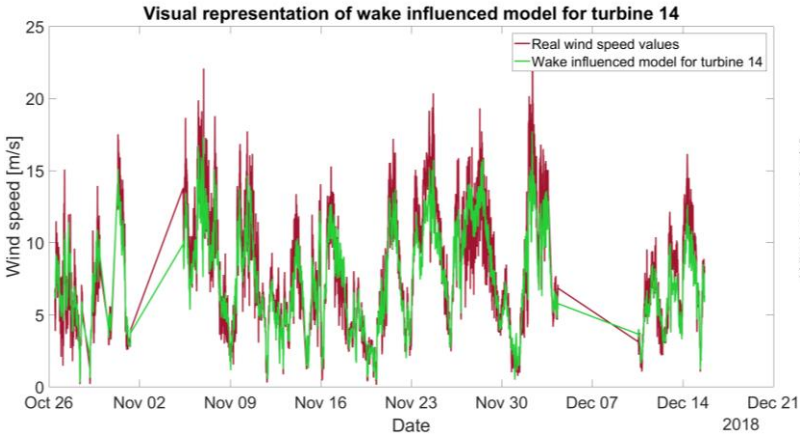


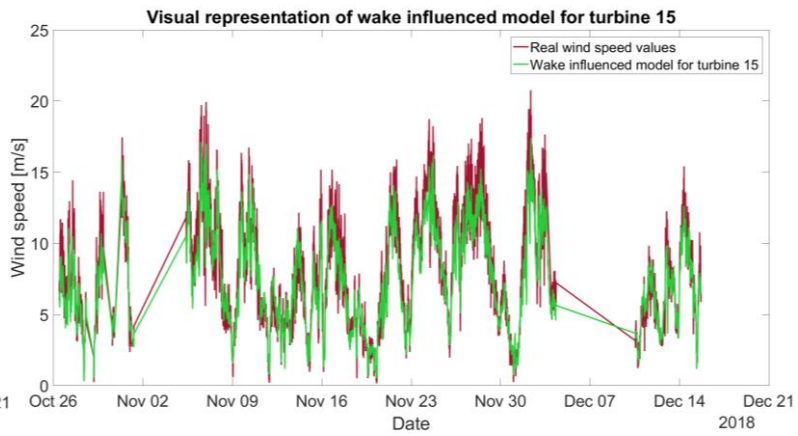

Visual representation of wake influenced model for turbine 9
Real wind speed values
Wake influenced model for turbine 9
Wind speed [m/s]
0
5
10
15
20
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
2018
Date

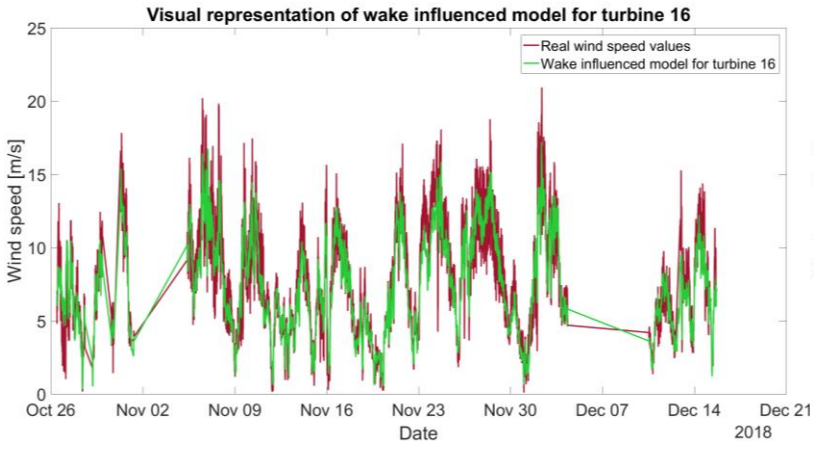

Visual representation of wake influenced model for turbine 10
Real wind speed values
Wake influenced model for turbine 10
Wind speed [m/s]
0
5
10
15
20
25
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
2018
Date

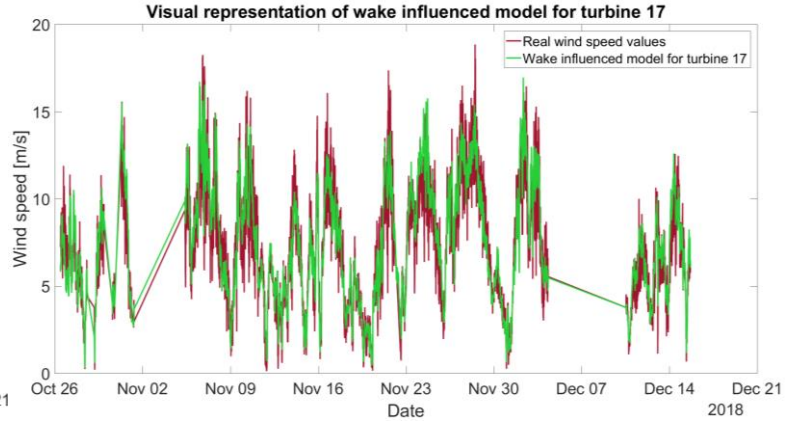

Visual representation of wake influenced model for turbine 12
Real wind speed values
Wake influenced model for turbine 12
Wind speed [m/s]
0
5
10
15
20
25
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
2018
Date

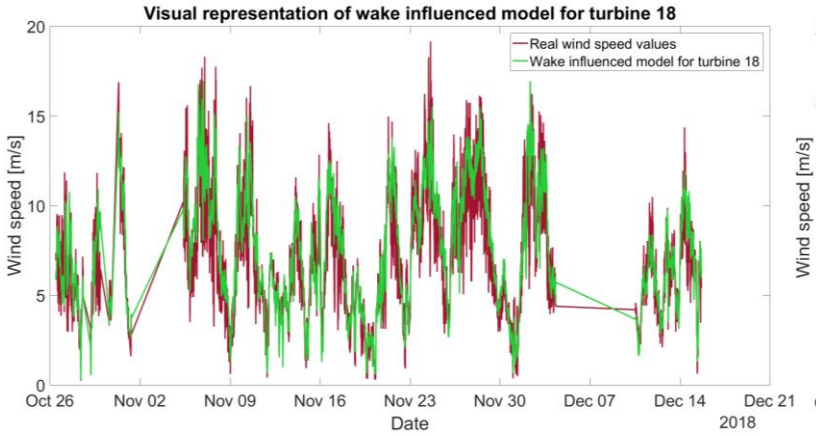

Visual representation of wake influenced model for turbine 13
Real wind speed values
Wake influenced model for turbine 13
Wind speed [m/s]
0
5
10
15
20
25
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
2018
Date

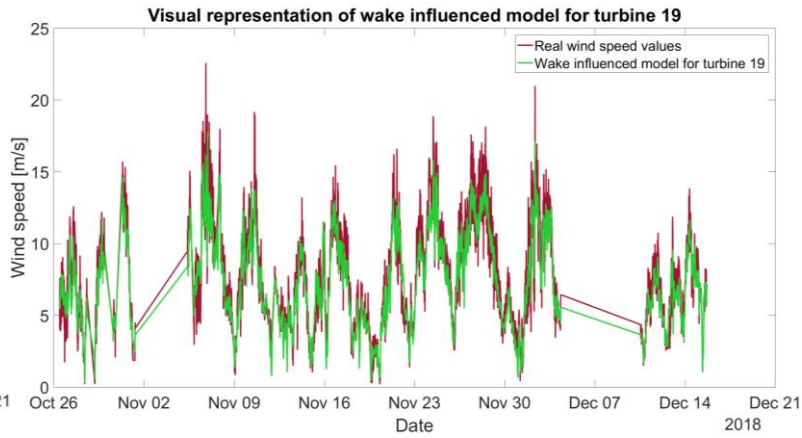

Visual representation of wake influenced model for turbine 14
Real wind speed values
Wake influenced model for turbine 14
Wind speed [m/s]
0
5
10
15
20
25
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
2018
Date

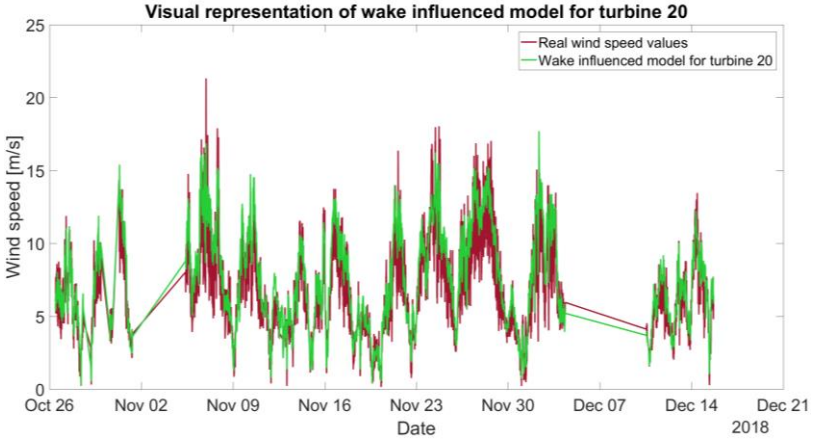

Visual representation of wake influenced model for turbine 15
Real wind speed values
Wake influenced model for turbine 15
Wind speed [m/s]
0
5
10
15
20
25
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
2018
Date


Visual representation of wake influenced model for turbine 16
Real wind speed values
Wake influenced model for turbine 16
Wind speed [m/s]
0
5
10
15
20
25
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
2018
Date

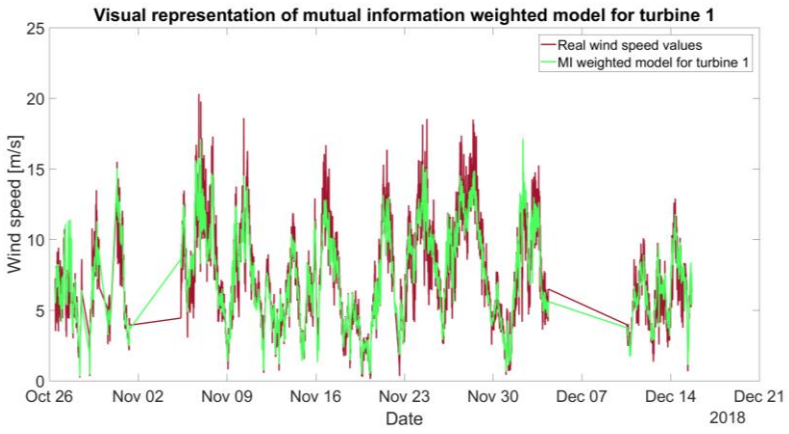

Visual representation of wake influenced model for turbine 17
Real wind speed values
Wake influenced model for turbine 17
Wind speed [m/s]
0
5
10
15
20
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
2018
Date

