CODEN:LUTEDX/(TEIE-5436)/1-70(2020)

Electrical Engineering and Automation

Industrial

Accuracy of EMD-WRF reanalyis for wind power estimations, using WindPRO

Mathieu Bruneau

Division of Industrial Electrical Engineering and Automation
Faculty of Engineering, Lund University

Master of Science in Engineering

Master's thesis

Accuracy of EMD-WRF reanalyis for wind power estimations, using WindPRO

Author:
Mathieu Bruneau

Supervisor:
Jörgen Svensson

Examiner:
Olof Samuelsson

Lund
March, 2020

LTH
FACULTY OF ENGINEERING

Division of Industrial Electrical Engineering and Automation
Faculty of Engineering
Lunds Tekniska Högskola

Abstract

Developing wind farms is a long and costly task. Assessing the potential of a site is a major part of the development of a wind project. Reanalysis, that are modelling of the wind speed made from scattered wind data across the world (from satellites, weather stations for example), can help to fulfill this step as they are improving their accuracy with each new release. Among them is EMD-WRF, a reanalysis dataset released in the summer 2019. It was chosen to assess this particular reanalysis due to its novelty and because no study was done on it yet. The reanalysis data is compared to 23 measurement masts located in France, that are at heights ranging from 42 m to 122 m above ground level. The comparison is performed on the software WindPRO which allows to handle wind data by moving them to one location to another, by the process called *downscaling*, and to do wind power estimations. The reanalysis overestimates the measurements by an average of 18.6 % and with a standard deviation quite high of 12.52 %. The overestimation does not seem to be linked to the correlation coefficient between the two dataset or the geographical proximity of the dataset, except for flat terrain with very few trees. However, the impact of the correlation between terrain complexity, ie. orography and roughness, and the overestimation is investigated and shows promising results. It shows that categorizing the sites based on terrain criteria can help to reduce the scattering of the results. The measurement sites with simple terrain are generally having the least overestimation from the reanalysis data. This led to test modifications, based on the type of terrain, applied to the studied reanalysis in order to calculated the wind power of known sites. The wind power estimation was improved in all the sites but it led to some underestimation of the site's potential.

Acknowledgements

First I would like to thank my supervisor Jörgen Svensson who was very supporting of my project. Then a big thank you is also due to my supervisor in *ENCIS Environnement* Valérian Cantegril for his support during this whole thesis.

I would like to thank my parents as well for their trust in my project.

Finally, a big thanks to my friends I made in Toulouse and Andrew Segar who helped throughout every steps of my internship.

Contents

List of Figures

List of Tables

Abbreviations and List of Symbols

Abbreviations

agl Above ground level

EU European Union

GW Giga-watt

Lidar Light Detection And Ranging

MW Mega-Watt

Sodar Sonic Detection And Ranging

TWh Terawatt hour

WAsP Wind Atlas Analysis and Application Program

WRF Weather Research and Forecasting

List of Symbols

ρ Density [kg/m3]

θ critical angle [°]

A Area [m^2]

C_p Coefficient of performance

C_t Drag coefficient

E_c Kinetic energy [J]

k Von Karman's constant

k Wake decay coefficient

m Mass [kg]

N_b Gar box bearing efficiency

N_g Generator efficiency

$U*$ Friction velocity [m/s]

CHAPTER 1

Introduction

Renewable energies are expanding worldwide: the global renewable energy capacity, excluding hydropower, went from 160 GW installed in 2004 to 1246 GW in 2018 (REN21 2019). Of those sources of energy, two come ahead: wind and solar, with respectively 591 GW and 505 GW installed worldwide. The EU is in second position for its installed capacity of renewable energy with 339 GW in 2018, behind China with 404 GW. This shift comes in response to the necessity to replace fossil-fuel based energies which are sources of greenhouse gases that accelerate the global warming of the planet. Scenarios of the climate change are more and more alarming regarding the consequences of increase of the global temperature: rise of the sea level, augmentation of catastrophic events such as hurricanes, famine and conflicts due to an increased scarcity of resources.

The growth of installed renewable energy is supposed to continue in the near future. The EU set targets and policy objectives for the period 2021 to 2030. Those are:

1. At least 40% cuts in greenhouse gas emissions (based on 1990 levels);
2. At least 32 % share for renewable energy;
3. At least 32.5 % improvement in energy efficiency

Regarding those targets the *Programmation Pluriannuelle de l'Energie* (Multiannual Energy Program), the document planning the ecological transition of France, set the objective to double the wind power energy production from 2016 (12 GW) by 2023 (24 GW) (ADEME 2017). Currently the installed power is 15 GW (*Wind in France: power installed* 2018), which is in line with the objectives. However, in Europe, finding good sites for wind farms is increasingly difficult. Indeed some restrictions, eg. being 500 m away from housings or avoiding military flight paths, are limiting the implantation of wind farms. Moreover, the sites with the best wind potential are already being used, as they were usually among the first to be set up, or are blocked by people refusing wind farms to be constructed. This led some developers to move to offshore where stronger winds are available and less restrictions are applied. But onshore development is still continuing. France has yet to install an offshore wind farm.

1.1 Background

Historically in Europe, wind has been first widely used in wind mills in order to transform wind force into mechanical force. Their appearance in Northern Europe is situated around the 12th century. They were a source of energy until the Industrial Revolution: coal and steam engines had the advantage to provide energy where it was needed and used at a desired time while the wind was considered non-dispatchable and non-transportable. The emergence of the wind turbine, ie. transforming wind forces into electricity, is due to five factors according to Manwell, McGowan, and Rogers 2009:

- An awareness of the limitation of the main resource of energy at the time: fossil fuel. As well as the awareness of their side effects on the environment;
- A strong potential of wind energy in numerous places on earth;
- The rise of new technologies capable of transforming the wind into electricity (material sciences, computer sciences, aerodynamic understanding, power electronics and so on);
- A vision of a new way to use the wind as the source of energy;
- Policies in favor of the rise of these technologies.

Around the 1960s and 1970s, with books such as *Silent Spring* (Carson et al. 2002), describing the impacts of human activities on the environment, and the Oil crises, the world saw the arrival of wind turbines as an alternative to fossil fuel. At first this technology was expensive and had trouble to reach good efficiency. But with governmental support and technology improvement the wind industry developed itself from the 1990s. In Europe, Denmark and Germany were the first ones to show interest in a wind-based source of electricity. This was partly due to concerns about both climate change and nuclear power. They are still two of the major actors of the wind industry with respectively 5.48 GW and 55.55 GW of capacity installed in 2017 (*Annual electricity generation in Germany* 2019).

Meanwhile, the development of wind power in France came later and slower, mainly due to a large share of nuclear power developed after the second World War. In 1996, the program *Eole 2005* (Benedict Jourdier 2015) was launched with the objective to install 250 MW to 500 MW of wind power until the end of 2005, by choosing suitable projects. Unfortunately, most of the 55 selected projects did not end up being constructed. By comparison Germany had already 18 GW installed at the time. In 2008, France pledged to reach 23 % consumption of renewable energy by 2020. The investments were supposed to be for the years 2009-2020: 19 GW of onshore power and 6 GW of offshore power. The rate of annual installation of wind power was of 1.2 GW/year in 2009 and 2010 but steadily decreased to 0.6 GW/year installed in

2013. At the time it seemed difficult to reach the objectives mentioned above. But the rate increased and reached 1.79 GW/year of power installed in the year 2017. 15 GW were installed by the end of 2018. The drop in number of installed wind farms can be explained by the continual increase in the length of administrative paperwork to do before being allowed to start a construction, as well as a strong lobby against the wind farms. Indeed, regarding offshore wind projects, no farms has been built near the French coasts yet due to association protests.

In 2018, the electricity production of France from wind was 27.8 TWh which represented roughly 5 % of the national consumption. It is the second source of renewable energy behind hydropower in the country. France is ranked 4th in Europe in wind-based electricity production; Germany being ranked first with 111.6 TWh.

In addition, wind farm development is expensive. Poyri 2016 assesses the costs of wind farms to be on average 1.4 million €/MW installed onshore in France. The costs of wind projects are often taken by banks in the form of loans. Developers and banks base their economic models according to the potential production, or wind power estimation, that has been estimated. It allows the calculation of the project financing and its financial viability. If the wind turbines produce less than the estimation, it can be financially difficult for the developer to make a profit while reimbursing the bank loan. If the estimation underestimated the reality, then it can be a source of refusal of loans from the banks that see it as an non-viable or non-profitable project. But the time span of the development also has to be taken into account. Wind projects take months to years to be build: legal procedures can hinder the project.

Consequently, wind developers have to be certain of the economic viability of their future project. The sooner the developers know the potential of the site, the sooner they can take measures to move forward or cancel the project. That is why wind estimation has a critical role in the wind farms development.

1.2 Wind estimations

Generally, wind estimations are usually done by the developers to assess the economic viability of the project or they can be done by independent companies for banks in the perspective to provide a loan. Most commonly, it is done by installing a measurement mast on the site of the project. This mast records wind data for a full year, ie. wind speed and direction, pressure and temperature. These data are used to calculate the electricity production. For those calculations, some inputs are necessary: a fine topography of the terrain, model and coordinates of the wind turbines. Furthermore, the losses have to be estimated in these calculations as well. Finally, the uncertainties of the calculations made are also taken into account.

Another wind estimation can be performed, without a measurement mast. It is done

only with wind data modelled from real measurements (such as weather stations or satellites measurements) that were extrapolated on a geographical grid by a prediction model and are called reanalyses. The same method as the one described previously is used and the wind software WindPRO allows to handle reanalysis data through a process called *downscaling* which is explained on section 3.4.2.1. However, such reanalyses data are known to provide a less accurate result, as it is shown in Chapter 4. It is done at the early stages of the wind development. One main advantage is that it is less costly: reanalysis data can be purchased online from providers for a few hundreds euros while a measurement mast costs around €100 000. It can prevent a developer from investing in a measurement mast and losing time if the site does not seem to be producing enough power.

1.3 Aim and research questions

Reanalyses come from different providers that use different real-life data and various extrapolation methods. In this master thesis, it was decided to focus on a newly released reanalysis: EMD-WRF, released by EMD International A/S with the Weather Research and Forecasting (WRF) model. This master thesis aims to assess the accuracy of this new reanalysis by comparing it to measurement masts data provided by *ENCIS Environnement*, a French company working, among other areas, in wind power estimations. It also proposes to do a correction of the wind speed depending on the terrain of the site. Then the estimation of production are compared to real-life wind farms production.

The research questions are:

- What are the methods used to assess the energy production of a wind farm?
- What is the accuracy of the EMD-WRF reanalysis?
- Is this accuracy depending on some parameters?
- Which solution can be proposed to improve the accuracy?

1.4 Limitations

The possibility to compare wind power estimation made by developers to our method was considered. Unfortunately, energy estimation methods are usually kept secret by developers that use their own database and own methods. Therefore it was not possible to get a wind power estimation directly from a developer.