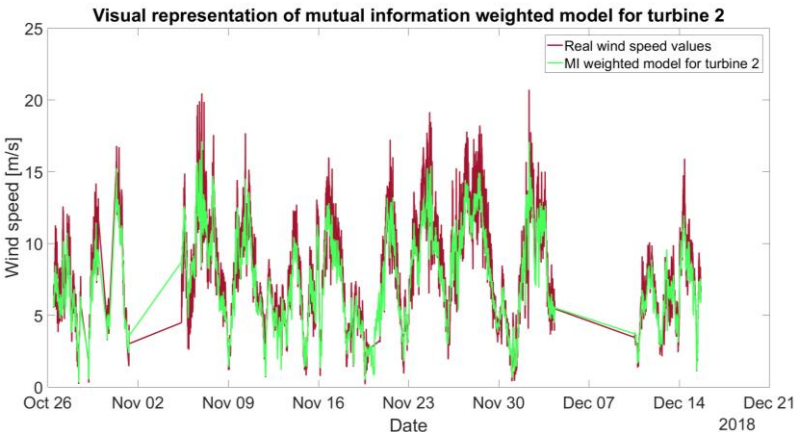


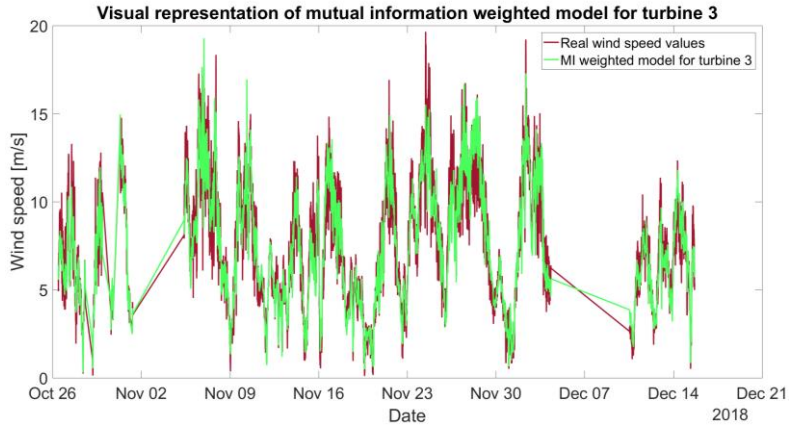


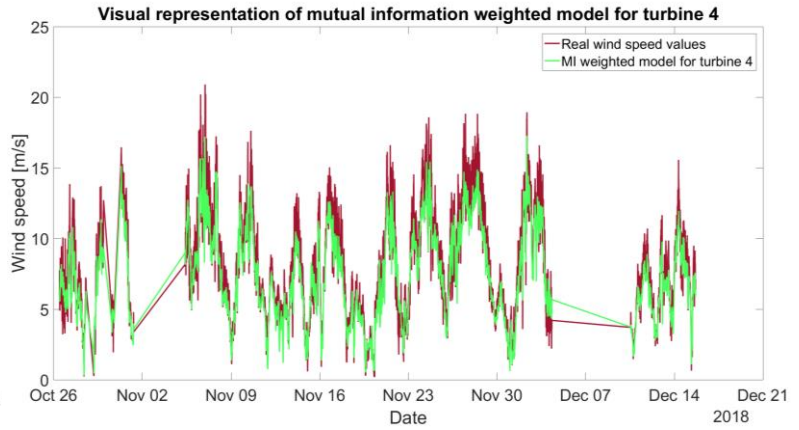


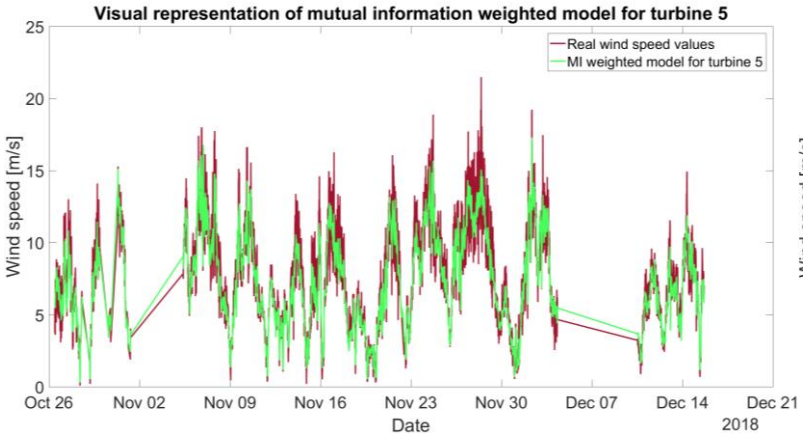
Mutual information weighted model

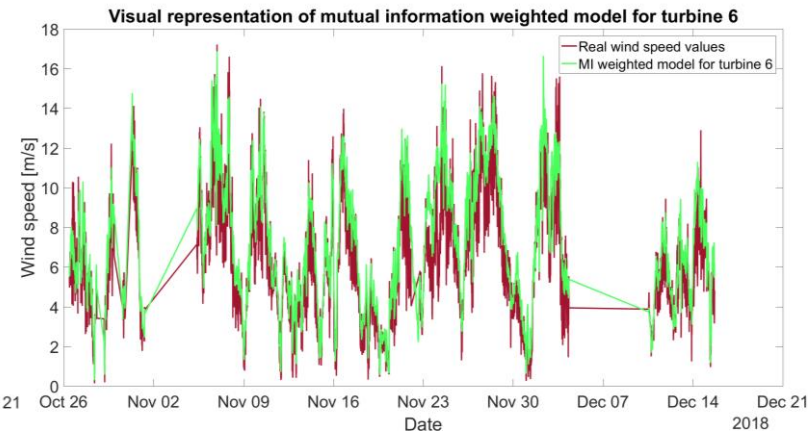


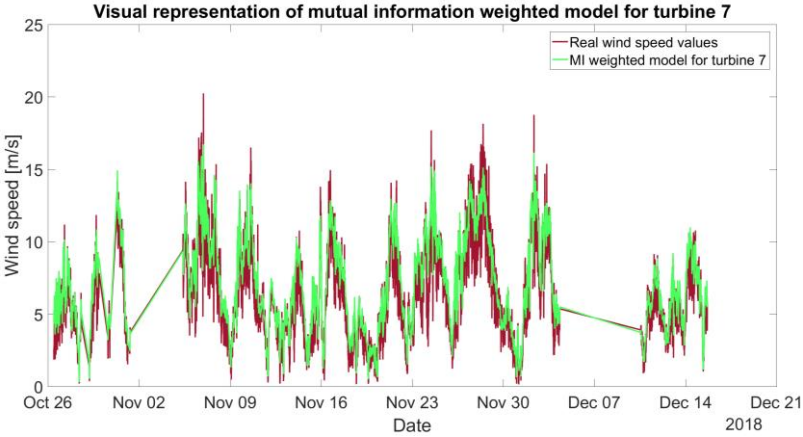


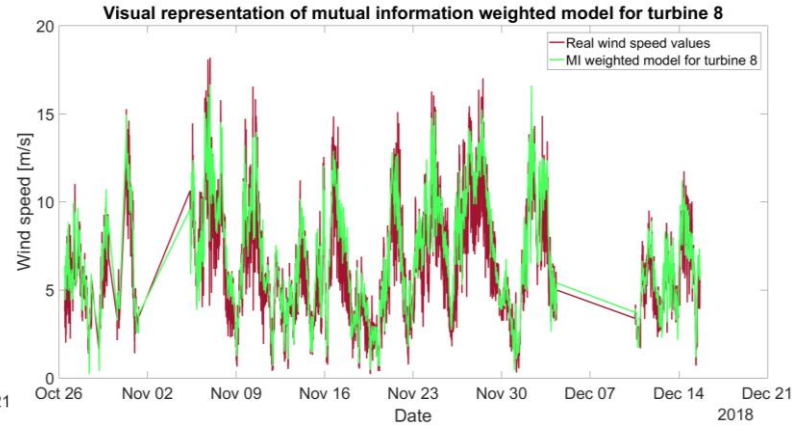


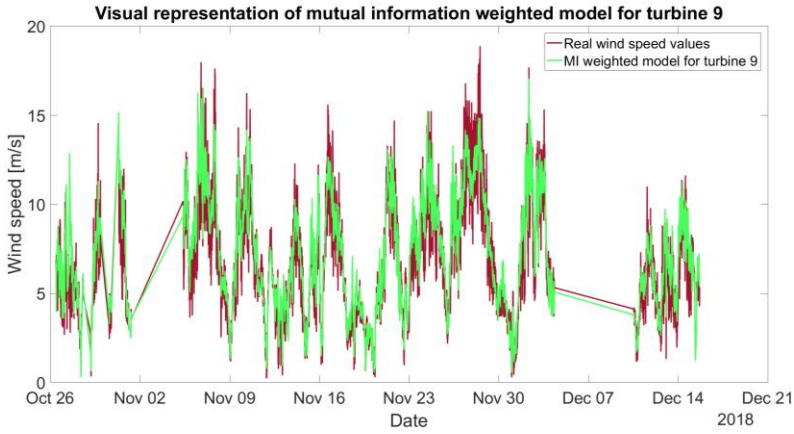


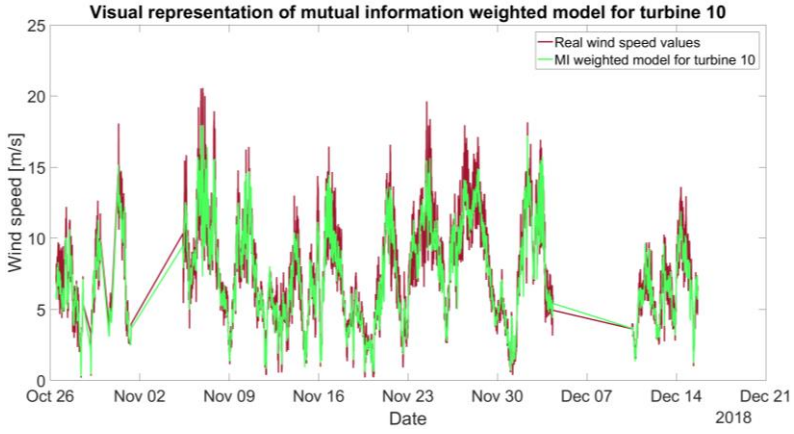


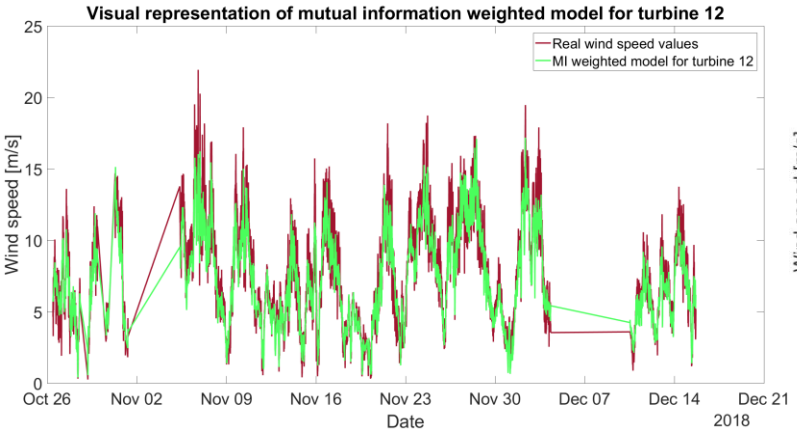


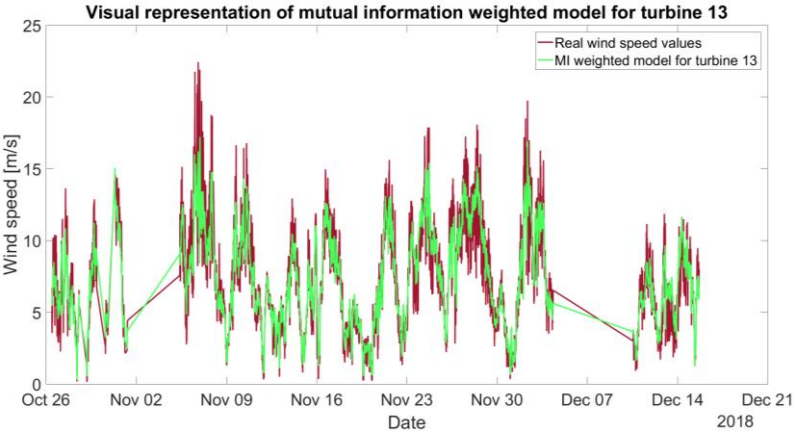


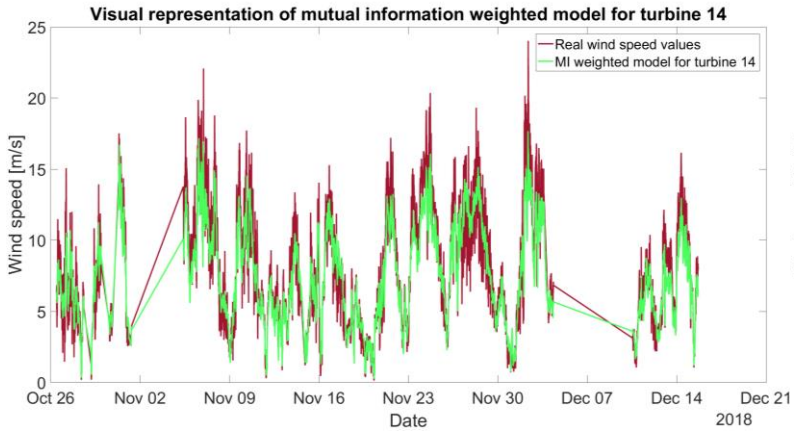


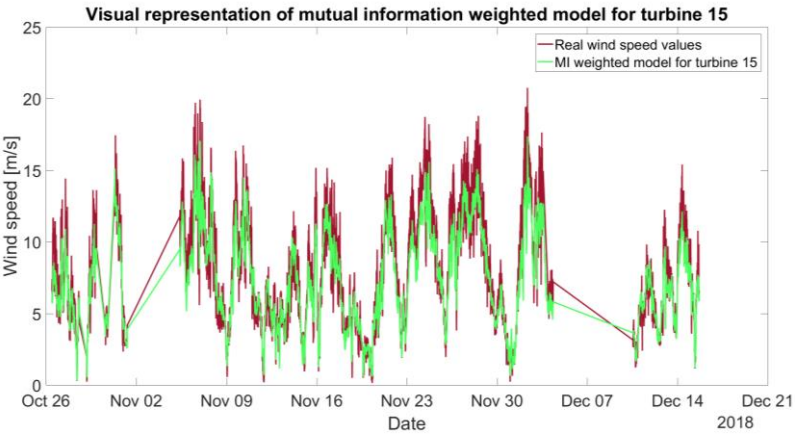


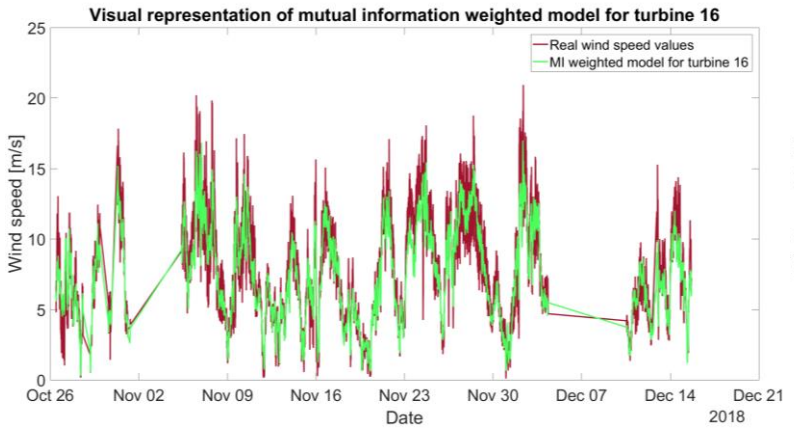


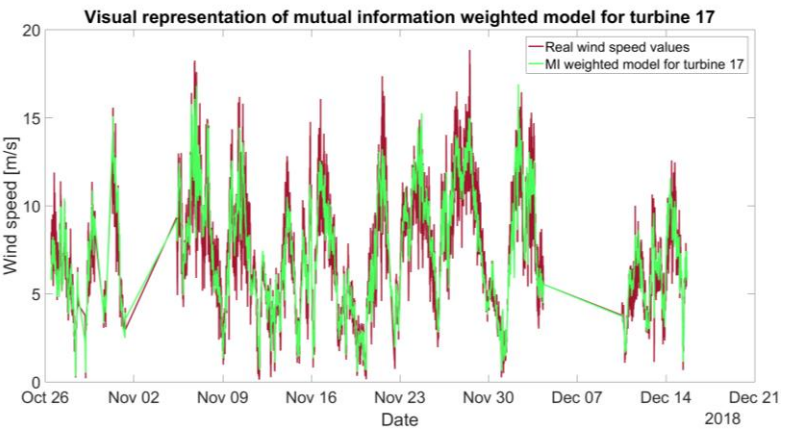


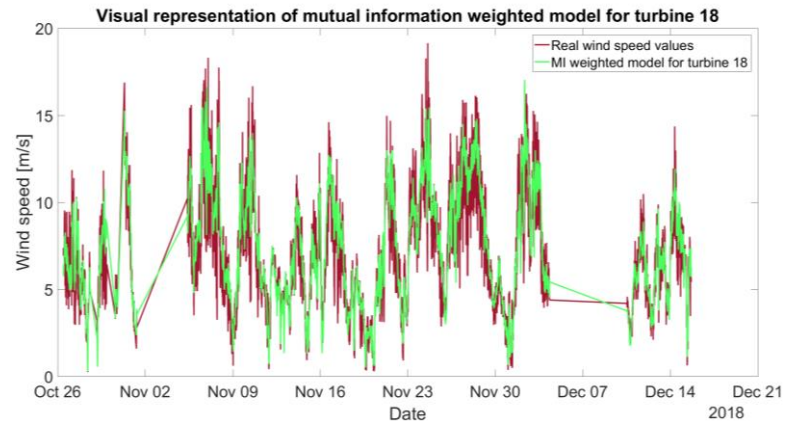


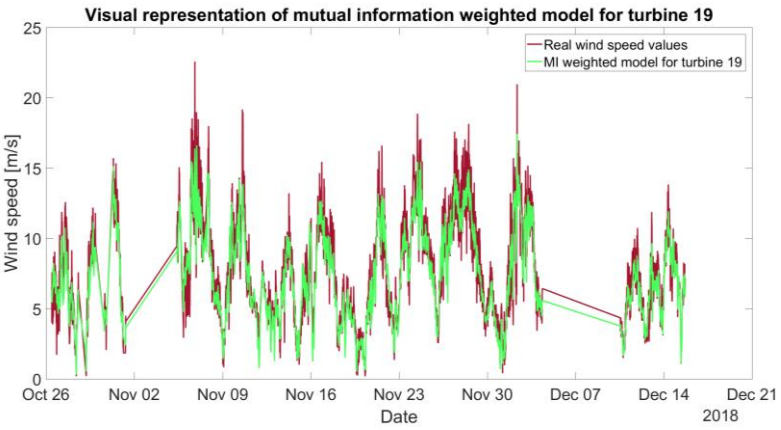


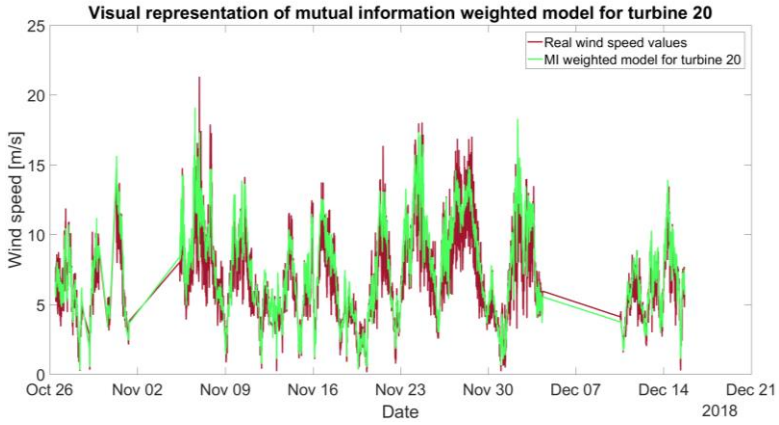


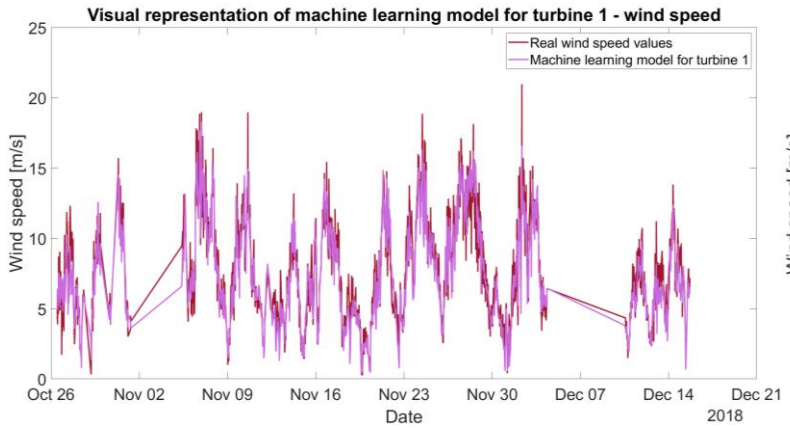


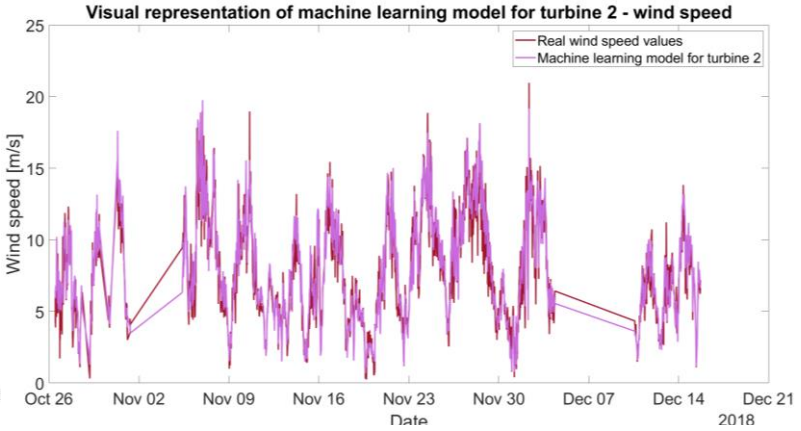


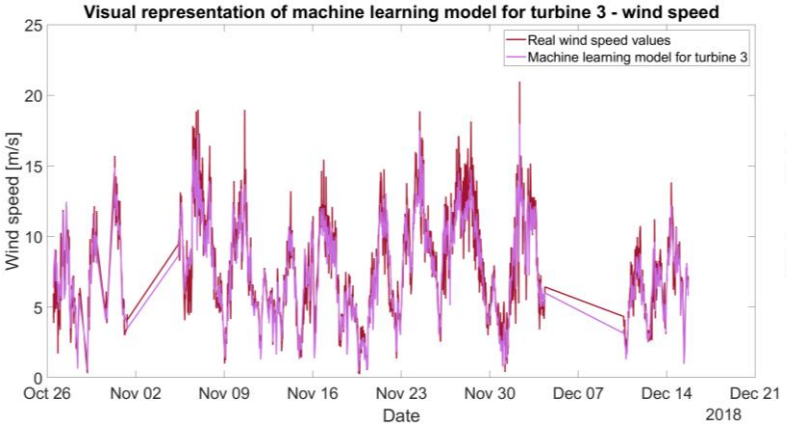


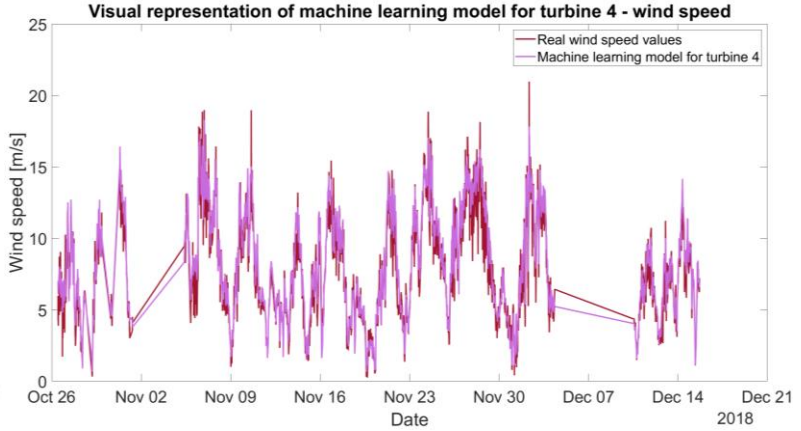


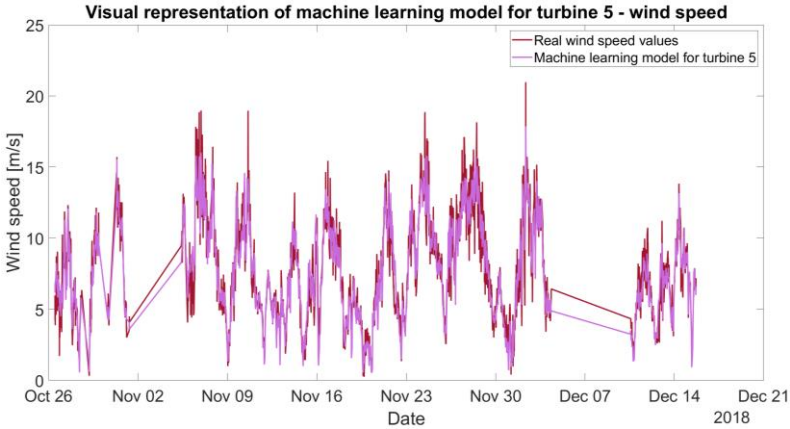


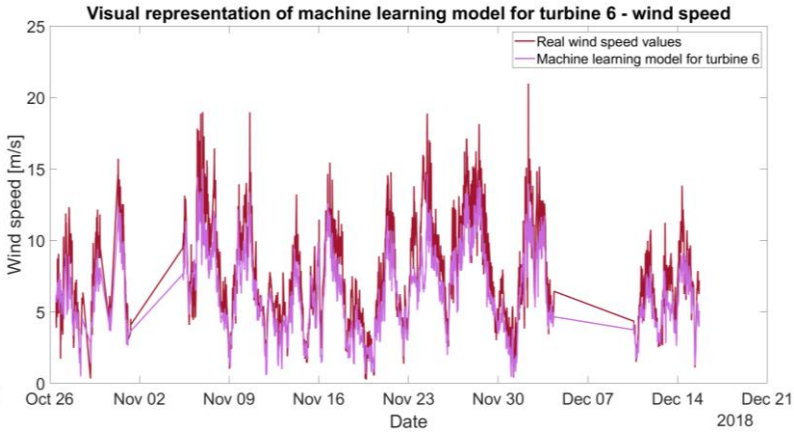

Visual representation of machine learning model for turbine 1 - wind speed
Real wind speed values
Machine learning model for turbine 1
Wind speed [m/s]
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
Date
2018

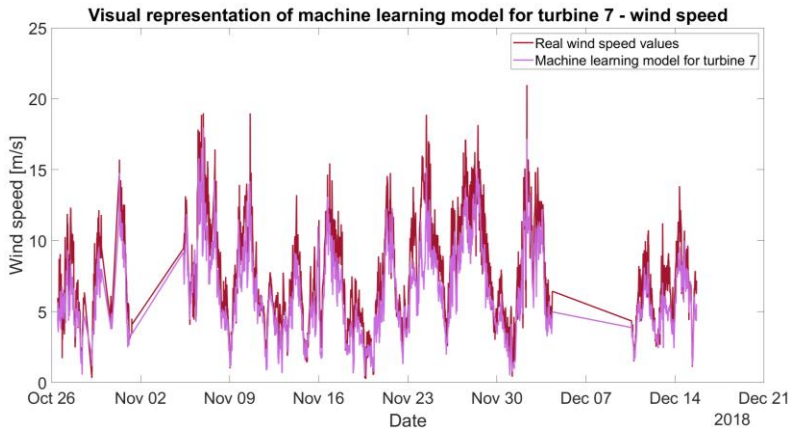

Visual representation of machine learning model for turbine 2 - wind speed
Real wind speed values
Machine learning model for turbine 2
Wind speed [m/s]
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
Date
2018

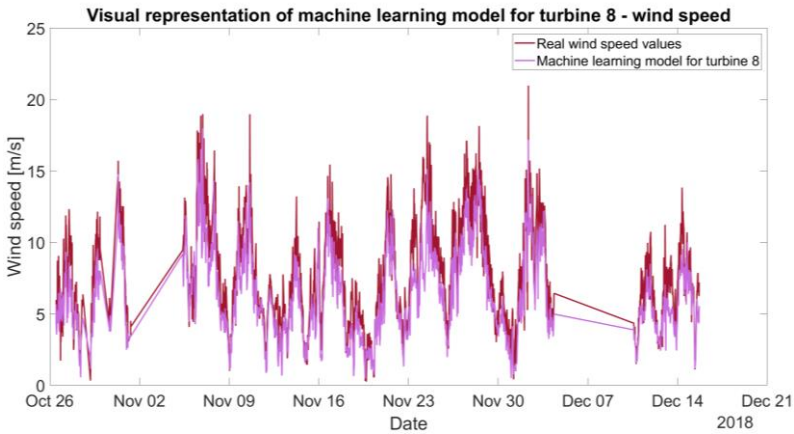

Visual representation of machine learning model for turbine 3 - wind speed
Real wind speed values
Machine learning model for turbine 3
Wind speed [m/s]
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
Date
2018

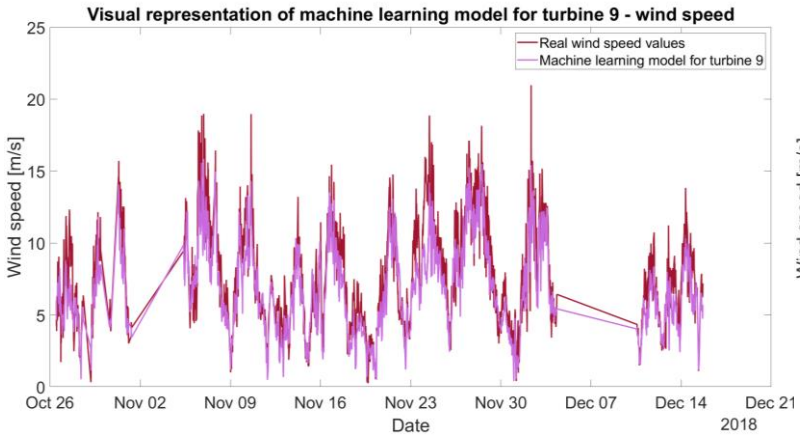

Visual representation of machine learning model for turbine 4 - wind speed
Real wind speed values
Machine learning model for turbine 4
Wind speed [m/s]
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
Date
2018

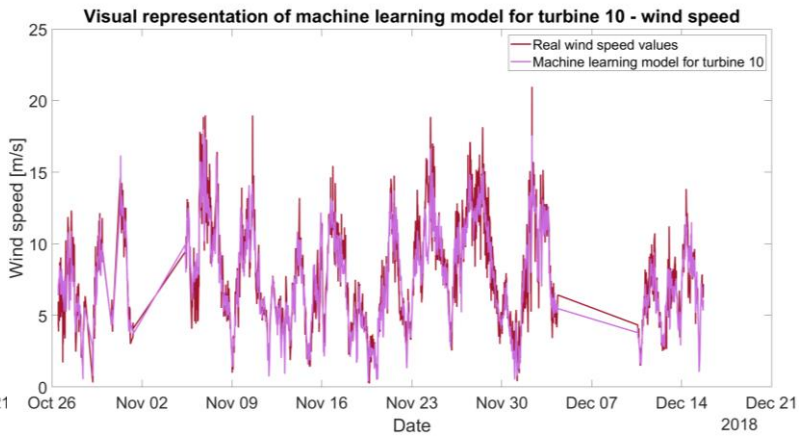

Visual representation of machine learning model for turbine 5 - wind speed
Real wind speed values
Machine learning model for turbine 5
Wind speed [m/s]
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
Date
2018

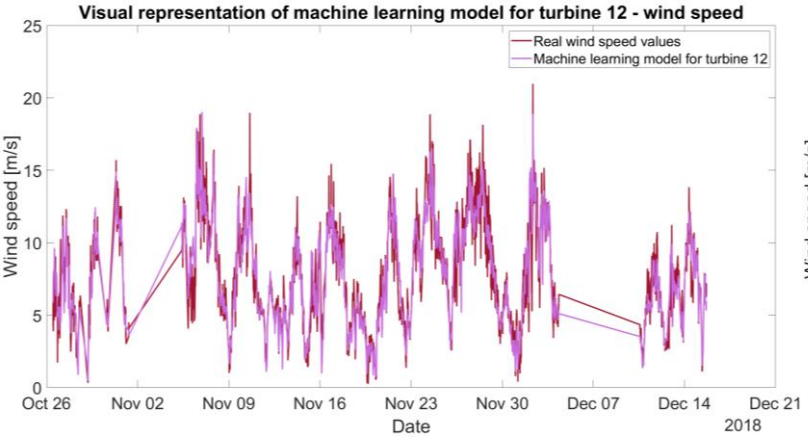

Visual representation of machine learning model for turbine 6 - wind speed
Real wind speed values
Machine learning model for turbine 6
Wind speed [m/s]
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
Date
2018

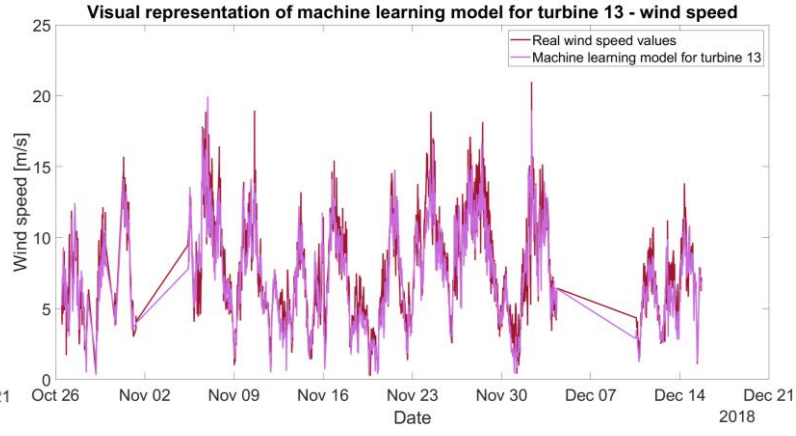

Visual representation of machine learning model for turbine 7 - wind speed
Real wind speed values
Machine learning model for turbine 7
Wind speed [m/s]
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
Date
2018

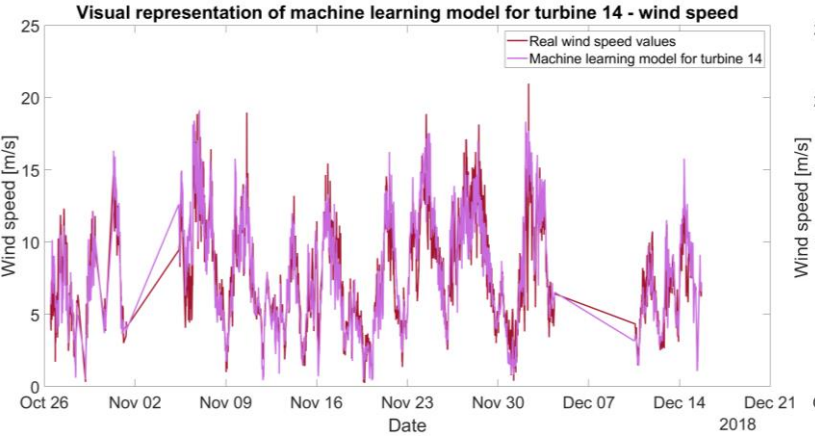

Visual representation of machine learning model for turbine 8 - wind speed
Real wind speed values
Machine learning model for turbine 8
Wind speed [m/s]
Oct 26
Nov 02
Nov 09
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
Date
2018