1.5 Outline of the report

Chapter 2: *Wind power plant overview*
Presentation of the different steps of a wind power plant development with a focus on micrositing. These steps go from site prospecting to construction and maintenance phases. Then a presentation of the method used to calculate the electric production of a wind power plant is done. The losses and uncertainties that are used in an estimation of wind project potential are also presented.

Chapter 3: *Wind analysis*
Focus on the wind theory at a micro scale, ie. within a few meters. The main parameters to take into account when doing a wind estimation are presented: orography, roughness and wind profile. It also presents the wind measurement instruments that are used in relation to the thesis work. The software WindPRO and WAsP, later used in the experimental part of the thesis are presented as well as their limitations.

Chapter 4: *Wind reanalysis*
The concept of reanalysis is defined. The major dataset of global reanalysis are presented: MERRA 2, CFSRv2 and ERA-5. Their accuracy is described through a literature review. Then the mesoscale reanalysis EMD-ConWx and EMD-WRF are displayed and their accuracy is also introduced.

Chapter 5: *Reanalysis and measurement comparison*
Description of the methodology used. The studied sites are also presented. The reanalysis wind speed data are compared to the measurement masts wind speeds. The monthly variability are also presented. Parameters such as correlation coefficients, altitude of the site and height of the measurement mast are also compared.

Chapter 6: *Categorizing data*
Description of the method and criteria used to separate the site depending on their terrain. The possibility of a link between accuracy and geographical location is also presented.

Chapter 7: *Wind power calculation*
Calculation of the wind power production of five sites using reanalysis wind data. This calculation assesses the usefulness of adding a monthly modification to the reanalysis.

Chapter 8: *Conclusion and future work*
In the conclusion, the research questions are answered while presenting some possible future work.

CHAPTER 2
Wind Power Plant Overview

This chapter aims to present to the reader the different steps of a wind farm installation, from prospecting a favorable site to the operation and maintenance of the wind farm. Then the method of calculation of wind production is displayed. It is done by explaining the statistical modelling of the wind and the equations needed to calculate the power from the wind speed received. Then the different losses that can occur in the power plant are listed, as well as the uncertainties coming from the modelling of the wind farm.

2.1 Steps of wind power plant development

Before installing wind turbines, the best location has to be determined. The main goal of this process is to maximize the income of the wind farm while minimizing the impacts on the environment namely the noise and visual pollution. The major steps of this process are:

- Prospecting promising sites
- Micrositing
- Permiting
- Construction
- Operation and maintenance

2.1.1 Prospecting promising sites

Wind developers start their wind power plant projects by looking for sites that are suitable for them. They search sites with strong wind, but they cannot have wind measurements that are both on the entirety of the country and that recorded a full year (which allow to see the annual variability of the site). So, as a first step they can have an estimation of the site potential using wind maps and reanalyses wind speed data (detailed in Chapter 4) and then install a measurement mast in order to record the wind for at least a full year.

2.1.1.1 Wind maps

First wind resource maps or wind atlases are available to see a general potential of a region. They contain Weibull parameters which are parameters of a standard wind distribution (see section 2.2.1), direction and mean wind speed over a certain period (Manwell, McGowan, and Rogers 2009). One of the first atlases in Europe is the European Wind Atlas published in 1989. Figure 2.1a shows a global wind map of data made by the European Wind Atlas. It was the first to provide an overview of the wind potential of the European countries (Troen and Lundtang Petersen 1989).

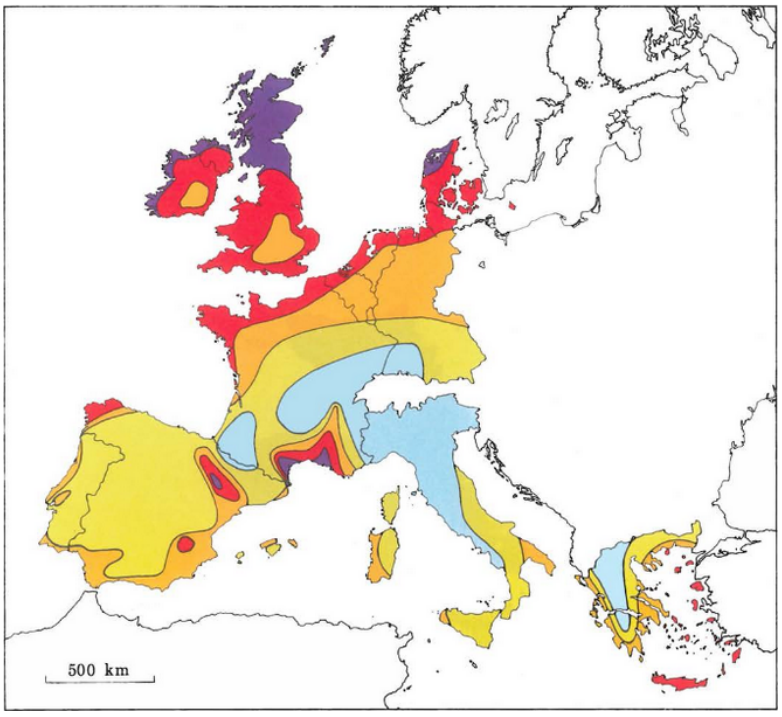
Other wind maps are now available online such as Global Wind Atlas and Windprospecting (DTU Wind Energy and World Bank Group 2018 ; *WindProspecting* 2019). Their geographical resolution varies between hundreds of kilometers to kilometers. Figure 2.1b and figure 2.1c show the current accuracy of the Global Wind Atlas and Windprospecting, respectively. They often do not account for variation on the micro-scale. Indeed on such a large scale only global variations of the terrain are taken into account.

2.1.1.2 Wind data

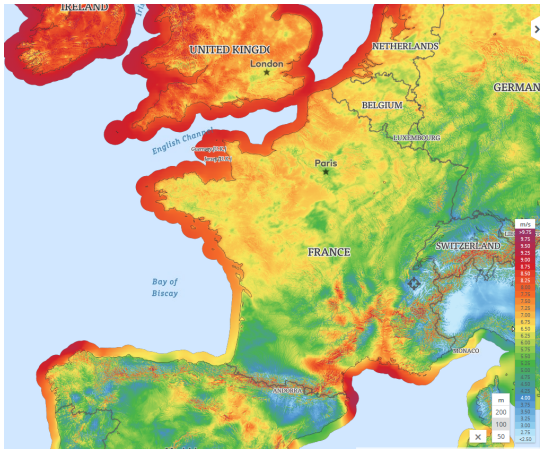
To improve the accuracy, wind data are available. In France, *Météo France* (official national meteorological service) can provide wind measurements. Their weather stations are installed at usually 10 m above ground level (agl) and at different locations in France. Since they record real measurements they can give very precise and accurate data of the wind with a six-minute time step (*Meteo France* 2012). Those wind data can be used to do a rough wind power estimation, usually with software such as WindPRO, WAsP or in-house programs. The first two software are presented in section 3.4.

However, 10 m agl is relatively low compared to the sizes of wind turbines that are at around 100-200 m nowadays. Indeed, with the altitude, the wind speed generally increases, but the rate at which it does depends on what is called the wind profile and it is presented in section 3.2.3. Besides, at 10 m agl the wind data is heavily influenced by the surrounding terrain such as trees or buildings. The roughness of the terrain can significantly hinder the flow of the wind. Another disadvantage is that the data can be several kilometers away from the studied site, since the weather stations are scattered. Indeed, wind can have different speed and direction for two locations a few kilometers away, eg. if they are separated by a hilly terrain. This parameter of the terrain topography is called the orography. Wind profile, roughness and orography have their dedicated part in chapter 3 explaining their importance for wind power estimation. For all these reasons, the result of the estimation using this type of data is generally deemed too uncertain.

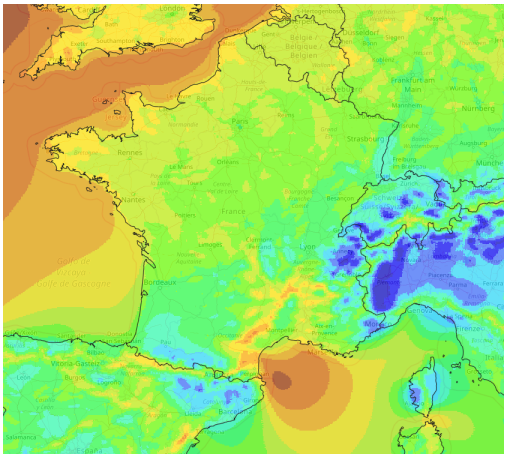
Wind reanalyses data can be used instead to estimate the wind of an area. They are not real measurements and can also be far from the site. But they have available heights



(a) Wind Map of Europe, from Troen and Lundtang Petersen 1989



(b) Wind Map of France, from DTU Wind Energy and World Bank Group 2018



(c) Wind Map of France, from *WindProspecting* 2019

Figure 2.1: Wind maps

that are near to the future wind turbines mast. Chapter 4 presents more in depth the reanalysis data accuracy and its definition.

At this stage of the wind farm development, the choice of the turbine should be narrowed down to a few models and a few layouts of the turbines should be selected.

2.1.2 Micrositing

When a site has been pre-selected, it is necessary to do a more in-depth analysis of the future production of the wind farms. The goal is to select the turbine model and the specific layout that is going to maximize production. The estimated electricity production that takes into account the losses and the uncertainties has to be calculated.

First of all, it is highly recommended to set up a measurement mast at the given site. It should record data for at least a year due to seasonal variation. The measurement mast should be at a height near the future wind turbine height: the Measnet report (MEASNET 2016), which gives guidelines for micrositing, advises for 2/3 of the wind turbine's height. It is done to reduce to any uncertainties on the future production of the wind farm.

Then softwares (eg. WindPRO) are used to calculate and optimize the estimated average production of the wind farms. The different layouts and turbines are tested to maximize the revenue. One main focus is to model the aerodynamic interactions between turbines that affect the energy captured, namely the Wake Effect presented on section 2.3.1. As wind turbines harness kinetic energy from the wind, the turbines downstream are going to receive less energy. It is recommended to limit the number of wind turbines aligned with the dominant winds on a given site and preferably align the turbine perpendicularly to the dominant wind.

The needed data to model a wind farm production, in addition to wind data, are the roughness and the orography data (presented on sections 3.2.1 and 3.2.2) as well as the characteristics of the wind turbines. Indeed, different turbines can be compared, some have different power curves (presented section 2.2.1), different hub height and different diameter of rotor. Different power curves can favor specific wind speeds. An higher hub height can harness a stronger wind, as wind speed tend to increase with height. A larger rotor diameter can harness more wind but it will create an important Wake Effect downstream.

In addition, a thorough analysis of the restriction of this particular site needs to be done. The project has to respect noise limitation, not threaten important species of the fauna and the flora, be accepted by locals, and have good geological condition for the implementation of a wind turbine, and so on depending on the legislation of the country.