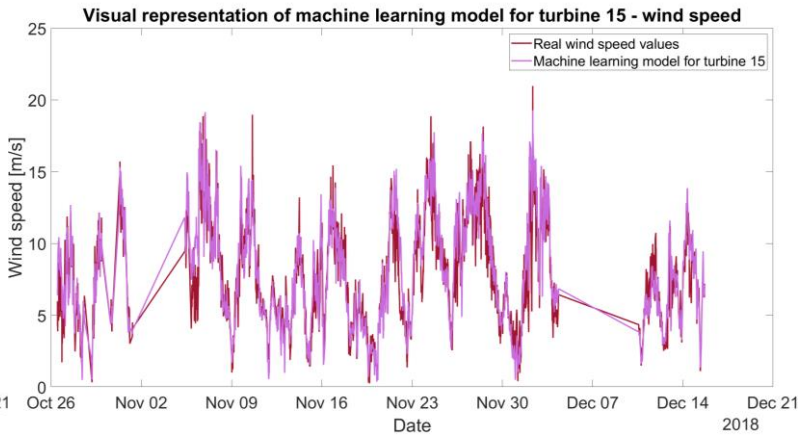


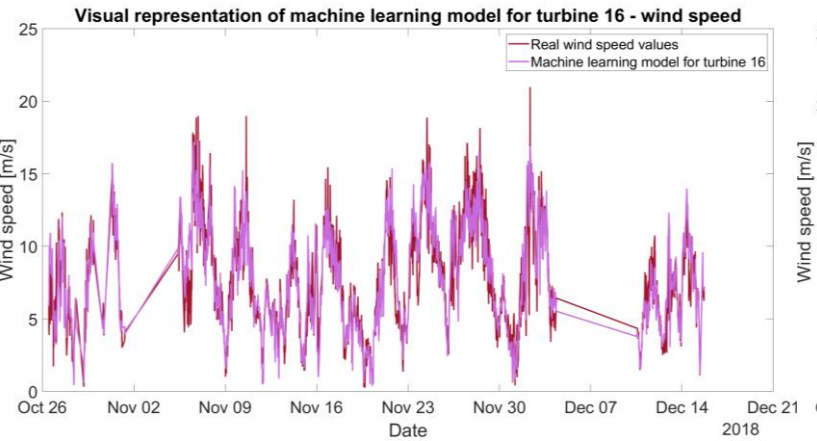


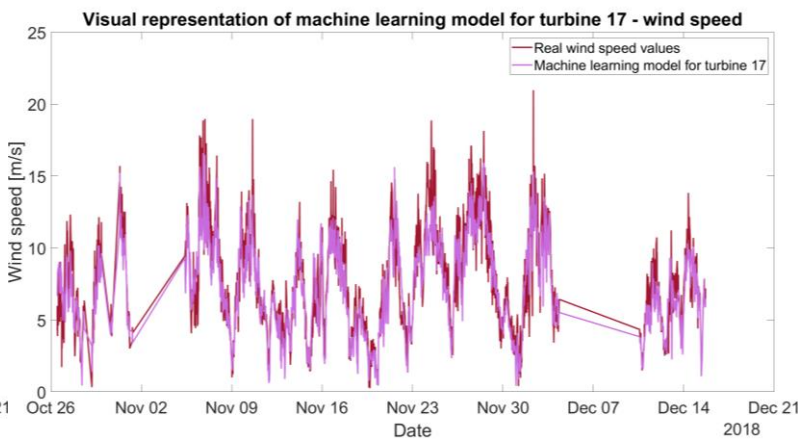


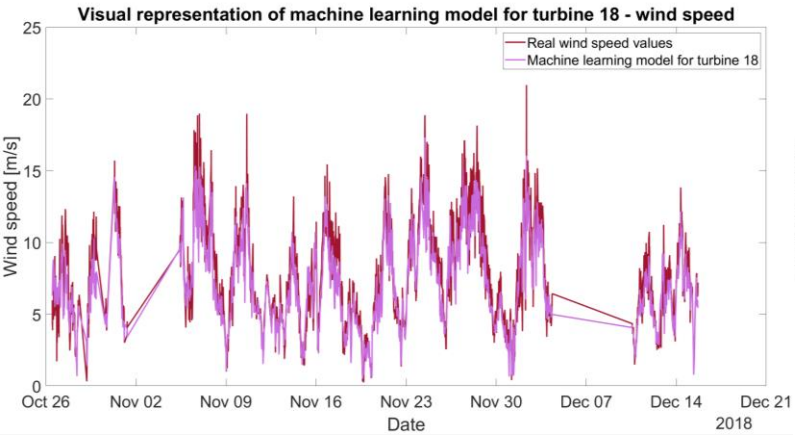


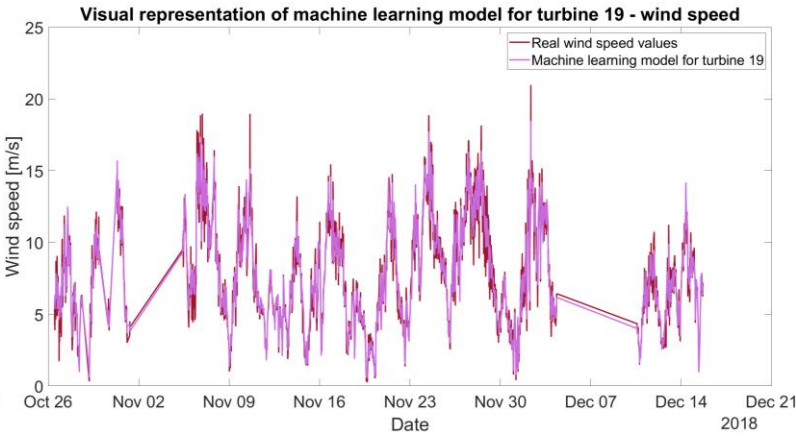


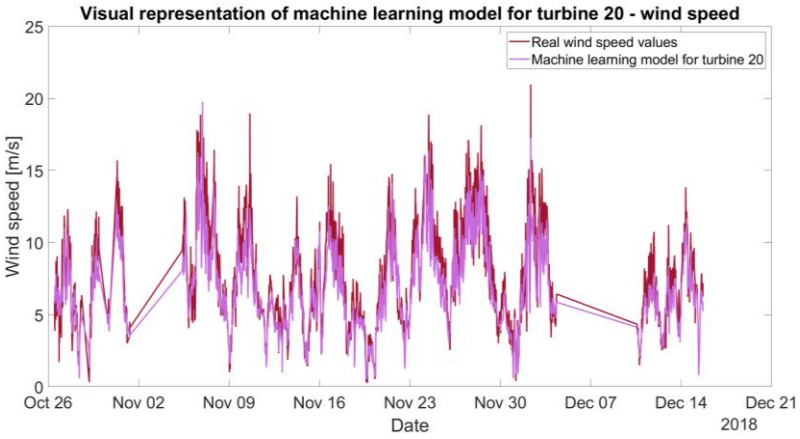




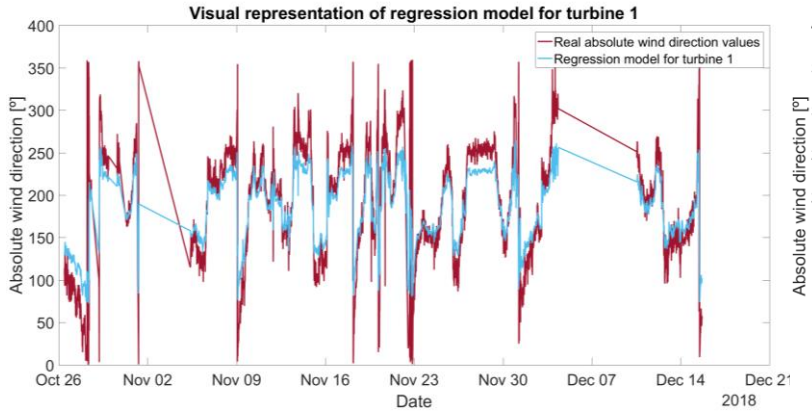


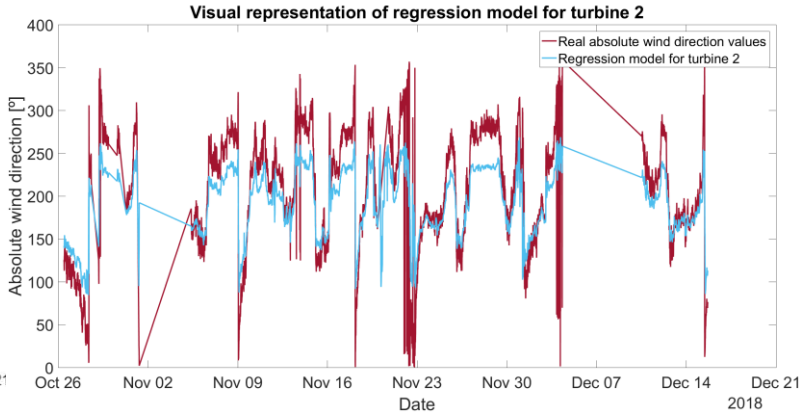


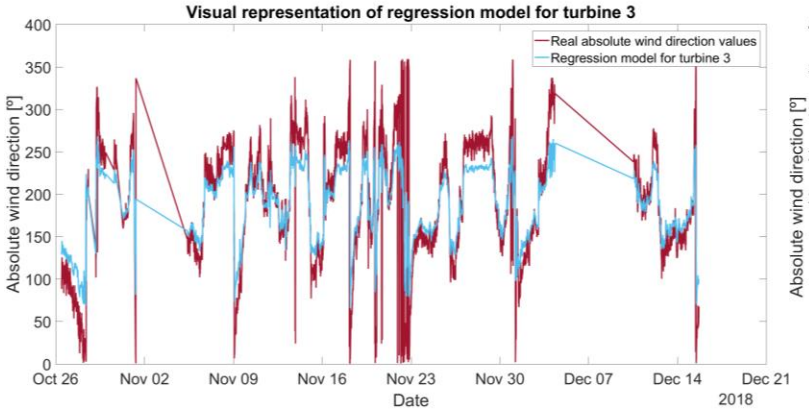


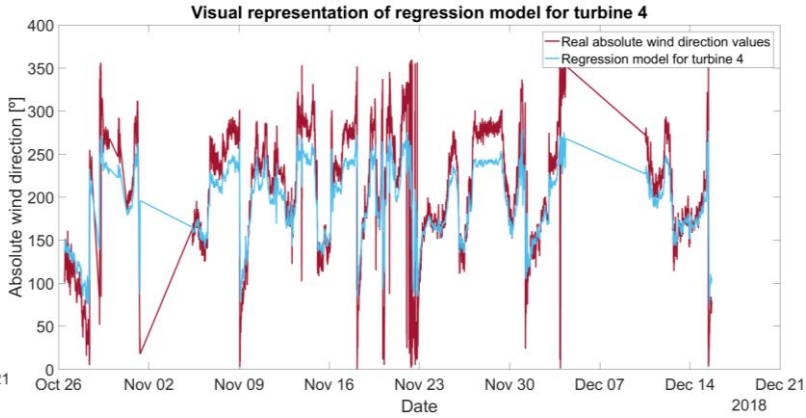


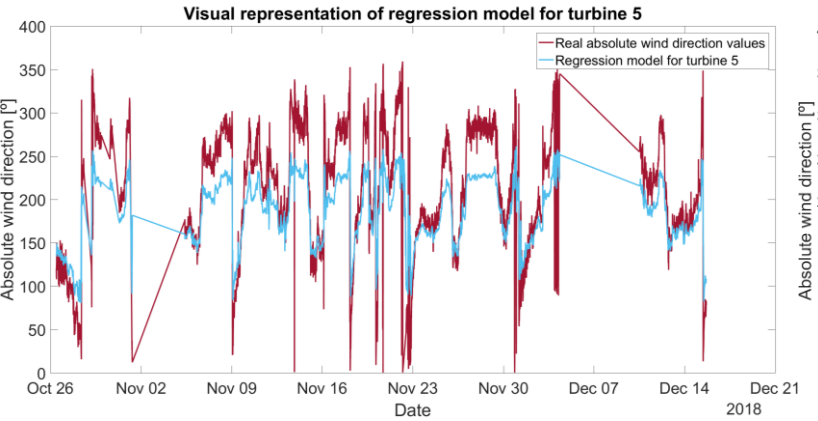
Absolute wind direction: Regression model