2.1.3 Permitting, construction, operation and maintenance

Depending on the countries, several legal steps need to be performed. Those can be land rights agreement, obtaining power purchase agreement, financing, public support or procuring turbines (Manwell, McGowan, and Rogers 2009). Once everything is completed, the project is filed to an administration authority whose task is to verify the good compliance of the project with the regulations. Then nearby citizens/inhabitants are consulted about their inquiries on the project (eg. noise, visual impact). It can lead to further studies being done about more specific or forgotten topics eg. a specific building nearby that was not known by the developer and/or was not mentioned in the studies. Then, if not satisfied, citizens can still pursue legal action to prevent the farm. Once and if it is over, the construction phase can begin.

The construction phase starts by preparing the site for the engines that will arrive and help to install the wind turbines. Then the building of the wind turbine and its foundation can take around 2-3 weeks depending on the weather conditions. Last, the wind turbines can be connected to the grid.

The owner realizes operation and maintenance tasks throughout the lifetime of the wind farm.

2.2 Annual Electrical Production calculations

2.2.1 Statistical modelling of the wind

The Weibull distribution is a statistical distribution that is often used to model the wind data. It depends on two parameters:

- A, scale parameter, that is close to the mean wind speed;
- k, shape parameter

The probability density function is, for a wind speed V [m/s]:

$$p(V) = \frac{k}{A} \cdot (\frac{V}{A})^{k-1} \cdot exp[-(\frac{V}{A})^{k}] \tag{2.1}$$

Different examples of distribution are given in the figure 2.2. In the case where k =2, it is called a Rayleigh distribution.

Historically, wind statistics has been used for energy production of wind farms since the end of the 1970s and rapidly became the norm for wind production and was the

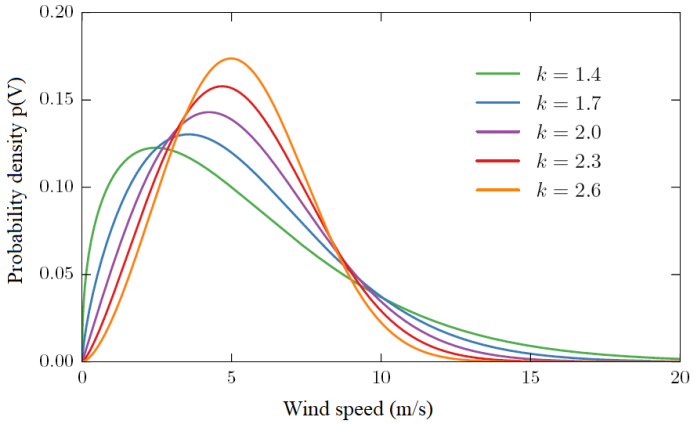


Figure 2.2: Weibull distribution, taken and adapted from Benedict Jourdier 2015

only distribution implemented in the wind power software (Benedict Jourdier 2015). It consists in converting the wind data into its corresponding Weibull distribution and use this to do the calculation. This was mainly due to restrictions on the calculation potential of the computers at the time. It is the method used in the European Wind Atlas in 1989 and used by default by the software WAsP and WindPRO.

The link between wind speed and output power can be obtained in two ways. The method used in the WindPRO sofware consists in having directly the power curves already loaded in the software. The power curves give the output power depending on the wind speed as presented in figure 2.3. The cut-in speed is the minimum wind speed at which the machine starts to deliver power, the rated wind speed is the wind speed at which the maximum power is reached, the cut-out speed is the maximum wind speed at which the turbine is allowed to deliver power. The power curve is calculated by the turbines' manufacturers.

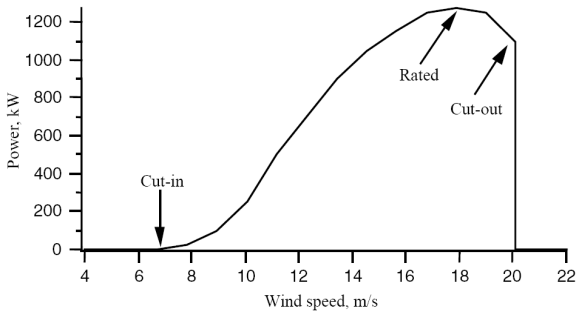


Figure 2.3: Typical wind turbine power curve, taken from Manwell, McGowan, and Rogers 2009

For statistical data, with a Weibull probability density function $p(U)$ and a known power curve $P_w(U)$ the wind turbine power P_w is equal to:

$$P_w = \int_0^\infty P_w(U)p(U)dU \tag{2.2}$$

If the power curve is not available, the power can be theoretically obtained as follows.

The wind with a mass *m* and a speed *V* has a kinetic energy of:

$$E_c = \frac{1}{2} \cdot m \cdot v^2 \tag{2.3}$$

The air has a density of around ρ = 1.23 kg/m^3 for standard conditions. So the mass of air going through the wind turbine, with a swept area A, each second is:

$$Mass/sec = v \cdot A \cdot \rho \tag{2.4}$$

The power is energy per second. So its formula is:

$$Power_{wind} = \frac{1}{2} \cdot A \cdot \rho \cdot v^3 \tag{2.5}$$

To pass on electrical power we use the following formula:

$$Power_{el} = \frac{1}{2} \cdot \rho \cdot A \cdot C_p \cdot v^3 \cdot N_g \cdot N_b \tag{2.6}$$

where:

- C_p is the coefficient of performance;
- N_g is the generator efficiency;
- N_b is the gear box bearing efficiency

2.3 Losses

The previous part showed how to get the gross energy production. It does not account for the different losses that occur in the park. The different losses that can happen are:

- Wake effect
- Unavailability losses
- Curtailment losses
- Environmental losses
- Electrical losses

2.3.1 Wake effect

The wake effect or array losses, occur when turbines are downstream to another. The wind turbine harness the kinetic energy from the wind, therefore a wind turbine downstream will receive less energy. This effect disappears after a certain distance as the wind returns gradually to its stable state. Therefore in a wind farm siting, it is important to position the turbines in order to limit the wake effect. They can range between 0 % to more than 10 % (for badly designed layouts).

Those losses depend on the number of wind turbine and their distance to each others. It is recommended to align them perpendicularly to the dominant wind if possible. The wake effect can be calculated in WindPRO and WAsP (presented section 3.4). The N.O Jensen model is used, it models the wind as presented on figure 2.4.

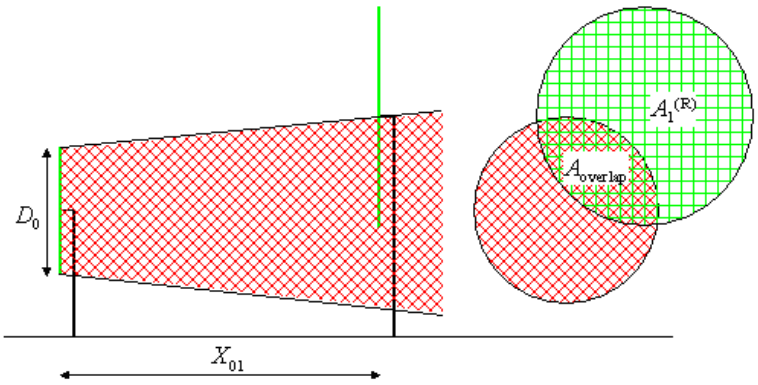


Figure 2.4: Wind used in the N.O. Jensen model, from WASP 2019

The wind losses, δV_{01}, at the second wind turbine (named A_1 here) are given by the equation 2.7

$$\delta V_{01} = U_0(1 - \sqrt{1 - C_t})(\frac{D_0}{D_0 + 2kX_{01}})^2 \frac{A_{overlap}}{A_1^R} \tag{2.7}$$

- U_0 is the wind speed at the first wind turbine
- D_0 is the diameter of the first wind turbine
- C_t is the drag coefficient
- X_{01} is the distance between the two wind turbines
- *k* is the Wake Decay Coefficient, that can be fixed by the user and depend on the type of terrain surrounding the turbines.

2.3.2 Unavailability losses

Unavailability losses are due to downtime of the wind turbines that can be caused by maintenance operations, technical incidents or due to power grid downtime that can happen a few times a year. They are usually estimated and can be improved with knowledge of the turbines, the reactivity of the maintenance service and the frequency of downtime events on the power grid. Table 2.1 presents the different availability reported by some of the main manufacturers of wind turbines.

Manufacturer	**Availability reported**
ENERCON	98,5% (Enercon 2007)
General Electric	98 % (General Electric 2020)
Goldwind	98 % (GoldWind 2020)
NORDEX	97,7 % (Nordex 2018)
SIEMENS GAMESA	98 % (SIEMENS GAMESA 2019)
VESTAS	98,4 % (Nielsen 2012)

Table 2.1: Availability of the wind turbine from different manufacturers

2.3.3 Other losses

- **Curtailment losses**: usually due to restrictions concerning the noise impact or the protection of the biodiversity, eg. bird or bat protection. The noise curtailment and biodiversity protection can lead to losses between 2 % and 3 %.
- **Grid unavailability**: some failures can happen to the grid to which the wind turbines are connected. Consequently, the electricity cannot be transferred to the grid and is lost. Those losses are usually small, around 0.75 % (Martin et al. 2016).
- **Electrical losses**: occur in the electric components such as cables or transformers. They are usually around 0.5 %. Another electrical loss is the park consumption. Each turbines consumes electricity for the transformers and for heating. This losses can go up to 2.5 %.

2.4 Uncertainties in annual estimation

The wind estimation needs to account for the different sources of uncertainties. Not all of the inputs are certain and the methods used above do not aim to know exactly the future production but the most likely estimation. A standard deviation is given to some parameters used in the calculation of the production. Then the general uncertainty is calculated as presented in equation 2.8.

$$\text{Global uncertainties} = \sqrt{\text{Uncertainty}_1^2 + \text{Uncertainty}_2^2 + \ldots} \tag{2.8}$$

Here are presented some sources of uncertainties and their usual standard deviation:

- **Wind data quality**: comes from the uncertainties due to the equipment calibration, it is around 2 %-5 % if the data is coming from wind measuring devices, called anemometers, that are calibrated or not (EMD International A/S 2013). If a reanalysis is used instead of a real wind data, WindPRO manual recommends to use a 20 % uncertainty of the result;
- **Annual variability**: comes from the variability of the wind from one year to another. Its value changes from one site to another, but a general value can be 6 % of uncertainty (EMD International A/S 2013);
- **Wind modelling**: the different extrapolations of the wind done in the calculation model add uncertainties to the project. For example WindPRO advises for 1 % of uncertainties for each 10 m between the mast's height and the wind turbine's height when a vertical extrapolation is performed (EMD International A/S 2013). Also for a horizontal extrapolation, WindPRO recommends to use a 1 % uncertainty for each kilometer between the mast and the wind turbine.

The uncertainties are often taken as default values, but if extra data is available it can help to reduce them. For example, the production of a nearby park can help to estimate the wind profile more accurately or help test the hypothesis made in the wind power estimation.

2.5 Summary

Chapter 2 of this thesis presented the different steps to develop a wind farm and how to estimate the production of it. A lot of uncertainties can hinder the accuracy of the estimation made, which can lead to a inefficient park or a refusal of the project from the investors. Therefore, a better estimation of uncertainties have to be made. Wind speed estimation has to be better and one major factor of uncertainties is the use of reanalysis as it will be explained in Chapter 4. But first, the wind speed itself and its influence at a microscale is discussed.

CHAPTER 3

Wind analysis

This chapter takes a closer look on wind in the nearest layer of the atmosphere, called the boundary layer, and presents the impact of two parameters of the terrain: roughness and orography. Then, two models allowing to extrapolate wind speed at different heights are presented: logarithmic law and power law. Their validity in the literature is displayed. Finally, explanations of the wind measurement devices and the wind modeling softwares, WindPRO and WAsP, are made.

3.1 Generalities

Meteorological phenomena happen in the troposphere, the lowest layer of Earth's atmosphere. The air does not have the same temperature, pressure or density at all points, which creates wind flows. Cold air is denser so it sinks which is a phenomenon called anticyclone. Conversely, hot air rises upward and it provokes a depression. The flow naturally makes the anticyclone willing to fill the depression. Additionally, the Coriolis force, due to earth's rotation, provokes a deviation of the wind flows: in Europe the wind leaves the anticyclone by spinning clockwise and enters the depression anti-clockwise.

The troposphere represents a 13 km layer agl. There are two layers of interest for wind analysis that are presented figure 3.1:

- **Geostrophic winds** are the product of temperature and pressures differences. Their wind speed is not affected by the terrain and are constant with increasing altitude. They are at around 1000 m agl depending on the terrain;

- **Wind of the boundary layer** are influenced by the ground topography and roughness. They are at around 0 m to 200 m agl.

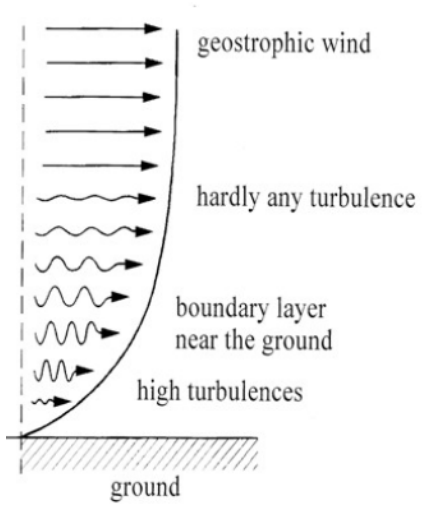


Figure 3.1: Geostrophic and boundary layer wind, from NTNU 2016

3.2 Wind change at micro scale

The following part aims to present the wind flow modifiers at a microscale, ie. a few meters of precision, and in the boundary layer. It does not present in-depth equations, such as Navier-Stokes', as it does not fall into the scope of the thesis. First the roughness and the orography impact on the wind at a micro scale is presented as they are two important factors in modifying the wind. Then two wind profiles that use the power law and the logarithmic law are displayed.

3.2.1 Roughness

The roughness of a given area is defined by the size and distribution of the different roughness elements present on the site. Roughness elements are all the elements that can hinder the wind's flow, eg. trees, cities, fields or hedges.

To quantify the level of a roughness element, it is common to use the empirical formula of the European Wind Atlas (Troen and Lundtang Petersen 1989). A roughness element is described by 3 parameters: its height h, its cross-section facing the wind S and the average horizontal area A_H. Then we have the relation 3.1 between the roughness element and the roughness length z_0. The roughness length is usually the parameter used to describe the roughness of a given area.

$$z_0 = 0.5 \cdot \frac{h \cdot S}{A_H} \tag{3.1}$$

To observe the effect of the roughness on the wind, the figure 3.2 was computed by the Global Wind Atlas (DTU Wind Energy and World Bank Group 2018). A terrain with more roughness slows down the wind flow more. Consequently the roughness can have an effect on a site that is a few kilometers away (Danish Wind Industry Association 2003).

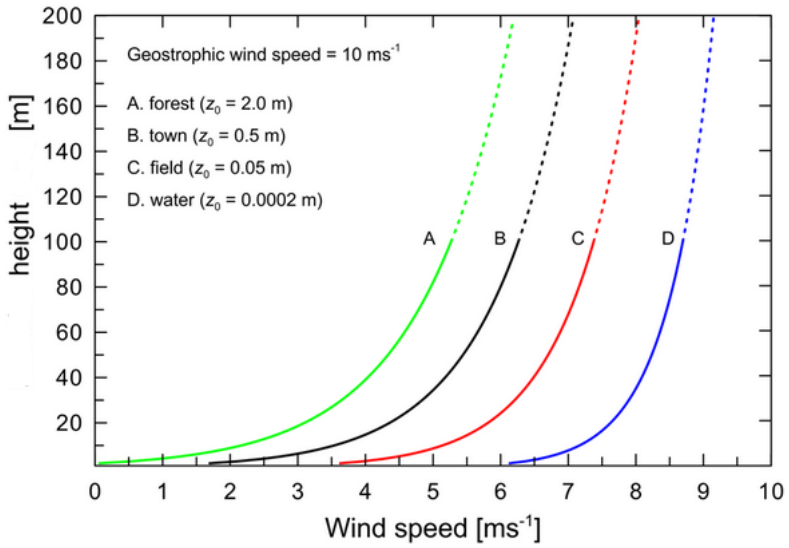


Figure 3.2: Effect of the roughness on wind speed

3.2.2 Orography

3.2.2.1 Definition

Orography is the study of the topographic relief of the terrain. It is also an important input of the wind estimation. Variations of the terrain change the wind flow properties. For example, the air going toward a rounded hill will be compressed and therefore will accelerate as it moves across the top of the hill, as shown in figure 3.3, the height of the maximum speed-up (l) is related to the geometry of the hill (L) (DTU Wind Energy and World Bank Group 2018). Depending on the slope and the height of the hill, the wind can become separated and turbulent.

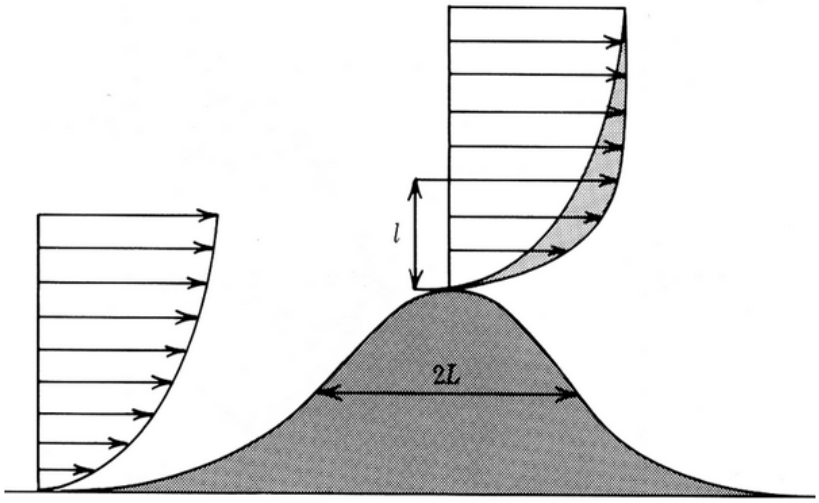


Figure 3.3: Visualisation of the impact of orography on a vertical wind profile upwind and on top of a wind, from DTU Wind Energy and World Bank Group 2018

3.2.2.2 Ruggedness index

The Ruggedness Index (RIX) is defined as ”the fractional extent of the surrounding terrain which is steeper than a certain critical angle” (Mortensen, Tindal, and Landberg 2008), ie. it is the percentage of terrain that has a slope exceeding a certain value, called the critical angle θ. The RIX is used to assess the orography's complexity of a terrain in the experimental study. Figure 3.4 presents a visual representation of the RIX. In the experimental part of the report, a critical angle of 10 % is used. For this value, the higher the RIX, the more hills or complex terrain are present in the area. In a radius of 1.5 km, a RIX of 20 % would mean that 20 % of the terrain in the radius has a slope exceeding 10 %.

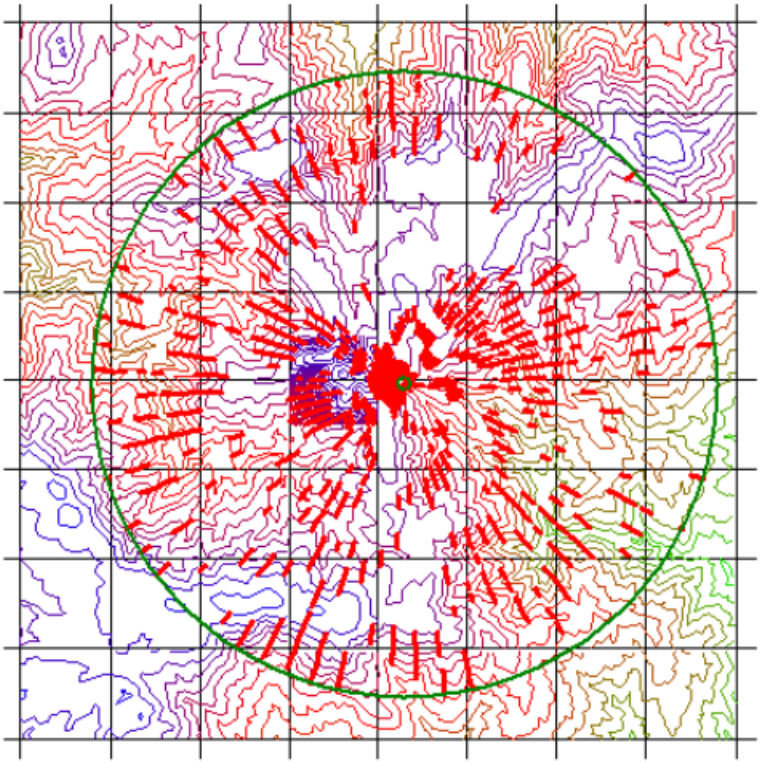


Figure 3.4: Representation of the RIX on a given terrain, each red line represents the terrain steeper than a critical angle θ, taken from Mortensen, Tindal, and Landberg 2008

3.2.3 Wind Profile

Roughness and orography have an influence toward the wind. Usually, the wind speed is small near the surface and increases with the height. But the rate of this increases changes a lot and it defines the wind profile. Different formulas were presented in wind theory to estimate the wind profile, also called the wind shear. Here two wind profiles commonly used in the wind studies are presented: the logarithmic law and the power law.