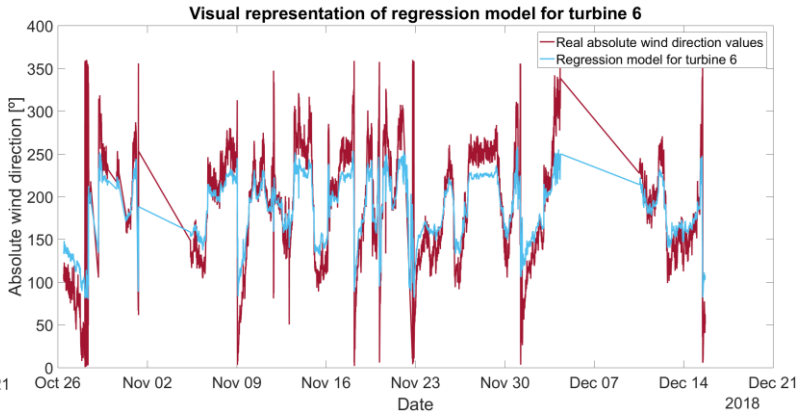


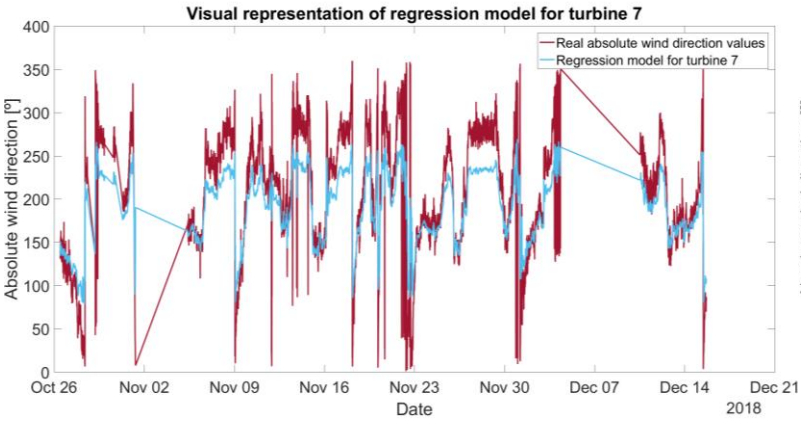


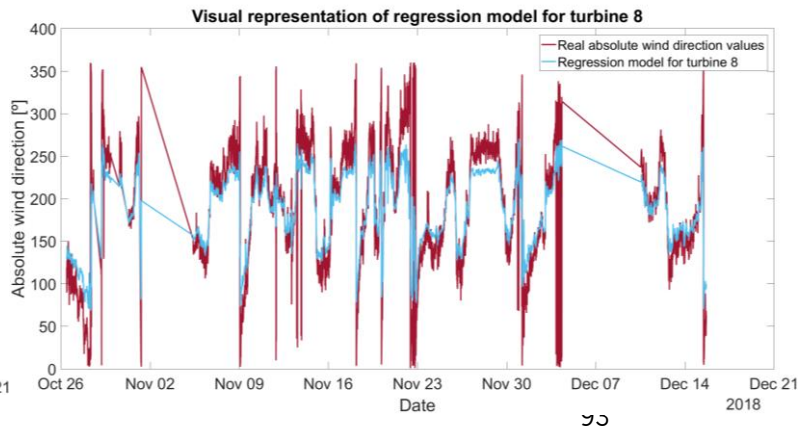


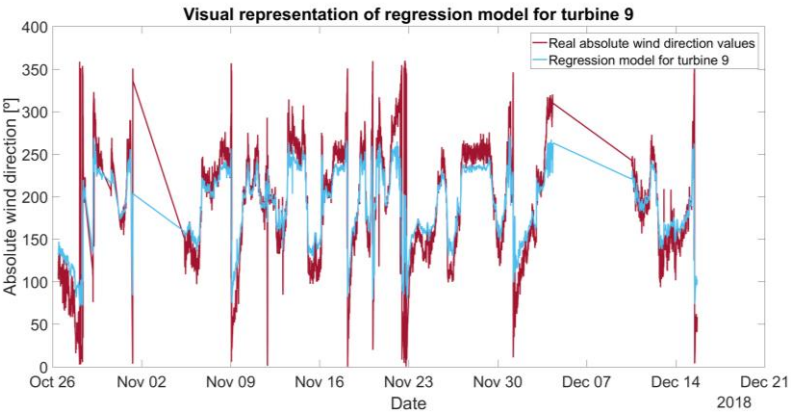


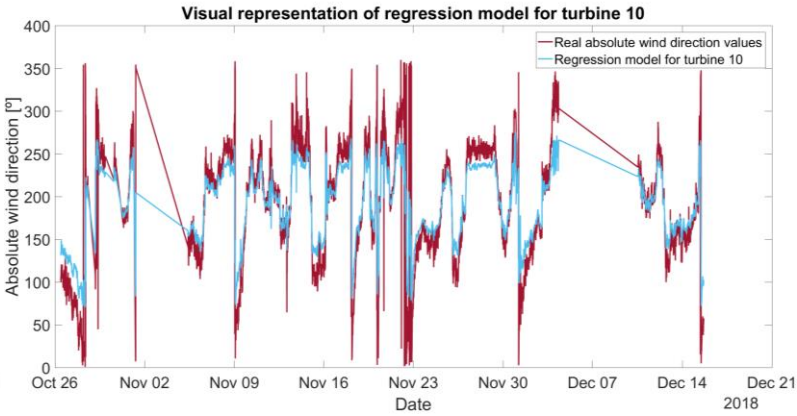


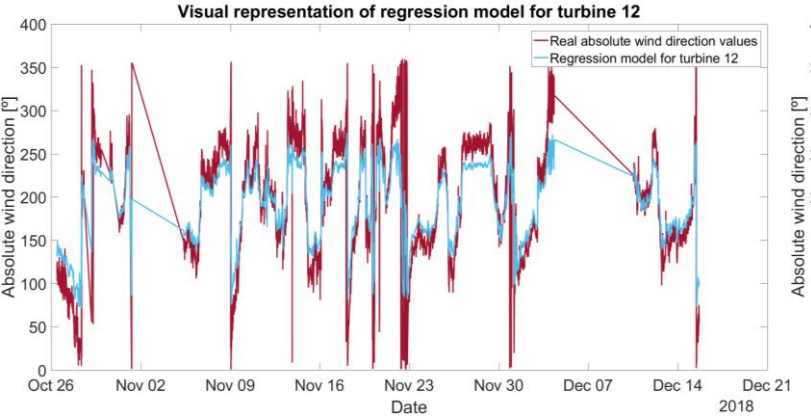


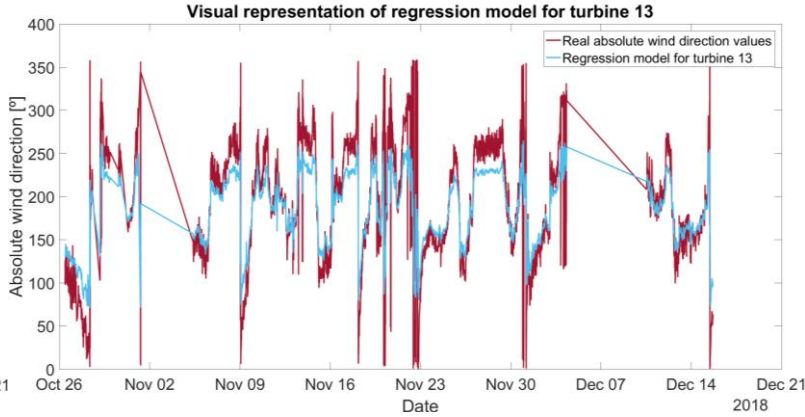


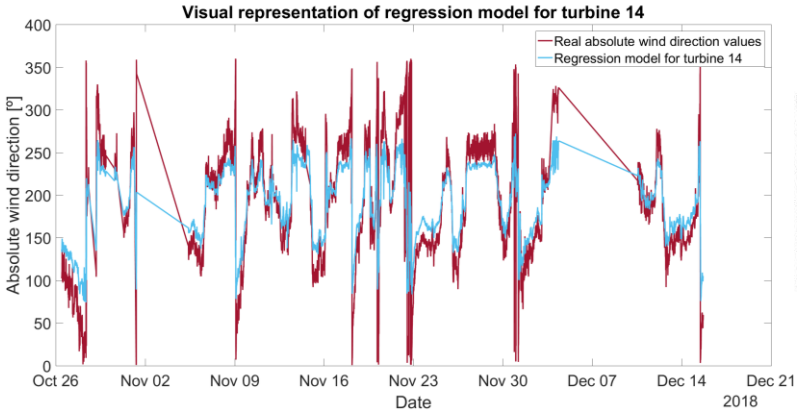


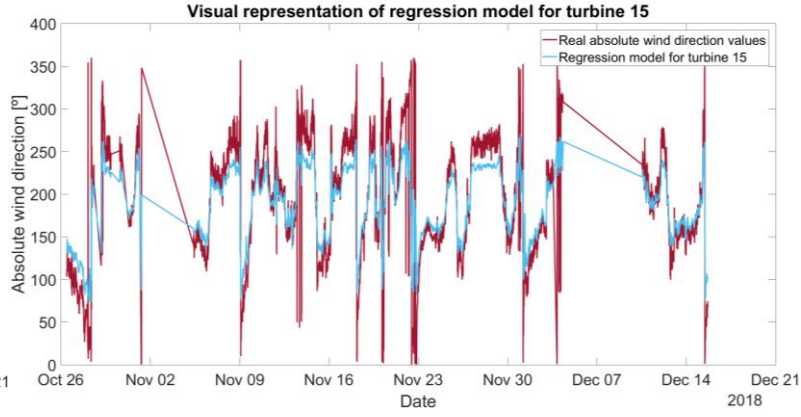


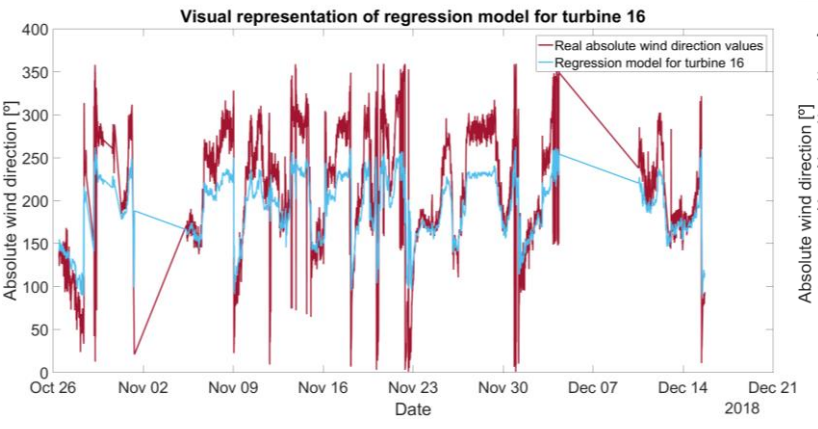


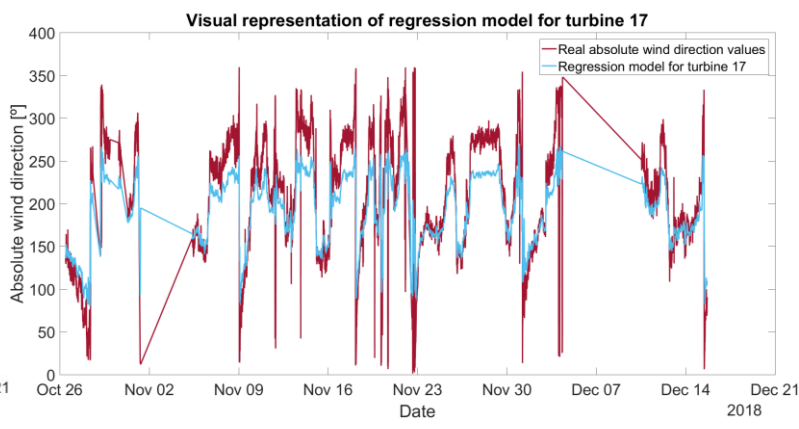


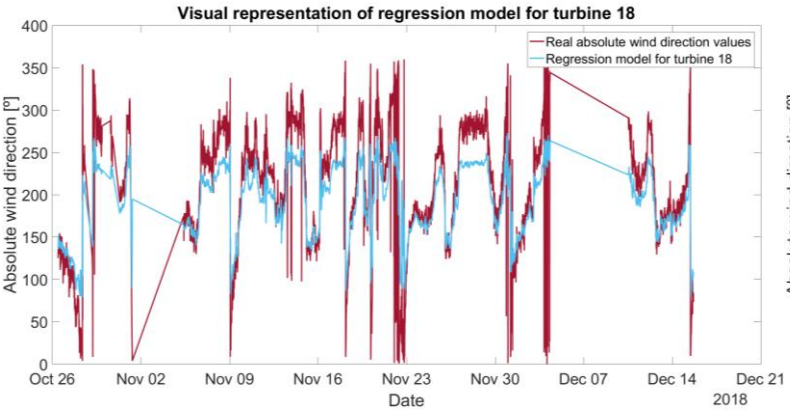


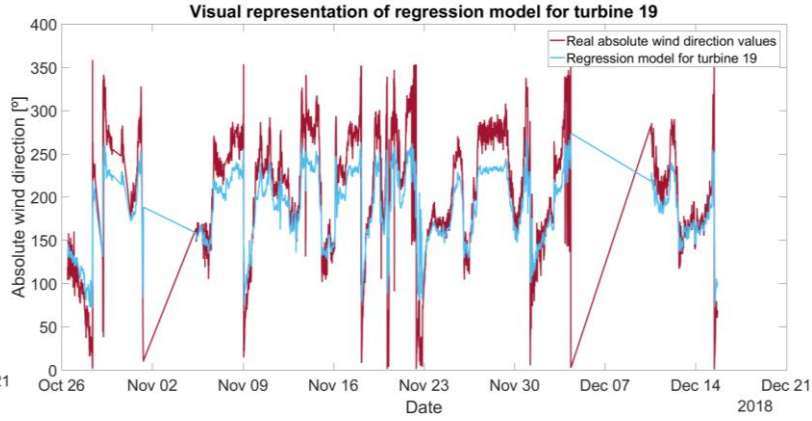


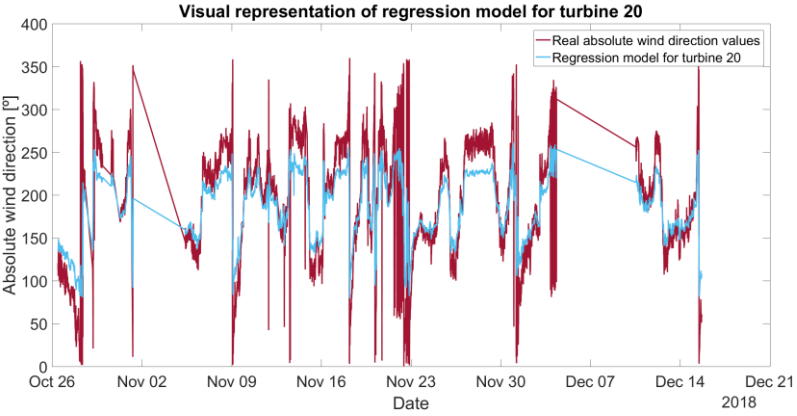




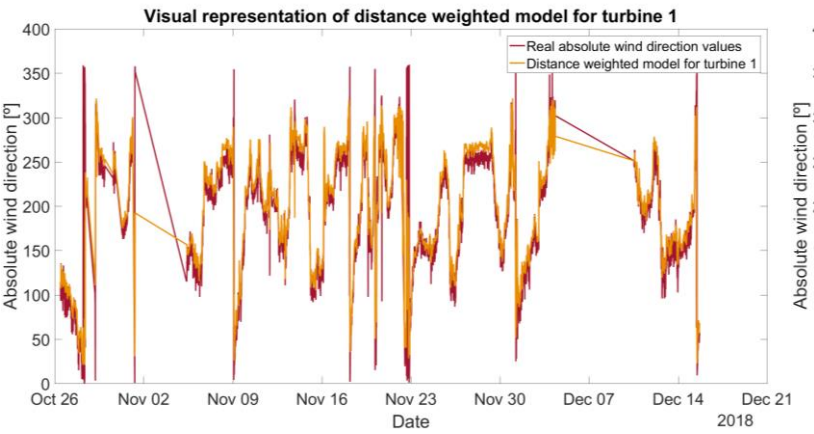


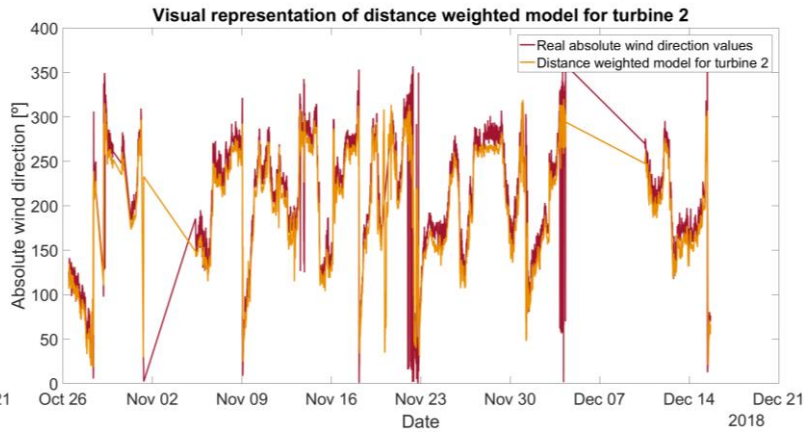


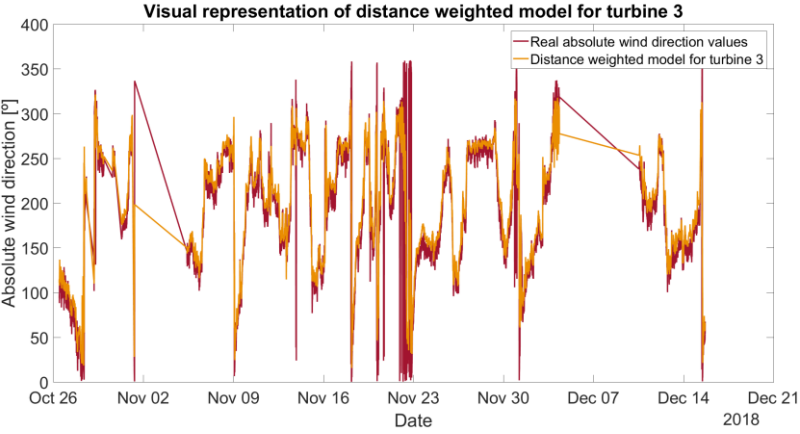


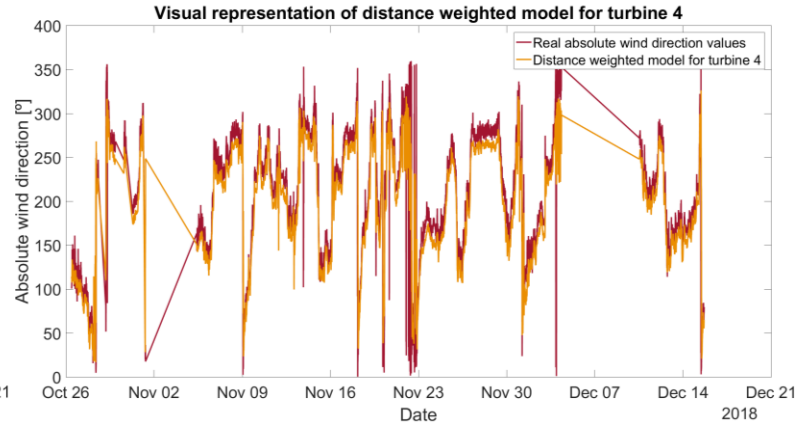


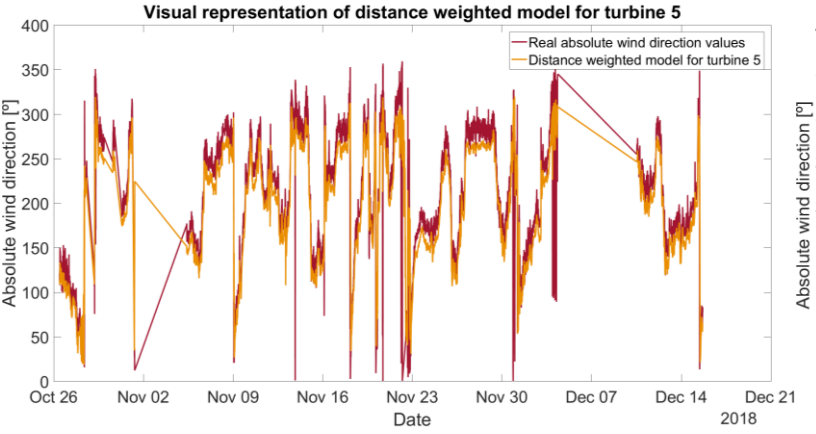
Distance weighted model

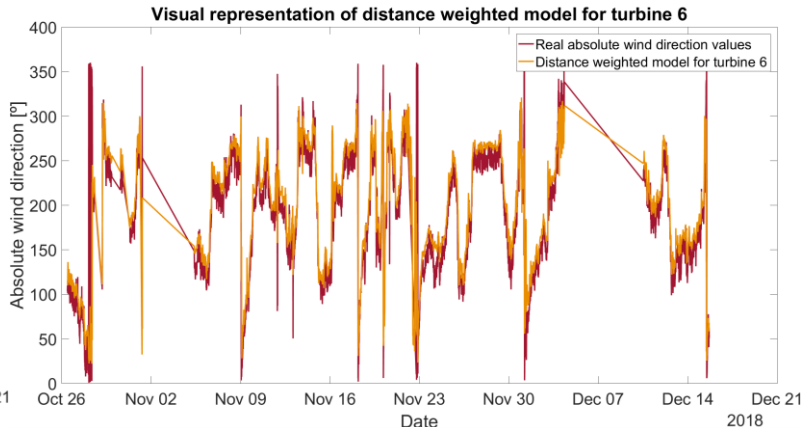


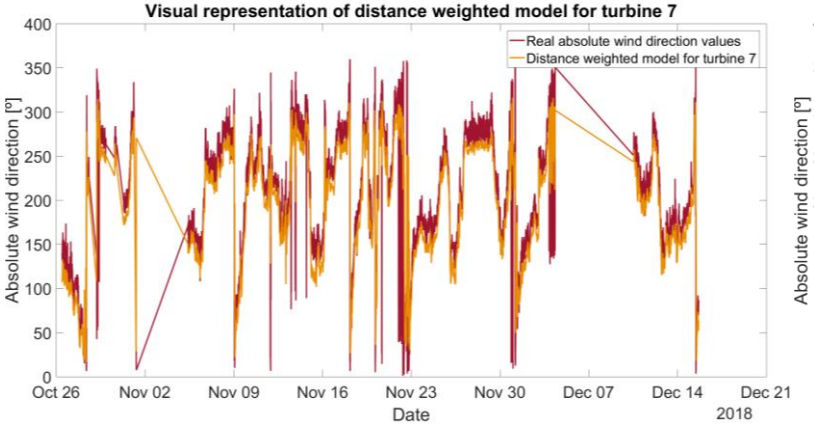


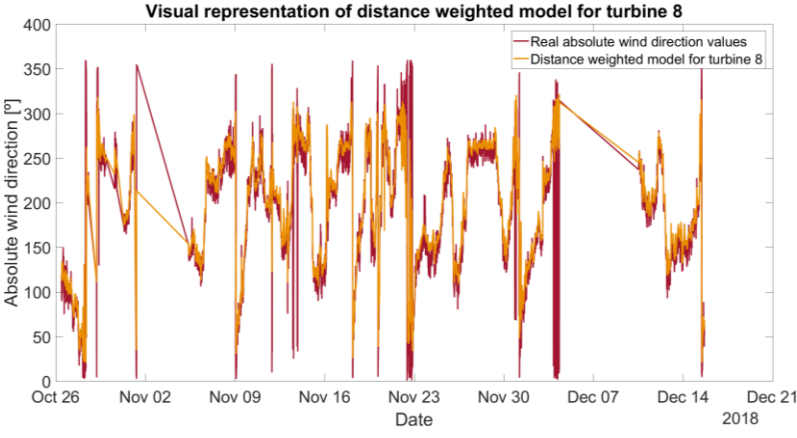


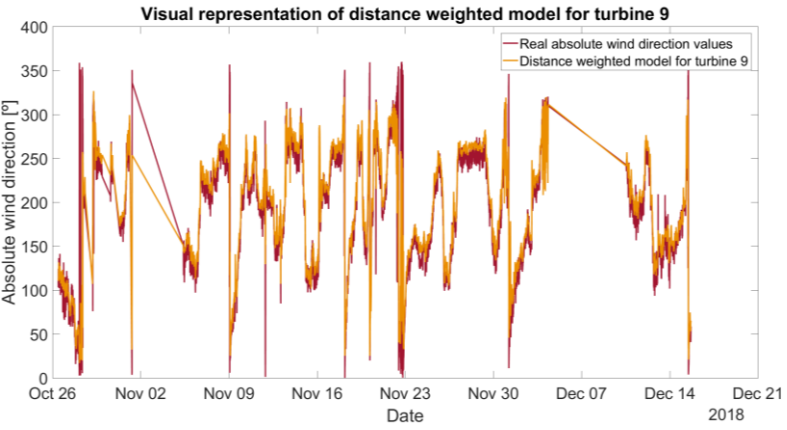


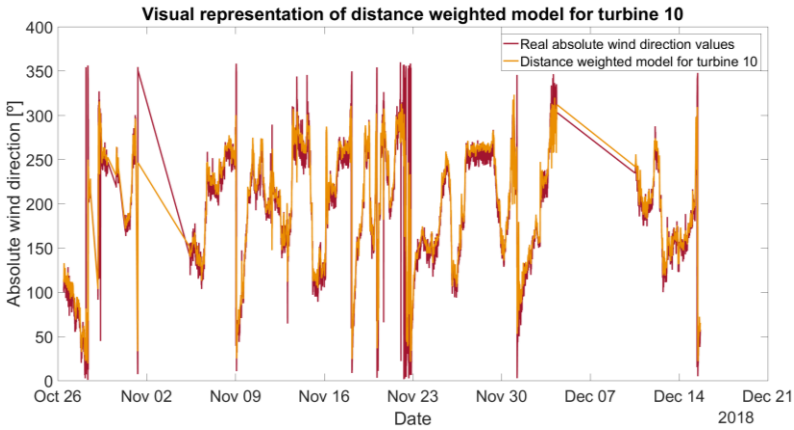


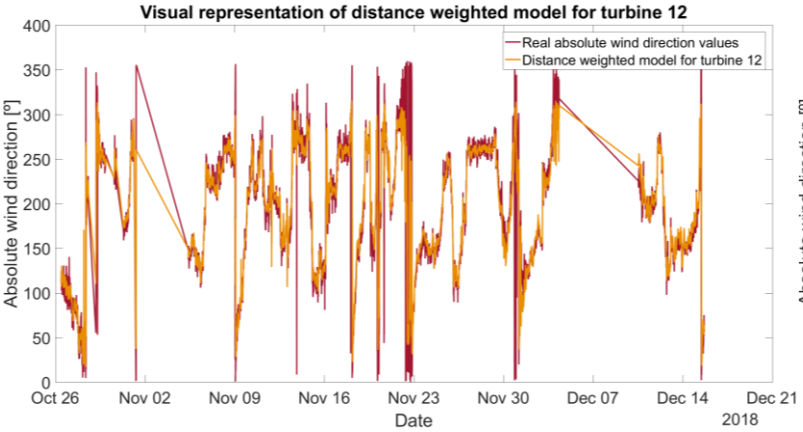


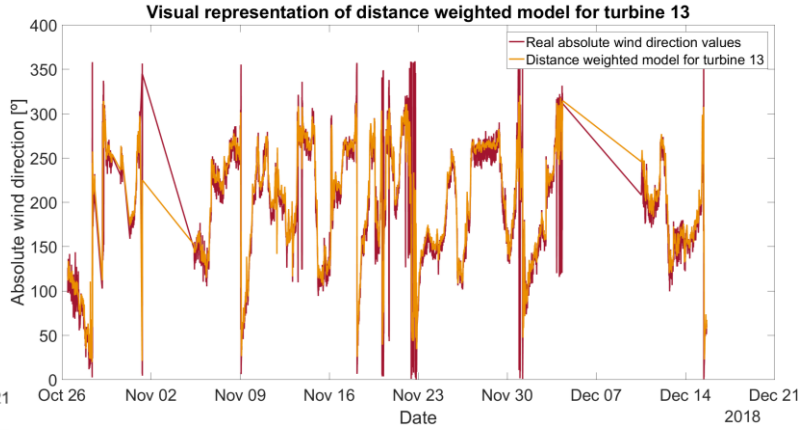


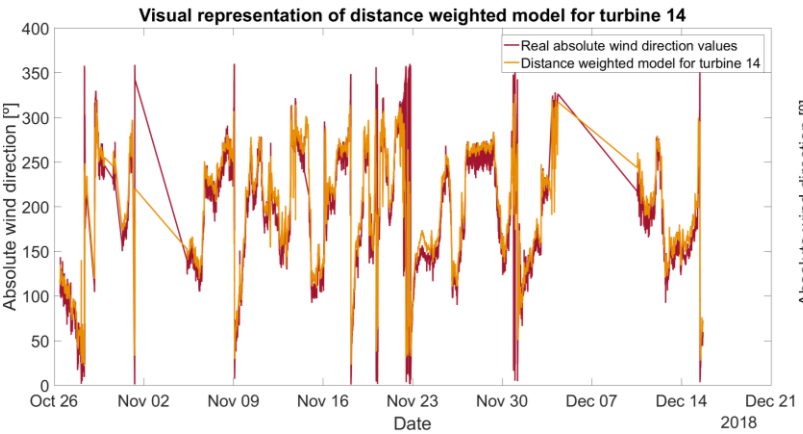


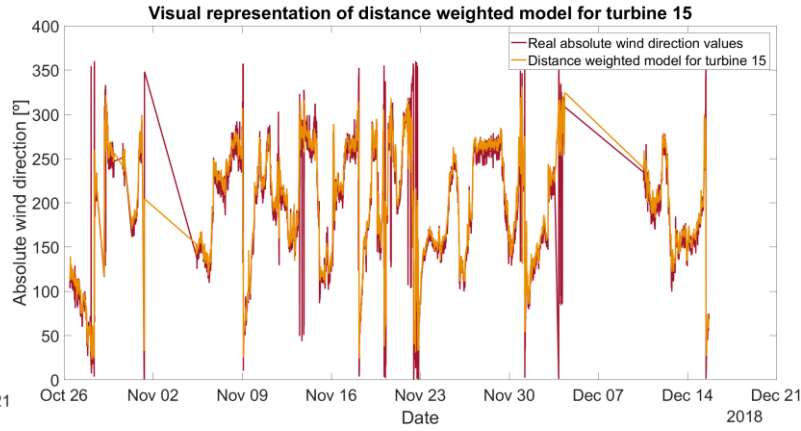


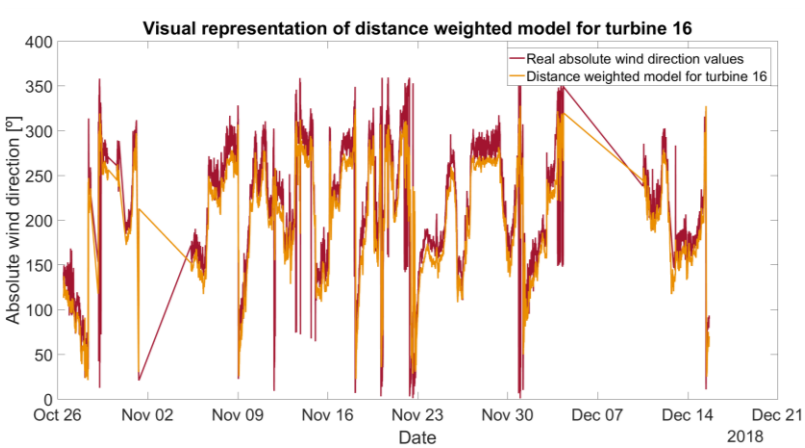


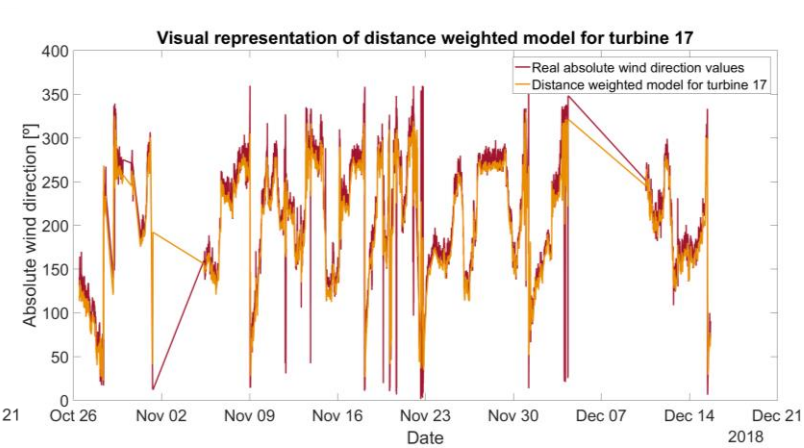


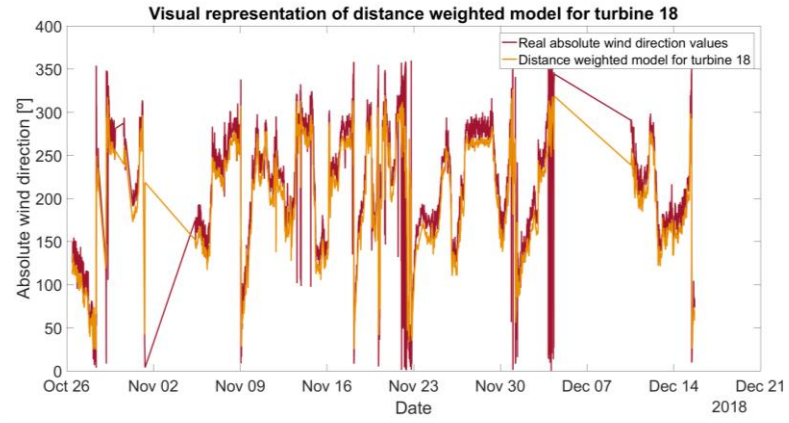


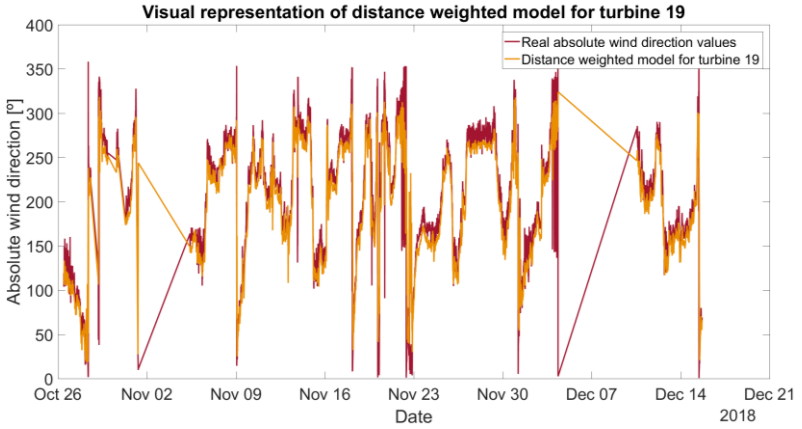


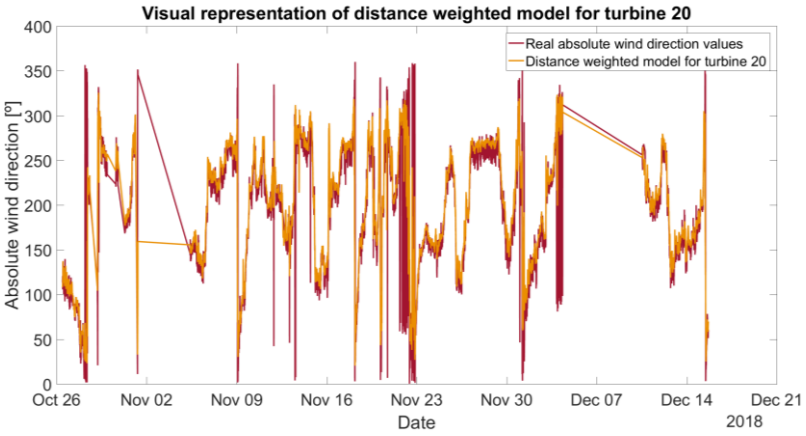




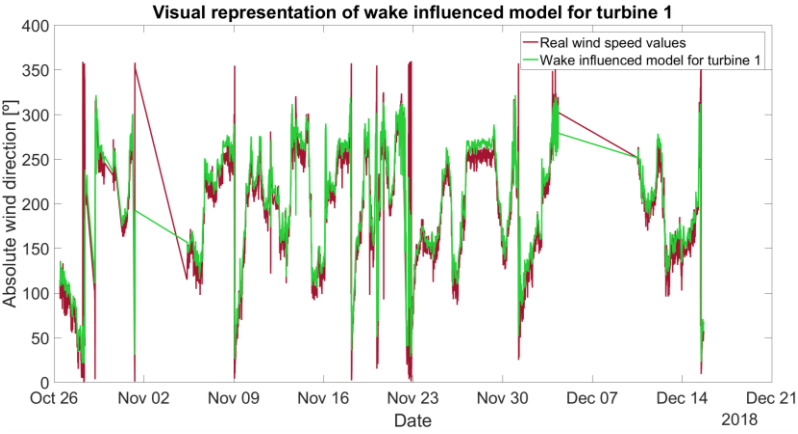


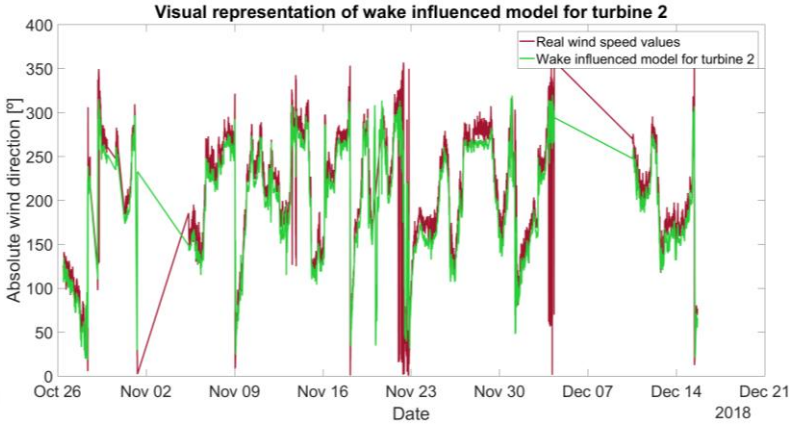


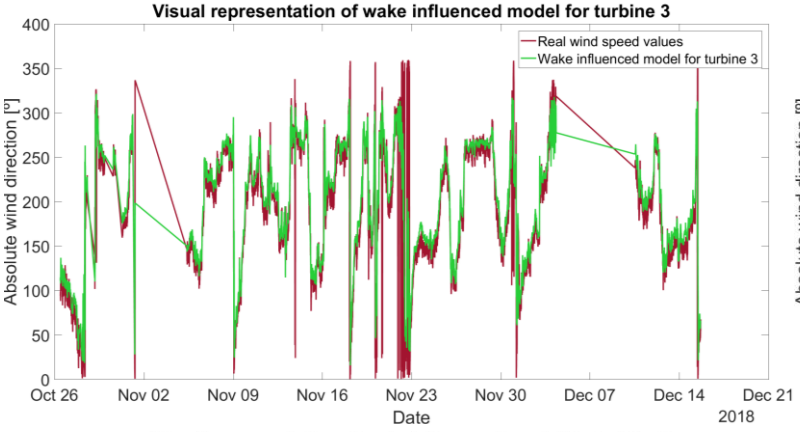