3.2.3.1 Logarithmic law

Logarithmic law's formula give the wind speed U at an altitude z

$$U(z) = \frac{U*}{k} ln(\frac{z}{z_0}) \tag{3.2}$$

Where $U*$ [m/s] is a constant called the friction velocity and k is the von Karman's constant. It is usually admitted that this formula is more relevant for low heights, ie. between 0 m to 50 m. From knowing one wind speed at one height and one roughness

length z_0 the wind speed can be deduced at another height by dividing the equation 3.2 by itself, to get the equation 3.3. If the roughness length is not known, tables such as Figure 3.5 can be taken.

$$\frac{U(z_1)}{U(z_2)} = \frac{ln(\frac{z_1}{z_0})}{ln(\frac{z_2}{z_0})} \tag{3.3}$$

Terrain description	z_0 (mm)
Very smooth, ice or mud	0.01
Calm open sea	0.20
Blown sea	0.50
Snow surface	3.00
Lawn grass	8.00
Rough pasture	10.00
Fallow field	30.00
Crops	50.00
Few trees	100.00
Many trees, hedges, few buildings	250.00
Forest and woodlands	500.00
Suburbs	1500.00
Centers of cities with tall buildings	3000.00

Figure 3.5: Values (approximate) of surface roughness length for various types of terrain, from Manwell, McGowan, and Rogers 2009

3.2.3.2 Power law

The power law's formula for an height z is:

$$U(z) = cz^{\alpha} \tag{3.4}$$

c is a constant and the α exponent is dependent of multiple parameters. Among them are the elevation, the time of day, season, nature of the terrain, wind speed, temperature and various thermal and mechanical mixing parameters (Manwell, McGowan, and Rogers 2009). Tables are available to estimate α and easily assess the wind speed. This profile is recommended for higher heights. By already knowing one wind speed at one height, all wind speed can be deduced as such:

$$\frac{U(z_1)}{U(z_2)} = (\frac{z_1}{z_2})^{\alpha} \tag{3.5}$$

3.2.3.3 Validity

Those models have multiple interests in the wind analysis community. They can be used to interpolate or extrapolate the wind speed at a desired height. Indeed, the measurement devices usually do not directly measure the wind speed at the wind turbine's

height for monetary reasons (the measurement mast price is linked to its height), but instead a few meters below. For example, in the experimental part of the thesis, wind speed needs to be interpolated at the right height as the reanalysis data give wind speed at a different height. The fact that they are still used for wind estimation show that they can provide a good enough estimation if the computer power and time available are taken into account .

However Elkinton, Rogers, and McGowan 2006 compared 11 measurement data sets in the United States for terrain that were either flat with no trees, hilly with no trees or forested and used the two models to extrapolate the wind to a known height. Flat terrain had a difference between modelled and measured wind speed of around -1%, the hilly terrain had a difference between 2% and 12%, and forested terrains had a difference between -0.7% and 6%. The authors conclude that both models, logarithmic and power law, give accurate results for flat sites. But they point out that the likelihood that either type of model gives a prediction close to the measured value at a site with complex terrain (hill or forest) is small.

3.2.4 Seasons

For seasonal bias, it is generally admitted that in Europe, the winds are stronger in the winter and weaker in the summer. Benedict Jourdier 2015 compared France's wind speed depending on four sections of the year. It appeared that in the winter and summer there was a significant variation of the wind speed. The main reasons are the change of wind regime throughout the year and on a smaller scale the change of vegetation, loss of leaves for example. Indeed the vegetation in winter generally has a smaller roughness length because the wind is less slowed down than in summer by the trees.

3.3 Wind Measurement

The following part describes the measurement devices on a measurement mast installed on a potential site for a wind farm, as well as a closer look on cup anemometers that measure the wind speed and that are used in the experiment part to record the wind data used.

3.3.1 Measurement mast

Measurement devices are used on site to estimate the wind potential of a limited area. Measurement masts are set up to assess the wind. It usually has the following types of meteorological sensors:

- Anemometers to measure wind velocity;
- Wind vanes to measure wind direction;
- Thermometers to measure the ambient air temperature;
- Barometers to measure air pressure.

A data-logger is in charge of providing the data, it records it with a 1-minute step-time and provides an average on 10 minutes. The mast has to be installed near to the investigated site, within 2 km of the wind turbines if the terrain is complex or within 10 km if the terrain does not have nearby complex terrain (MEASNET 2016). It is recommended to have its height at least 2/3 of the hub height considered, less would include too much uncertainties. As we have seen in section 3.2.3, wind profile can vary greatly over a few kilometers and long extrapolations could lead to errors.

Other measurement devices are also used: Lidar (Light Detection And Ranging) and Sodar (Sonic Detection And Ranging). They are more precise than a measurement mast as they can measure wind at higher heights, but they are more costly and are more prone to damage. As they are not used in the experimental part they are not looked on further.

3.3.2 Cup Anemometers

Cup anemometers are the most common anemometers used on measurement masts. They are usually more robust and less expensive than the alternatives. They are mounted on a vertical axis that can rotate freely. Cup anemometers' rotation varies with the wind speed and generates a signal. This signal in hertz is converted to m/s. It has to be calibrated before and after the measurement period in order to limit errors (MEASNET 2016). Usually 3 or 4 are mounted on the mast to have a better assessment of the wind. The advantage provided by having several anemometers is to observe the wind profile. The highest anemometer is set up at the very top of the measurement mast or it is positioned in order to be perpendicular to the dominating wind.

Other anemometers can also be used. Among among them are the ultrasonic anemometers, which were not used in the experimental part so it was chosen not to dwell on that technology.

3.4 WindPRO and WAsP software

Softwares are used to estimate the production of a wind farm, WindPRO and WAsP are two of them. They work together and allow to do production estimation rapidly.

3.4.1 WAsP

WAsP (Wind Atlas Analysis and Application Program) is a software commercialized by the Danish Risø National Laboratory and often used in the industry in order to predict wind climate and wind energy resources. It follows the recommendations and methods advocated in Troen and Lundtang Petersen 1989. It uses simplifications and is rather fast to compute and simple to handle. It contains simplifications from the Navier-Stokes equations generally used to compute flow and it uses semi-empirical linearized equations (Bayon-Barrachina et al. 2014). The wind flow model used by WAsP is the BZ model by Troen in 1990. It uses a zooming polar grid for the terrain description, where the grid has its finest mesh at the center. Therefore it is important to have an accurate and detailed description of the terrain's roughness and orography, especially near the wind turbines.

3.4.2 WindPRO

WindPRO is a module-based software developed by EMD International A/S (Energi- og MiljøData, a Danish-based company). It uses WAsP for some of the wind flow modeling. It can also be used for wind project design, annual power production and uncertainty evaluation. The module-base system allows the purchase of modules with different purposes, allowing the user to choose them in agreement with its needs. The modules used in the experimental part are :

- **METEO module**: allows to import an analyze wind data files;
- **PARK module**: used to make the power calculation. It needs the wind as input as well as the terrain's data. The wake effect is taken into account. The wakes are assumed to expand linearly with the distance from the turbine and the rule for overlapping wakes is simplified according to Rathmann et al. 2006;
- **Loss & uncertainties module**: accounts for all the other losses, eg. mechanical or electrical. They are entered manually. The uncertainties are also entered manually or WindPRO has help to estimate some of them, eg. for the distance and height extrapolation.

3.4.2.1 Downscaling

The main feature used in the experimental section is the downscaling process, or the horizontal extrapolation. It is a feature available in the current version of WindPRO used, version 3.3. Its goal is to move mesoscale wind data, eg. reanalysis, to the microscale terrain by taking into account its specificity. Indeed, it was shown in the section 3.2 that roughness and orography are important to describe the wind in the boundary layer. The speed-ups due to terrain in the microscale model are different to

the ones due to the mesoscale model because the mesoscale model uses mesoscale terrain data that are less precise than the one of the microscale model. This is illustrated in the figure 3.6 where the grid represents the accuracy of the mesoscale terrain.

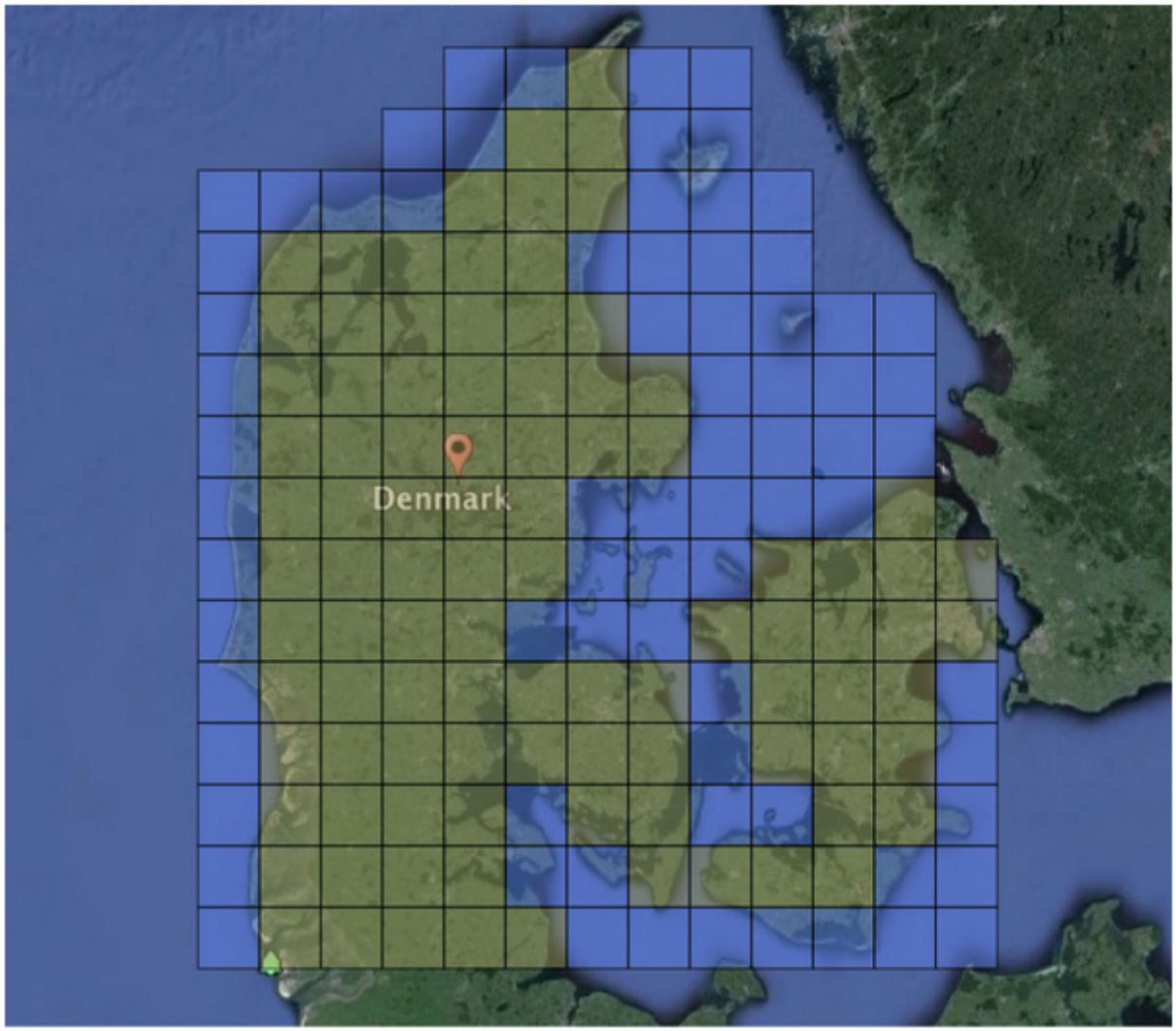
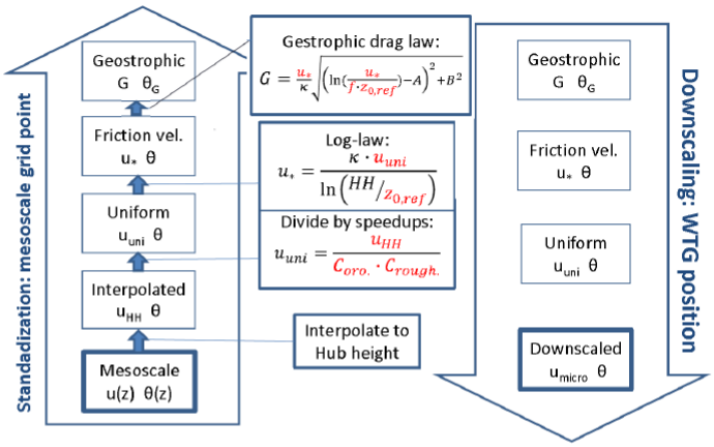