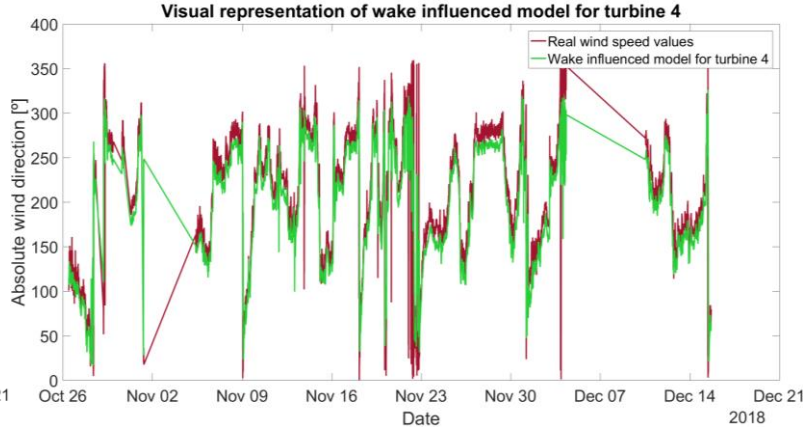


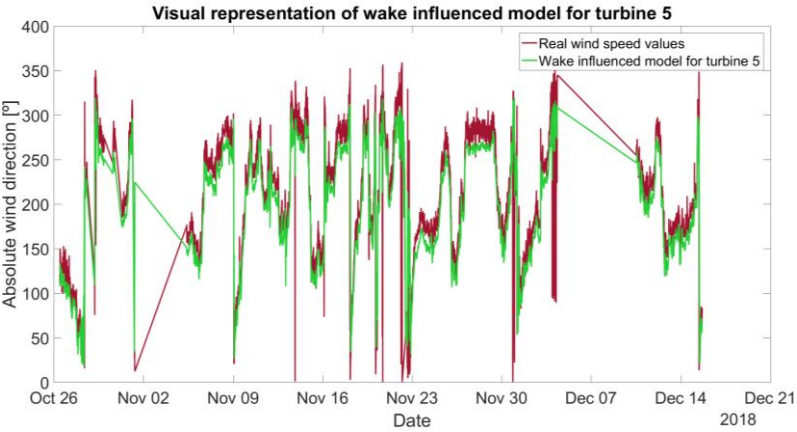
Wake influence model

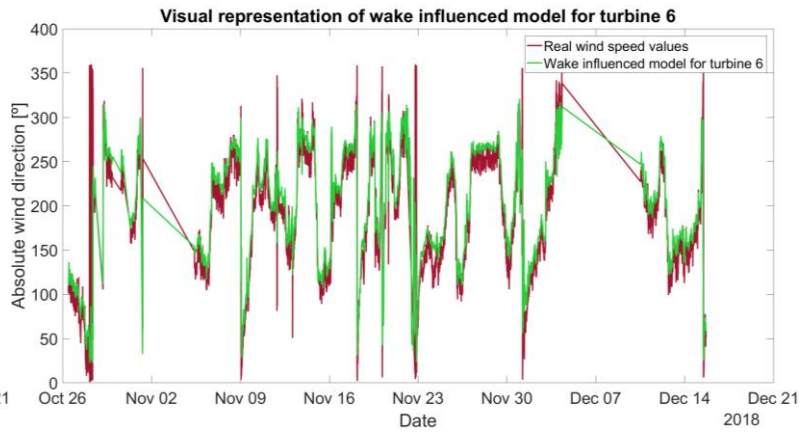


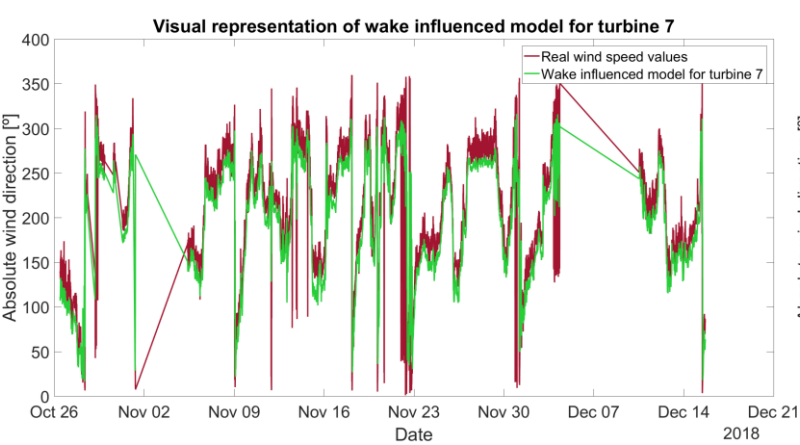


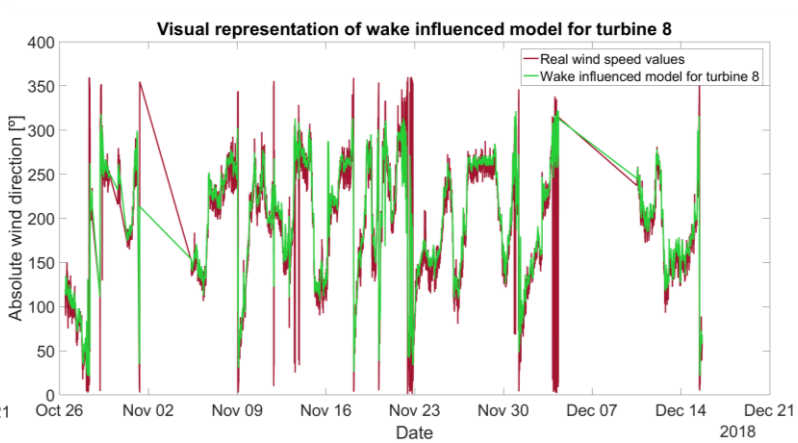


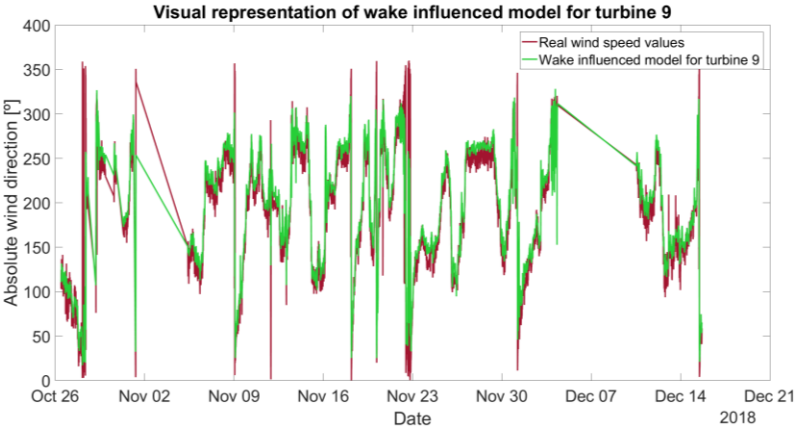


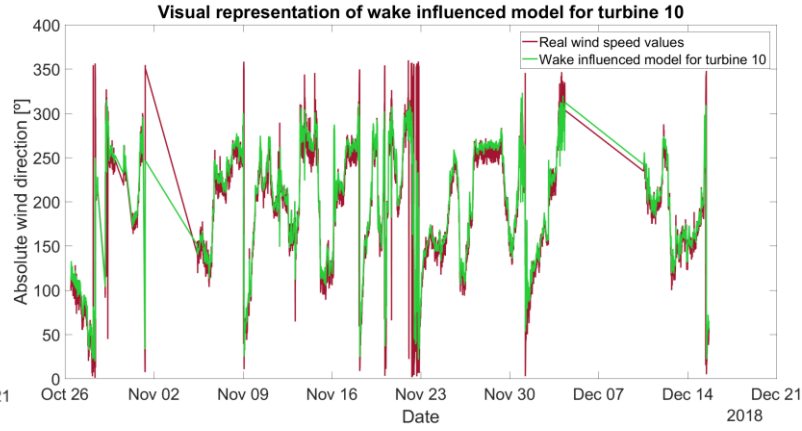


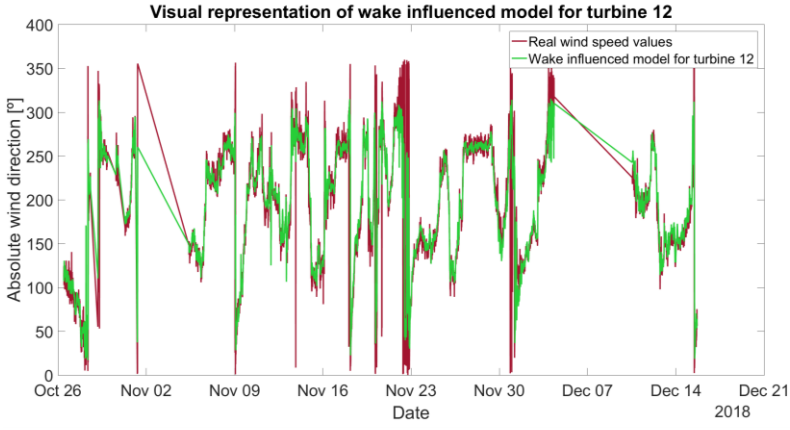


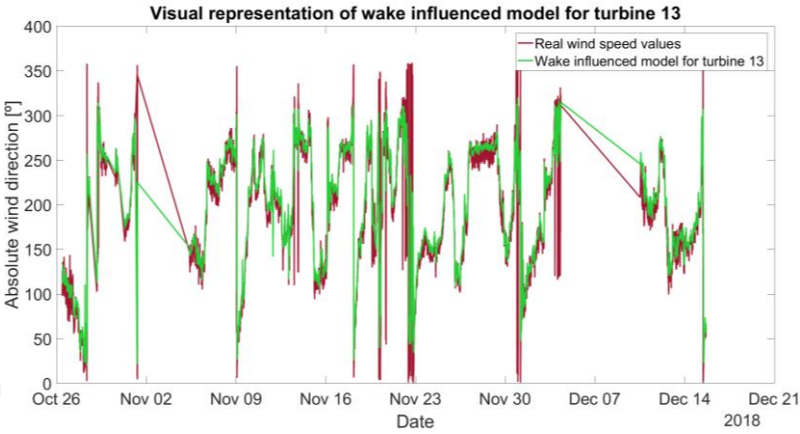


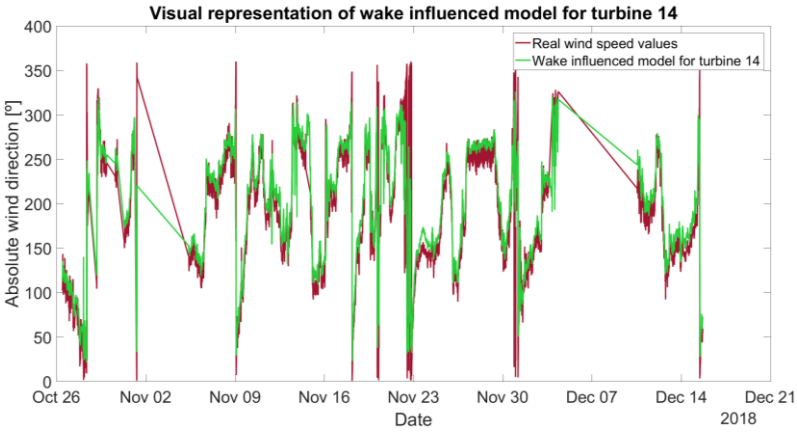


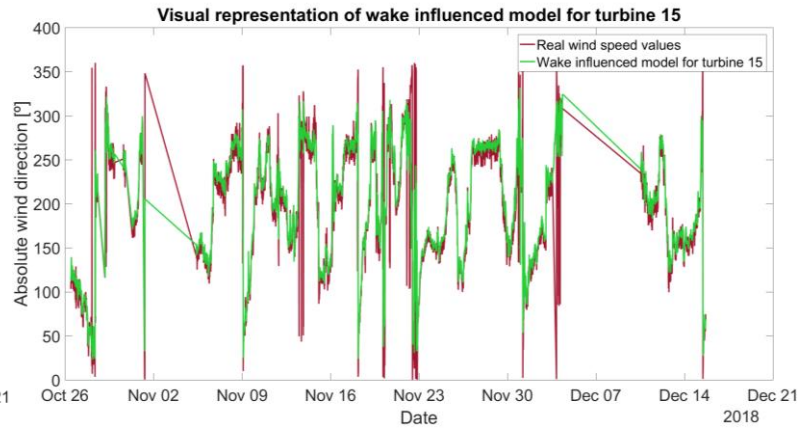


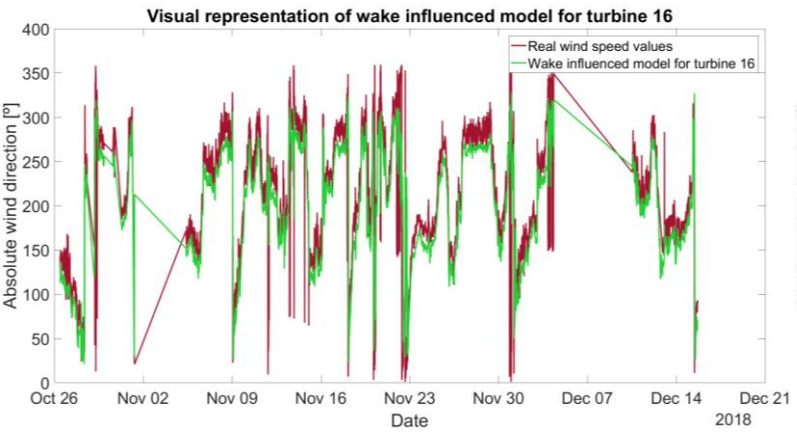


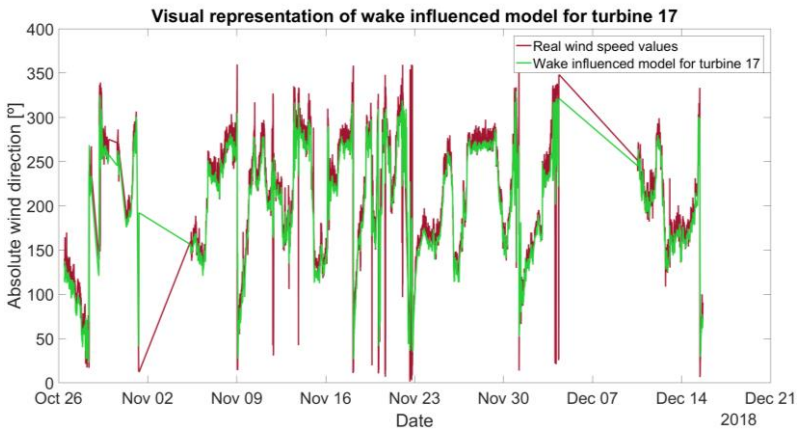


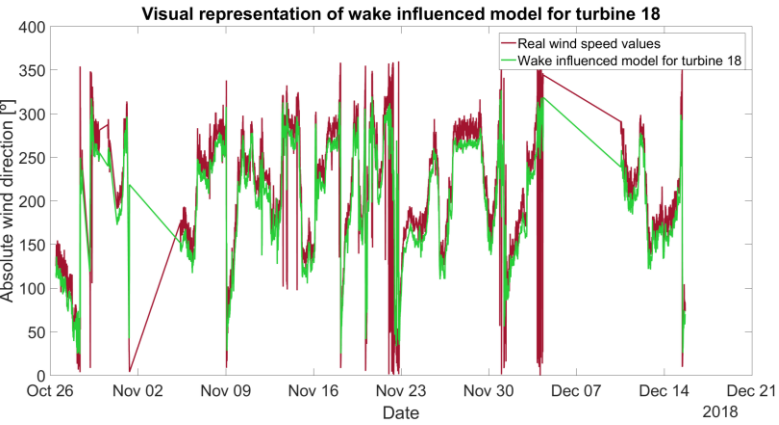


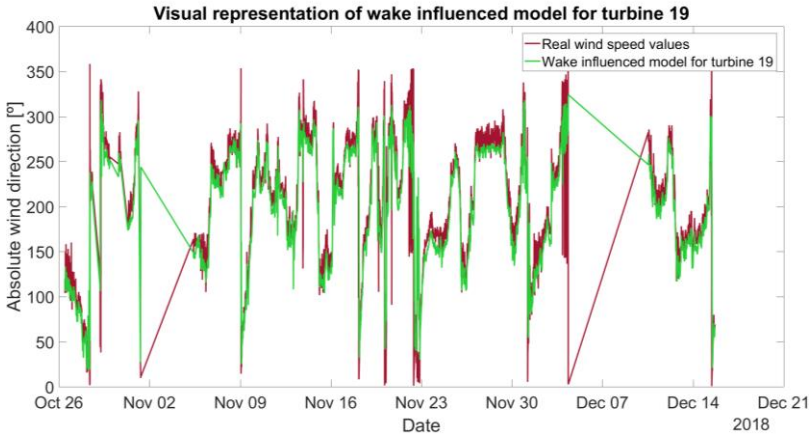


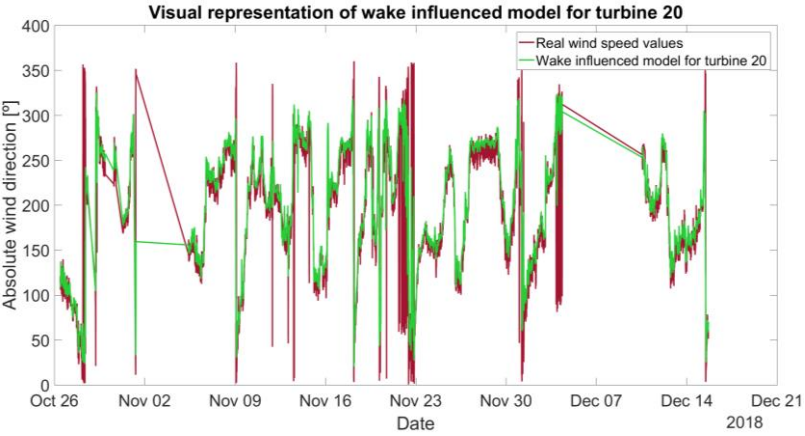




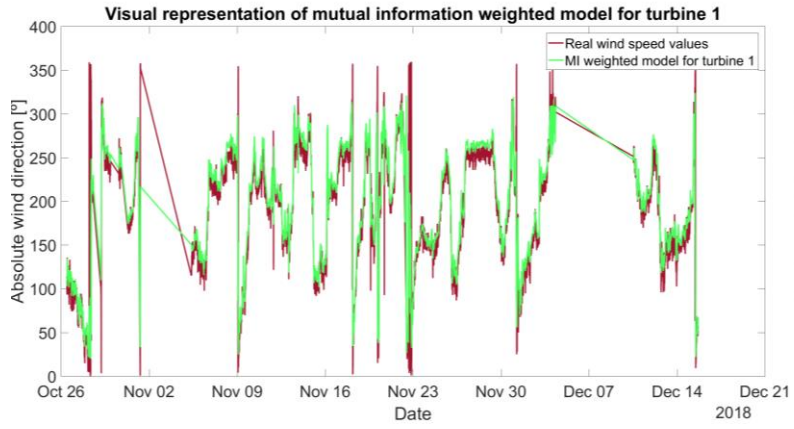


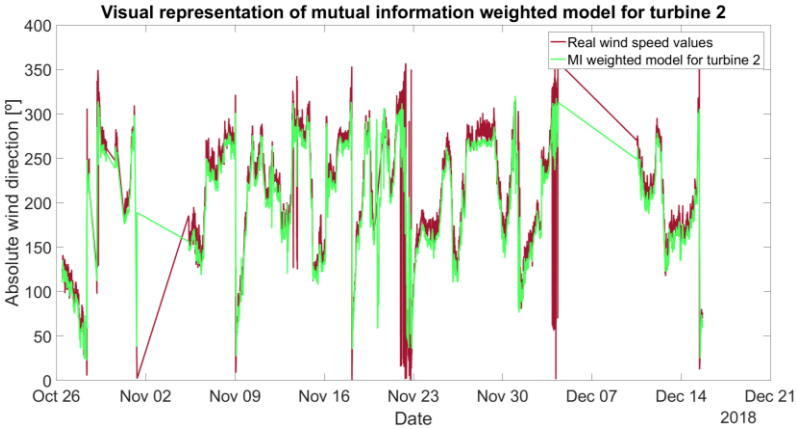