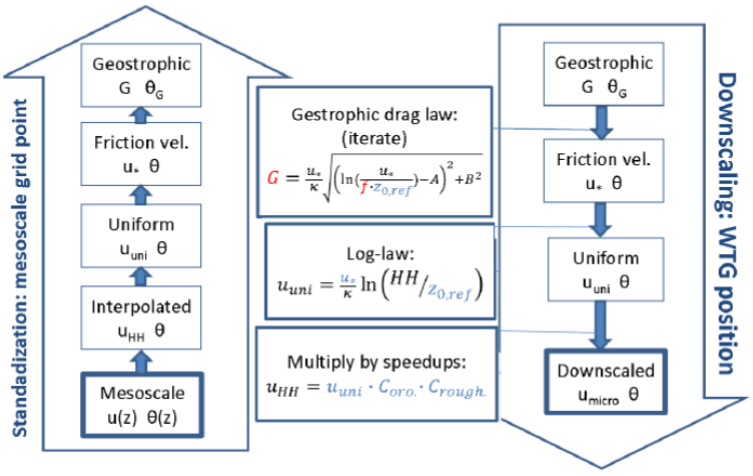
Figure 3.6: Illustration of mesoscale terrain, from DTU Wind Energy and World Bank Group 2018

One assumption made by the downscaling method is that on a micro-scale, wind change is only affected by terrain specificity, ie. orography and roughness. Therefore, only changes depending on these parameters on both location are made, and *a priori* it does not depend on the distance between the two places. It is a several step process. Figure 3.7a and 3.7b represent the equations and steps used in the WindPRO software in order to downscale mesodata to a microscale terrain. First, at the location A, the wind data is risen to the geostrophic condition, ie. the terrain parameters are removed from the wind and then the logarithmic law is used to calculate the wind at

a geostrophic level. Then this data is downscaled to the desired height, by using a logarithmic law with the parameter z_0 corresponding to the new site and the impact of the orography and the roughness are taken into account to obtain the wind speed and direction at the location B. WindPRO does not give more information about that method and the precise equations used and no further explanation was found in the literature.



(a) First part



(b) Second part

Figure 3.7: Summary of the equations used by WindPRO and WAsP to downscale mesoscale wind data to the microscale, taken from the WindPRO manual (EMD International A/S 2019)

3.4.3 Limitations

WAsP is limited by the fact that it does not take into account the non-linear wind flow used. According to Jothiprakasam 2015 and Lars Landberg et al. 2003 WAsP is not equipped to handle complex terrain, as it was first developed for flat terrain or hills with moderate slopes. Complex terrain is defined as having a slope exceeding 30% over a significant area by Jothiprakasam 2015. WAsP problem is that steep terrain induces change in wind flow, which may include flow separation at abrupt change in slope, vertical wind and recirculation behind cliff. Those changes are not taken into account. In addition, WAsP overestimates the speed-up at the top of hills when the input data is a wind measurement on a near flat terrain. The WAsP model also ignores the effect of thermal stability and temperature gradients, still according to Jothiprakasam 2015. An alternative is to use computational fluid dynamics (CFD) to make an accurate modelling of the terrain using Navier-Stokes equations (Lars Landberg et al. 2003). However this method is more time-consuming and requires more powerful computers.

CHAPTER 4

Wind reanalysis

In this chapter, the concept of reanalysis is explained. The two classes of reanalysis are presented: global reanalysis that have a 20 km to 100 km precision and the meso-scale reanalysis that have a 3 km precision. For both cases their accuracy in the literature is reported.

4.1 Description

In meteorological forecasts, real life measurements (from satellites, aircrafts, weather balloons and surface stations) are assimilated, by a prediction model, to create a complete description of the atmosphere on a grid. It is used as a starting point in numerical prediction of the weather. The prediction model uses a complex mathematical model that assimilate the observations that can be from different points in time, incomplete and possibly containing errors. Then the model does a short prediction (between 6 hours to 12 hours) that will be the starting point of the next cycle of assimilation. Those are called $analysis$.

The choice of the prediction model is important as it is based on fluid mechanics equations and take into account phenomena such as convection, diffusion and boundary layer interactions. Those allow to extrapolate the weather data and the wind.

Unfortunately, the assimilation model always improving from one iteration to the other so the analysis are not homogeneous from one time to another. The reanalysis data are then created by the same method as the analysis but with a fixed model. This allows to have coherent data on a long term scale. They have several parameters available at different heights that are useful for wind analysis (see section 3.3.1):

- Wind speed;
- Wind direction;
- Temperature;
- Pressure

Figure 4.1 displays the spatial resolution of some reanalysis data sets that are presented in the following section.

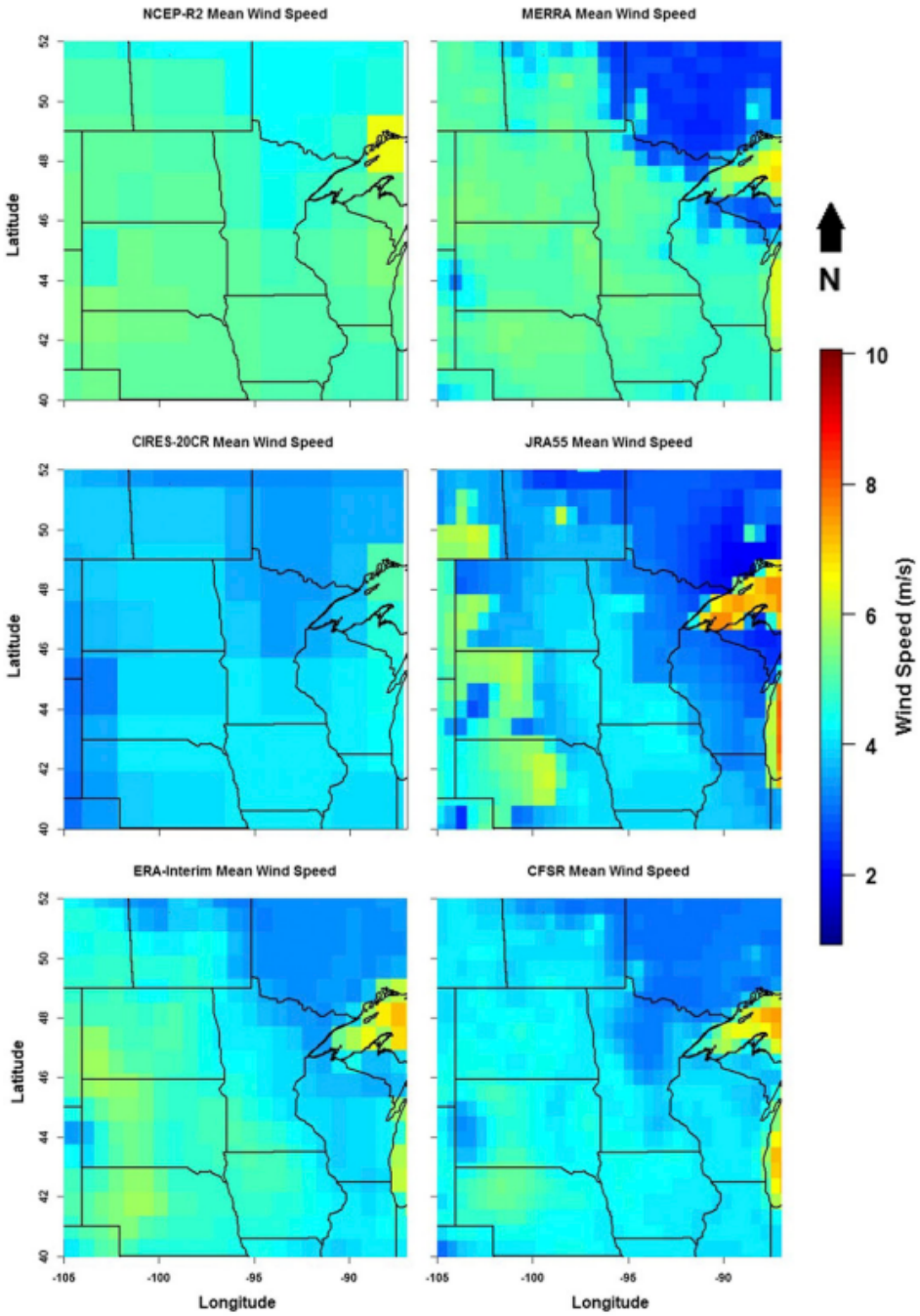


Figure 4.1: Maps of the average 10-m wind speeds (agl) over the 1979-2016 period for the upper Midwest, taken from Coburn 2019

4.2 Global reanalysis

There are several reanalysis data sets available. Most of them can be found on the WindPRO interface. Each uses different models of atmospheric flow and different data for the terrain. Here are presented the global scale reanalysis, ie. that has a precision between 20 km and 100 km.

- **MERRA-2**: The Modern Era Retrospective Analysis for Research and Application is modelled and distributed by the Global Modeling and Assimilation Office of NASA. A lot of observations from different data sets are assimilated by the GEOS-5 system (Goddard Earth Observation System Version 5) (Bosilovich et al. 2015). It has hourly values since 1992 until 2020. It gives wind speed and direction at 10 m and 50 m. It has a spatial resolution of around 55 km in latitude and 69 km in longitude.
- **CFSv2**: The Climate Forecast System (CFS) is developed at the Environmental Modeling Center at NCEP (National Centers for Environmental Prediction) and was released in 2011. A data assimilation algorithm based on GSI (Gridpoint Statistical Interpolation) is used to integrate a lot of observations in 6 hours time steps. It has hourly temporal resolution. It has a spatial resolution of around 22 km. It is an extension of the previous reanalysis data set CFSR since 2011. The height available is 10 m agl.
- **ERA-Interim**: ECMWF Re-Analysis Interim dataset is distributed by the ECMWF (European Centre for Medium-Range Weather Forecasts) and by its service the Copernicus Climate Change Service (C3S) (ECMWF 2019). It has only a six-hour time resolution and has an 80 km spatial resolution. It stopped being produced the 31st of August 2019 in order to be replaced by ERA-5.
- **ERA-5**: ECMWF Re-Analysis 5 is a climate reanalysis dataset developed by C3S. It replaces the previous ERA dataset ERA-Interim. This dataset has a spatial resolution of 31 km and a time resolution of 1-hour step.

4.3 Accuracy of global reanalysis

The accuracy of reanalysis data has been investigated due to their usefulness for wind power estimation: improvements in their accuracy allow a better estimation and therefore better sizing for wind farms projects. Reanalysis are modeled and not real measurements so they can contain a bias or deviation from the reality. Currently ERA-5 is considered one of the most accurate global reanalysis dataset as it is very recent (2016-2017) and it has been developed to replace ERA-interim, an already precise reanalysis dataset (Haxsen 2017). This part presents the state-of-the-art regarding the reanalysis evaluation of wind speed.

The values used for evaluating the reanalyses usually taken into account are bias of the wind speed's reanalysis, the correlation coefficient (Pearson's) and RMSE (root mean square error). The studies were selected depending on the dataset used, ie. preferably the ones cited above and the geographical areas investigated, ie. Europe or the Northern hemisphere. Indeed other countries might have specific weather conditions that can help or hinder the accuracy of the reanalysis but would not be likely to happen in Europe or France.

ERA-Interim, ERA-5 and MERRA-2 were compared in EMD International A/S 2019 to data from 108 measurements masts scattered around the globe. The mast's heights are between 60 m and 140 m agl. When replacing MERRA-2 with ERA-5, it resulted that the mean of the correlation coefficient, between the reanalysis data and the measurement data, went from 0.71 to 0.78 and the standard deviation went from 0.12 to 0.10.

Coburn 2019 investigated the accuracy of different reanalysis sets, notably ERA-Interim, CFSR and MERRA-2, comparing them to measurement from weather stations at 10 m agl. The results were compared annually and on the four seasons. The study takes place in the United States. The results for the concerned reanalysis data sets are summed up in the table 4.1 where the bias between the measurements and the reanalysis are calculated as well as the correlation coefficient (see section 5.1.2 for equations). A positive bias means that the reanalysis overestimates the real wind while a negative bias means that there is an underestimation. The coefficient of correlation quantifies how much the two data sets are varying in the same ways, it is a value between 0 and 1. A correlation coefficient of 1 means that the two data sets are the same. ERA-Interim has the smallest bias and MERRA-2 the best correlation overall. The month analysis showed more bias in the winter. Spring and summer are two seasons with the less bias.

Benedict Jourdier 2015 compared the MERRA-2 and ERA-Interim reanalysis to weather stations over the french territory. According to her, the mean bias is mostly positive for MERRA-2 and ERA-Interim except in some areas in the south of France as shown in figure 4.2. In the center and west part of France, regions of our study, the reanalysis bias with the weather station is between -0.2 m/s and 0.4 m/s. Those results are coherent with the ones made by Coburn 2019. Regarding seasonal variation, Benedict Jourdier 2015 showed that, in France, the MERRA-2 reanalysis overestimated the wind speed during all the months but specially during winter. ERA-Interim showed a more constant positive bias through the year.

Full period	Bias (m/s)	Correlation coefficient
ERA-Interim	0.127	0.85
MERRA 2	0.415	0.89
Spring		
ERA-Interim	- 0.154	0.86
MERRA 2	0.344	0.9
Summer		
ERA-Interim	0.135	0.83
MERRA 2	0.297	0.87
Autumn		
ERA-Interim	0.16	0.88
MERRA 2	0.459	0.91
Winter		
ERA-Interim	0.374	0.84
MERRA 2	0.563	0.89

Table 4.1: Summary of the results obtained in Coburn 2019

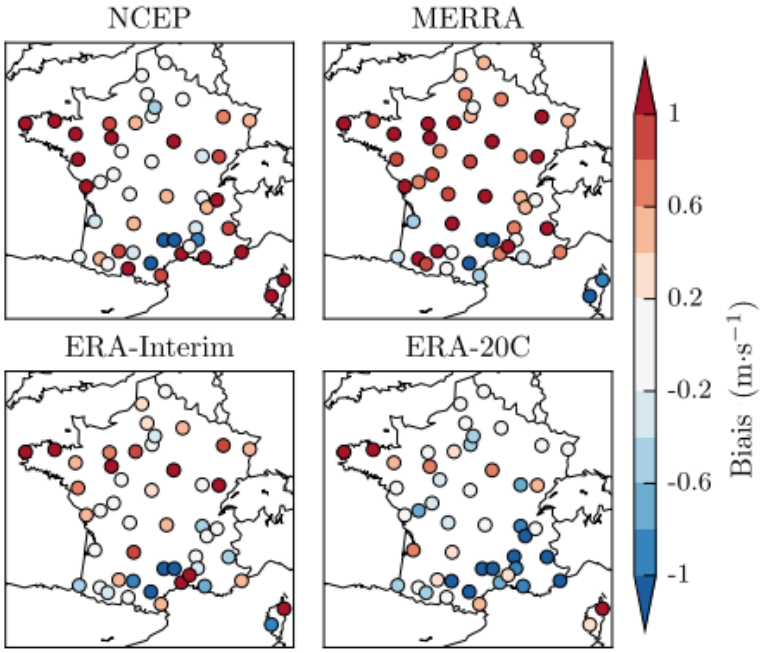


Figure 4.2: Bias between reanalysis datasets (MERRA, ERA-Interim, NCEP and ERA-20C) and weather station measurements (10 m agl), taken from Benedict Jourdier 2015

Another paper, Sharp et al. 2015, analyzed the accuracy of the CFSR reanalysis. Their bias with the wind speed observation MIDAS (Met Office Integrated Data Archive System, United Kingdom) are between -8 m/s to 5 m/s with a mean at 0.35 m/s. The bias of wind speed is evenly distributed in the positive and negative directions. The authors also point out that offshore sites have less extreme bias and it could be due to the homogeneity of the wind resources or because the offshore wind speeds are not as high as the windiest onshore ones. This paper pointed out that CFSR had larger errors with the altitude increasing, because CFSR had trouble to represent accurately the wind speeds above a certain threshold.

4.4 Studied meso-scale reanalysis: EMD-CONWX to EMD-WRF

The previous global reanalysis have the default to be on a not dense enough spatial grid, ie. their spatial resolution is between 22 km and 80 km. Also for the ERA-Interim dataset the time steps are too spaced, 6 hours. This default can impact the accuracy of a wind power estimation. The Measnet recommendation (MEASNET 2016) for a good wind power estimation regarding distance mast-wind turbine, during the micrositing phase, is 2 km for a complex terrain and 10 km for a simple terrain. The same scale is reasonable to hope for the reanalysis wind analysis. Therefore mesoscale reanalysis were created, that aims to have a denser spatial grid and better time resolution. This part presents two mesoscale reanalysis EMD-ConWx and EMD-WRF.

EMD-ConWX is the meso-scaled reanalysis developed in collaboration between EMD and ConWx, an expert in mesoscale modelling. It has the ERA-Interim reanalysis as input for the wind data. The spatial resolution goes from 80 km (ERA-Interim) to 3 km (EMD-ConWx). The temporal resolution goes from 6h steps to 1 hour steps. The heights available are from 10m to 200m.

EMD-WRF Europe+, or **EMD-WRF** is the new reanalysis from EMD, it began to be available in the summer 2019 in Europe on the WindPRO interface. It is stated on the EMD wiki (EMD 2019) that it is designed to replace the EMD-ConWx dataset. Indeed, it was modelled with the reanalysis ERA5. It has a spatial resolution of 3km and hourly time resolution. The heights available are from 10 m to 4000 m.

EMD-ConWx was investigated in Haxsen 2017. It was compared to Sodar (Sonic Detection And Ranging) measurements, in Germany. It allowed to have heights compared from 50 m to 200 m. On average, EMD-ConWx data overestimated the mean wind speed by 0.63 m/s at 200 m agl. up to 1.54 m/s at 50 m agl. The average wind speed deviation was 1.02 m/s and the standard deviation of the mean bias is 0.70 m/s. The average RMSE for all hourly wind speed data was of 1.91 m/s. The coefficient of correlation of the two data was 0.813 on average. Interestingly, the best site is one

Height [m]	Bias [m/s]	RMSE [m/s]	STD of bias [m/s]	Correlation
50	1.54	1.73	0.89	0.77
75	1.18	1.77	0.66	0.806
100	0.98	1.85	0.55	0.818
150	0.78	2.02	0.46	0.827
200	0.63	2.15	0.41	0.837
Average	**1.02**	**1.91**	**0.70**	**0.813**

Table 4.2: Summary of the results obtained in Haxsen 2017, STD: Standard deviation

with the flat terrain, with an average bias of 0.17 m/s, while the worst site is described as having a "forested and complex terrain", with an average bias of 2.51 m/s.

In order to improve the accuracy of the reanalysis, the author proposed to make a height shift of 50 m, ie. instead of taking the wind speed at 150 m on the reanalysis, the wind speed of the reanalysis at 100 m is chosen to be compared to the measurement data at 150 m. The results are presented in table 4.3. The bias mean value decreased to 0.18 m/s. Moreover it was noticed that wind speed between 0 m/s and 8 m/s had low overestimation. Wind speed between 8 m/s and 16 m/s had high overestimation below 100 m agl. The extreme wind speed, 16 m/s to 25 m/s have a high deviation as well. Concerning the correlation coefficient, it was found to be around 0.811 for hourly mean wind speed, it was deemed better than the global reanalysis range.

With this shift, a better mean bias is obtained which is close to the results obtained by Sharp et al. 2015 and Coburn 2019. But the measurements are from 50 m to 200 m agl. which is higher than the measurements devices used in those two papers, ie. weather stations at 10 m agl., so the bias might differ for this reason. Also, the results kept a important standard deviation, 0.6 m/s on average, it means that the results are spread a lot around the mean bias. If this method is used to do an estimation of production an important uncertainty will remain on the accuracy of the reanalysis used. It would be better to have the smallest standard deviation in order to reduce the uncertainty of calculation to a minimum.

Height [m]	Bias [m/s]	RMSE [m/s]	STD of bias [m/s]
50	0.63	1.57	0.85
75	-0.002	1.67	0.50
100	0.05	1.78	0.47
150	0.1	1.87	0.45
200	0.13	2.02	0.40
Average	**0.18**	**1.78**	**0.60**

Table 4.3: Summary of the results obtained in Haxsen 2017 with the vertical shift of data, STD: Standard deviation

Due to the recent release of the EMD-WRF Europe+ no independent study was found on its accuracy. Therefore this study is among the firsts to make a comparison of this dataset with real-life measurements.