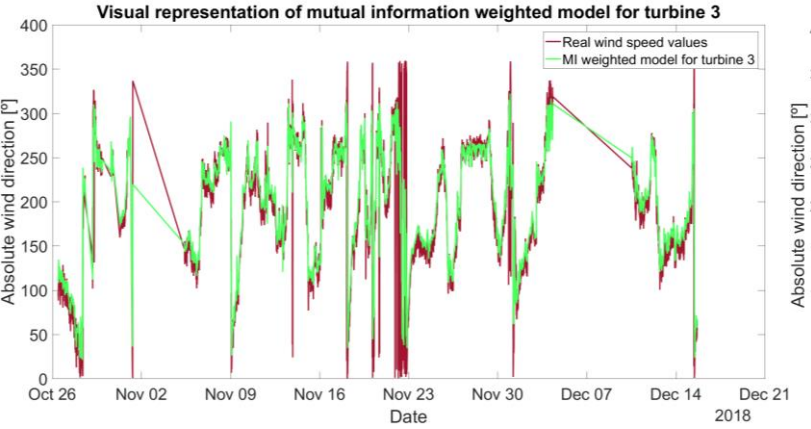


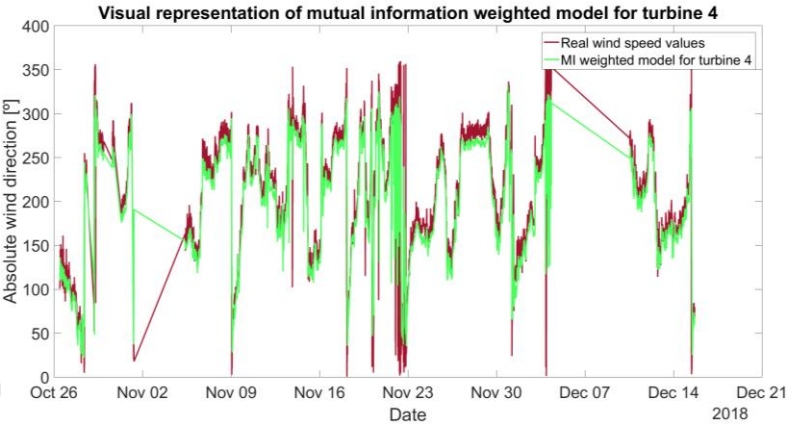


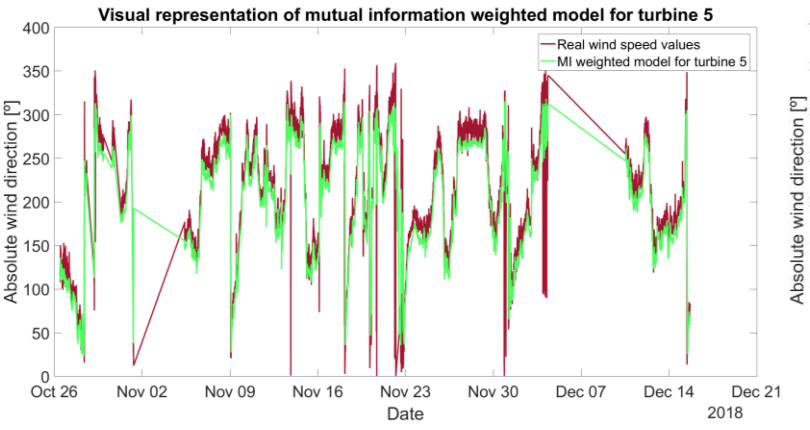
Mutual information weighted model

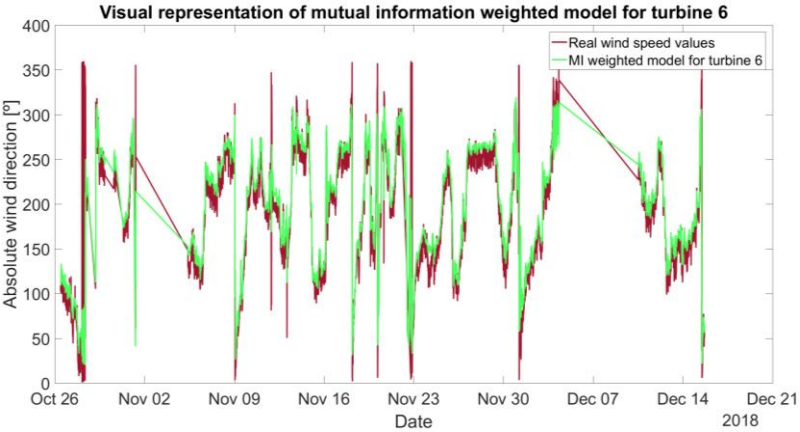


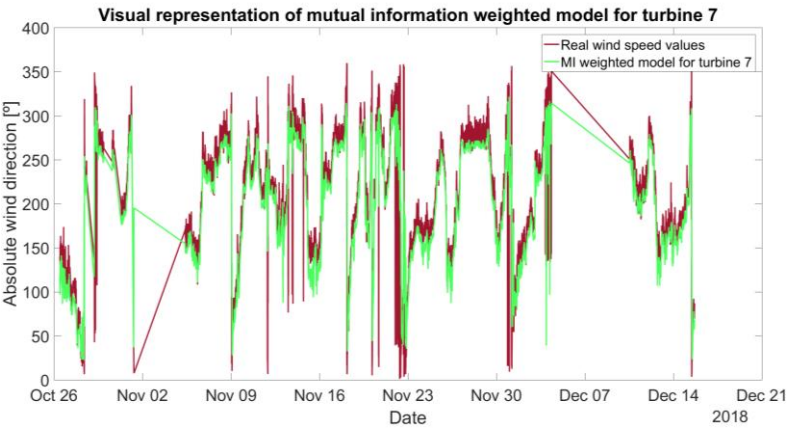


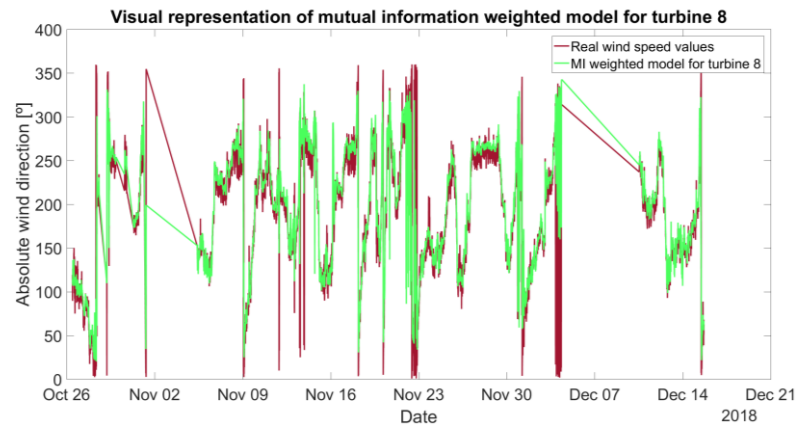


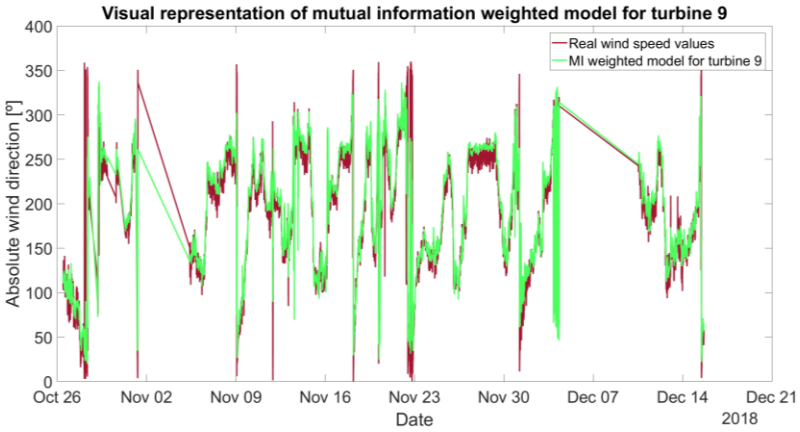


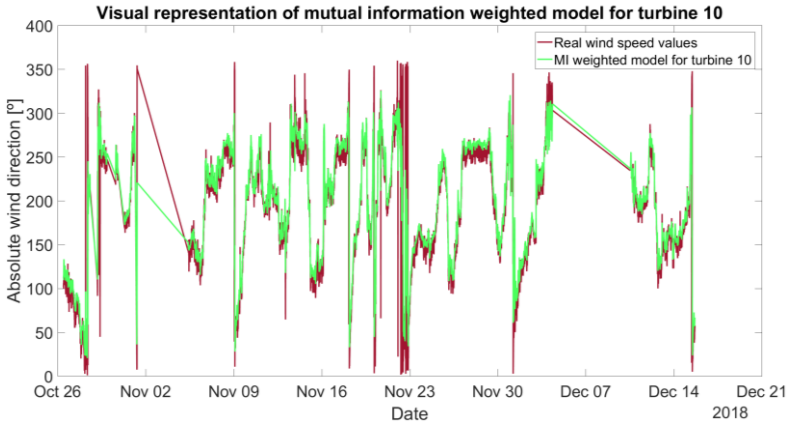


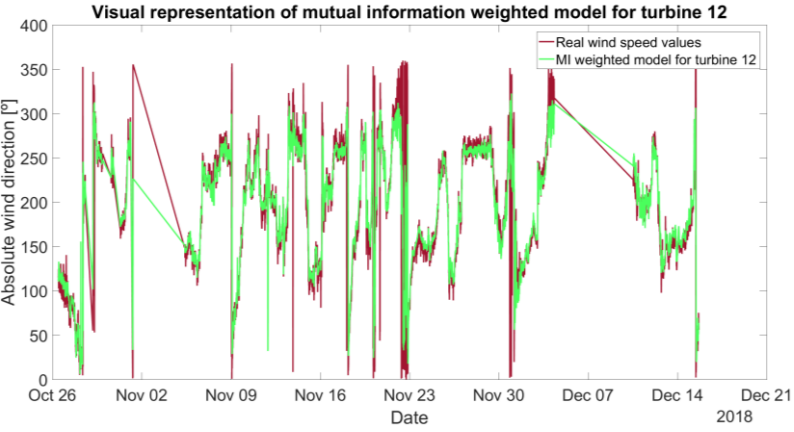


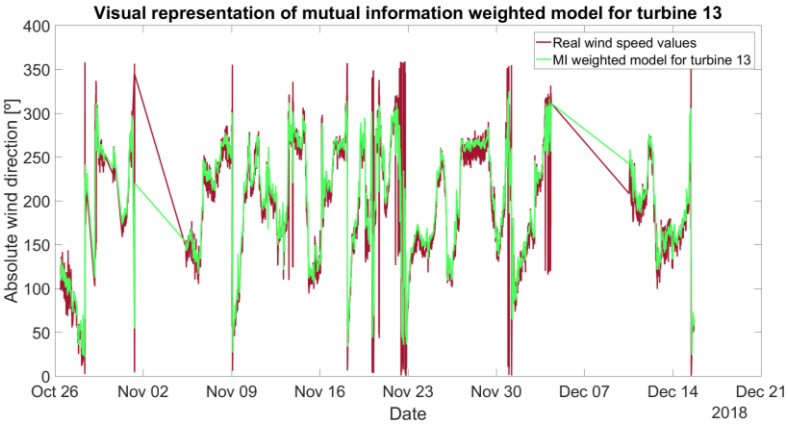


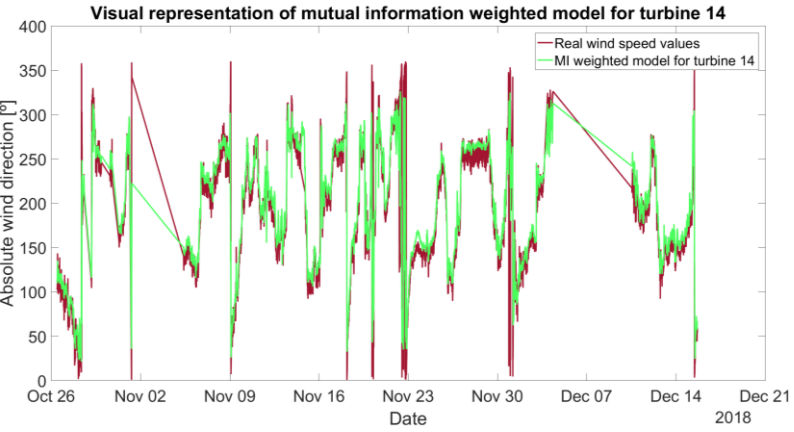


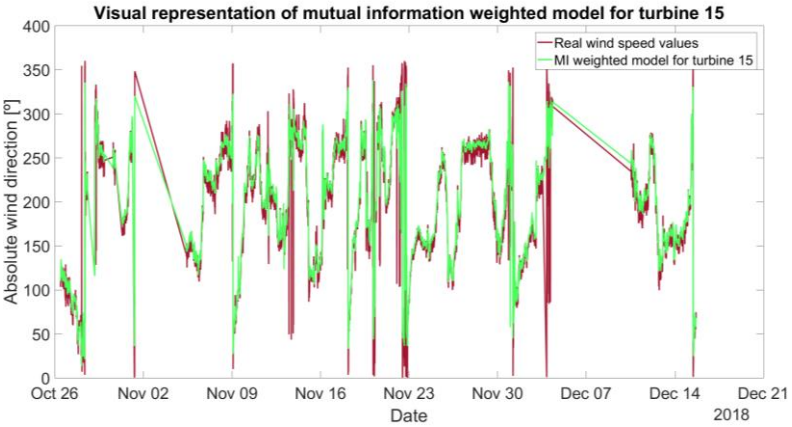


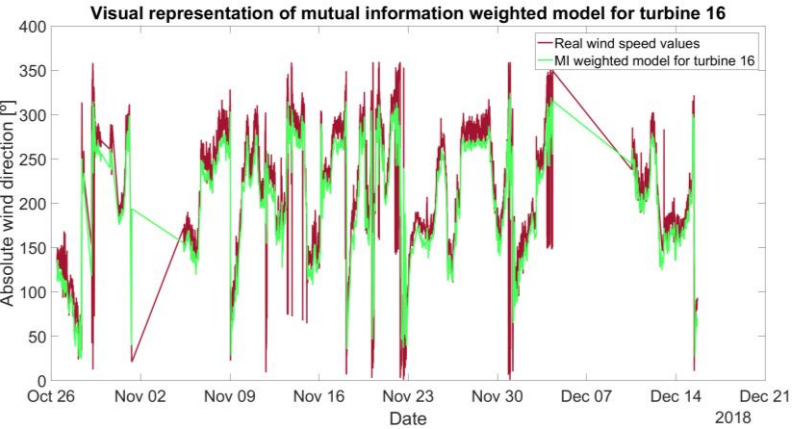


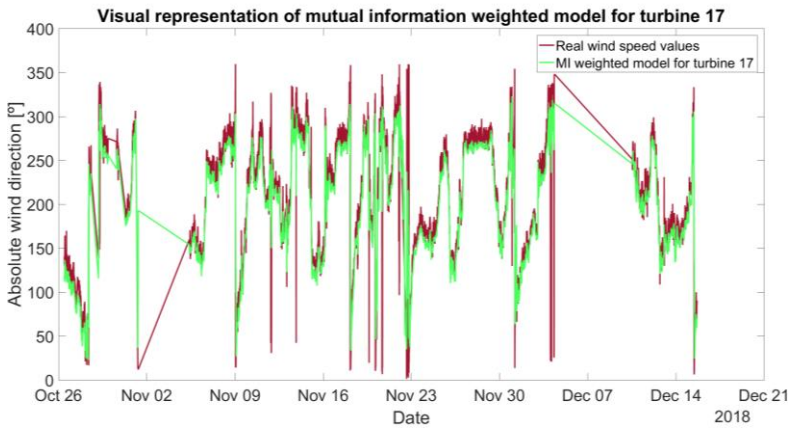


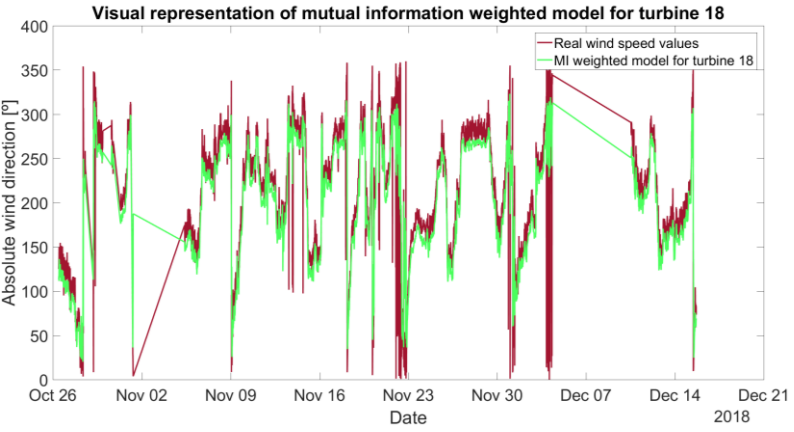


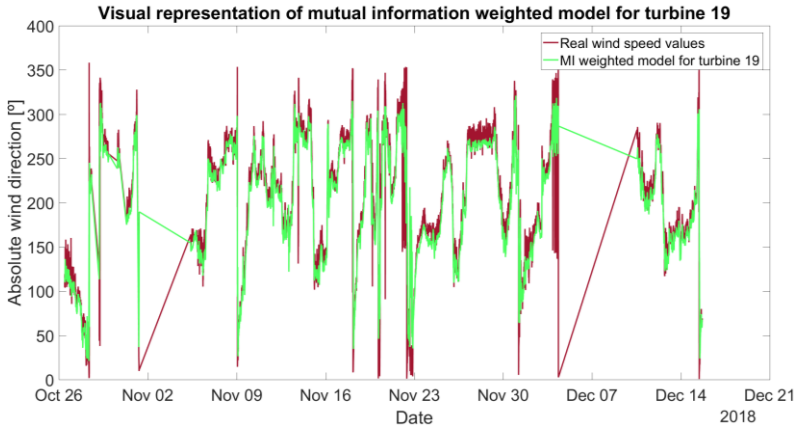


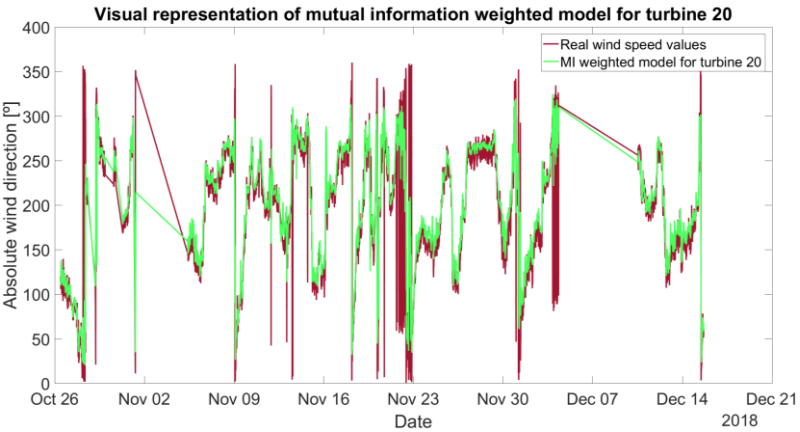












Machine learning model