CHAPTER 5

Reanalysis and measurement comparison

This chapter starts by explaining the method used to compare mast measurements to reanalysis data. The inputs used and the data handling are presented. Then the results are displayed. First, a general comparison is made between the measurement masts' wind speed and the reanalysis data. After, this comparison is made on a monthly basis to see if the accuracy is dependent on the season. Finally three parameters, altitude, height and coefficient of correlation between the data sets are studied to see if they are correlated to the reanalysis accuracy.

5.1 Method

This section presents the method applied in order to study the reanalysis wind speed. Each sites has to be set-up, ie. the terrain specificity have to be loaded into the WindPRO software. Then the reanalysis wind speed is moved from its location to the measurement mast place with the downscaling method, and it is interpolated to the correct height. After, the two data (wind speed from the reanalysis and from the mast) can be compared.

5.1.1 Inputs

This part is going to present the different inputs used in order to compare the wind speed of mast measurements to reanalysis data. Those are:

- Sites studied
- Reanalysis selection
- Terrain data: roughness and orography
- Method of interpolation: vertically and horizontally

5.1.1.1 Sites studied

ENCIS Environnement provided the measurement masts data from 23 sites for testing the reanalysis data accuracy, those data were anonymized for reasons of confidentiality and are presented in table 5.1. The data comes from cup anemometers. Twenty one are from measurement masts at different heights. Two, number 17 and 19, are from anemometers situated at the top of a wind turbine.

Site number	Height (m)	Length of recording (month)	Site number	Height (m)	Length of recording (month)
1	72	12	**13**	50	13
2	66	12	**14**	42	7
3	84	24	**15**	122	7
4	72.5	6	**16**	85	7
5	49	36	**17**	91.5	13
6	50	13	**18**	47	8
7	50	22	**19**	91.5	13
8	65	125	**20**	80	10.5
9	80	27	**21**	80.5	26
10	62	12	**22**	45	11
11	61	12	**23**	42	13
12	61	12			

Table 5.1: Height and length of recording for the different sites studied

The sites are located mostly in the West part of France and at the center of France. The measurements took place between 2006 and 2018 depending on the site. Their altitude are between 75 m and 500 m above sea level. Their height of the mast are between 42 m agl. and 122 m agl.

5.1.1.2 Reanalysis selection

The reanalysis EMD-WRF is available on a regular 3 km x 3 km mesh directly through the WindPRO interface, the data are available from 1999 to 2019. It was chosen to study the four reanalysis data coordinates nearest to the mast. They are named d1-d2-d3-d4 for the nearest to the furthest data point respectively.

5.1.1.3 Terrain data

The roughness data is loaded directly from the WindPRO interface. The dataset Corine Land Cover 2012 was deemed suitable for the experiment part. It has a resolution of 100 m and is available on the whole country.

For the orography, it was chosen to take the dataset called BD Alti. It represents each altitude separated by 75 m. The data was interpolated to a precision of 1 m, with the software QGIS, and its veracity was verified on maps of the IGN (*National institute of the geographic and forest information*) which are maps representing the different altitudes of France.

5.1.1.4 Extrapolation and wind profile

In most cases, the height of the mast was not directly available in the reanalysis dataset. In those cases, the wind was interpolated vertically with a wind shear matrix that uses the power law formula to calculated the wind speed at the desired height.

Each reanalysis data is situated at different coordinates than the measurement mast, so a horizontal extrapolation, ie. downscaling, is done directly by the WindPRO and WAsP software through the Meteo-analyzer function that downscale the mesoscale data to microscale using the process described in section 3.4.2.1.

5.1.2 Data handling

WindPRO offers to export the final wind data into a .txt files. Those data contains the wind speed of the reanalysis at the same location and the same height of the measurement height.

The data is then handled through a small Matlab code in order to get the following results:

$$\text{Percentage of increase} = \frac{V_{EMD-WRF} - V_{measurement}}{V_{measurement}} \tag{5.1}$$

$$\text{The bias} = V_{EMD-WRF} - V_{measurement} \tag{5.2}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(V_{EMD-WRF[i]} - V_{measurement[i]}\right)^2} \tag{5.3}$$

$$\text{Correlation coefficient R} = \frac{cov(measurements, reanalysis)}{\sigma_{measurements} \cdot \sigma_{reanalysis}} \tag{5.4}$$

The Matlab codes operate according to the following steps:

- The measurement mast data are averaged from 10-minutes steps to 1-hour step to match the temporal resolution of the reanalysis

- For each hour the percentage of increase, the bias are calculated
- Then an average is done for each month and for each years
- For each month the standard deviation and the correlation coefficient R are calculated

The correlation coefficient is a value between 0 and 1 that quantify the similarity between the two sets of data. From the literature study it was found that the reanalysis data can have a seasonal bias (Coburn 2019). The percentage of increase and the standard deviation that each months have are also calculated. Most of the measurements have at least one year of measurement period, 7 sites (numbers 4, 14, 15, 16, 18, 20 and 22) have between 6 and 11 month of data.

Other parameters have been chosen to be compared. The first one is the correlation coefficient R. Indeed, if the two dataset have a high coefficient of correlation R, then their variability is similar. It seems interesting to see if it means that the wind speed is less overestimated when the coefficient of correlation is increasing. The two other parameters are altitude of the site of measurement and the height of the measurement mast. In Sharp et al. 2015, it was noticed that wind estimation were worse at higher altitude, while Haxsen 2017 found that the bias of the reanalysis decreased when the measurement's height increased. It seems also coherent to verify if a trend is noticeable on the data provided.

To do so, the correlation coefficient R_2 is calculated between the percentage of increase obtained in each site and the three parameters corresponding to the sites. Its formula is given in 5.5. Each time two datasets are going to be used as input in this formula. For example for the parameter *correlation coefficient* R, the first dataset is the percentage of increase depending on the sites and the second dataset is going to be the coefficient correlation R depending on the sites since those parameters are calculated for each sites. It is done to see if a trend appear between those parameters and overestimation. If the coefficient of correlation R_2 is close to 1 it means that if the parameters increase then the overestimation also increases and that this parameter is correlated to the imprecision of the reanalysis data. If the coefficient of correlation R_2 is close to -1, it means that if the parameter increases then the overestimation decreases. If it is close to 0, it means that the parameter is not correlated to the accuracy of the reanalysis data.

$$\text{Correlation coefficient } R_2 = \frac{\text{cov(percentage of increase,parameter)}}{\sigma_{\text{percentage of increase}} \cdot \sigma_{\text{parameter}}} \tag{5.5}$$

5.2 Results

In this section, the results from the wind speed comparison are presented. First, the yearly accuracy of the reanalysis is displayed. Then, the monthly variations of the

reanalysis data accuracy is investigated. Finally, the correlation between the accuracy and three parameters, altitude, height and correlation coefficient R, is studied.

The results suggest that EMD-WRF reanalysis overestimates the wind speed by 18.6 % on average. Also, some months are more prone to overestimate (January, February, October, November and December) the wind speed. Lastly, the three parameters do not suggest to be correlated to the reanalysis data accuracy.

5.2.1 EMD-WRF accuracy

The EMD-WRF reanalysis has a clear trend to overestimate the wind speed. Figure 5.1a and figure 5.1b represent respectively the percentage of increased and the bias for the 23 sites. For each site, four types of markers represent the results for one of the four reanalysis location surrounding the measurement mast (noted from nearest to furthest d1, d2, d3 and d4 respectively). As an example, the results of the nearest reanalysis for the site 2 are presented in table 5.2, in figures 5.1a (22.32 %) and 5.1b (1.20 m/s) these results are represented by the blue cross (d1). For each month the mean wind speed of the measurement mast, the mean wind speed of the reanalysis data, their mean bias and their mean percentage of increase are presented.

Month	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sept**	**Oct**
MSM (m/s)	5.77	4.30	4.19	4.43	4.99	5.29	5.38
MSR (m/s)	7.15	5.06	5.05	5.49	6.06	6.68	6.51
MB (m/s)	1.37	0.77	0.86	1.06	1.07	1.39	1.12
MPoI (%)	23.81	17.85	20.49	23.96	21.52	26.31	20.88

Month	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**	**Average**
MSM (m/s)	5.58	5.88	6.23	5.66	6.35	5.34
MSR (m/s)	6.89	7.04	7.87	6.77	7.85	6.53
MB (m/s)	1.31	1.16	1.64	1.11	1.50	**1.20**
MPoI (%)	23.50	19.80	26.34	19.67	23.69	**22.32**

Table 5.2: Results for site 2 with the reanalysis location the nearest to the measurement mast - MSM: Mean speed mast - MSR: Mean speed reanalysis -MB: Mean bias - MPoI: Mean percentage of increase

Table 5.3 represents the averaged results for all the sites.

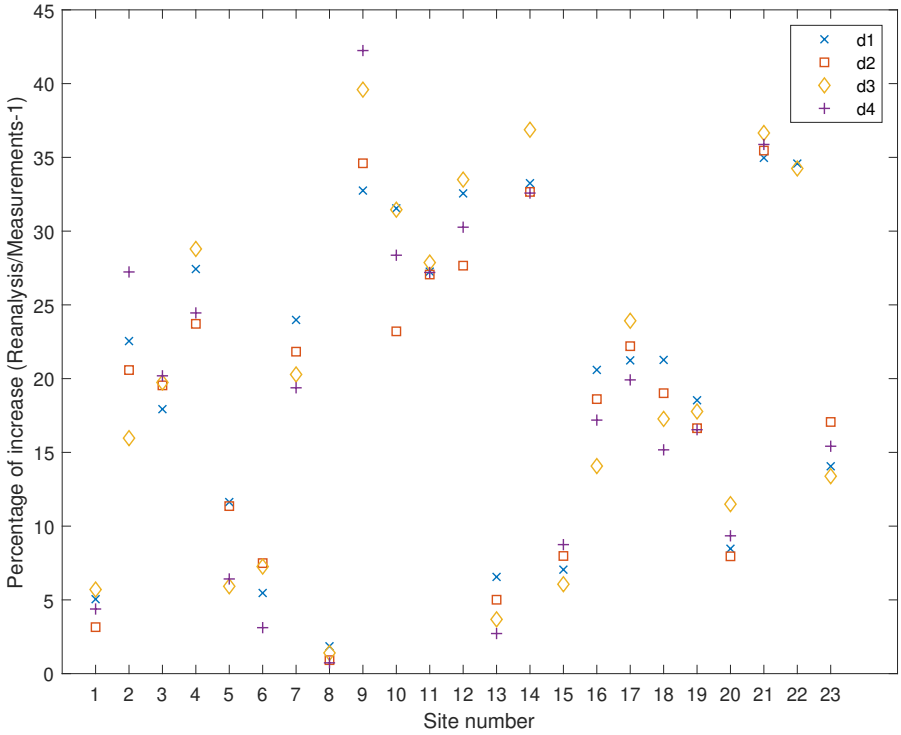
	Percentage of increase	**Bias**
Mean	18.6 %	1.02 m/s
Standard deviation	12.52 %	0.57 m/s

Table 5.3: Global results

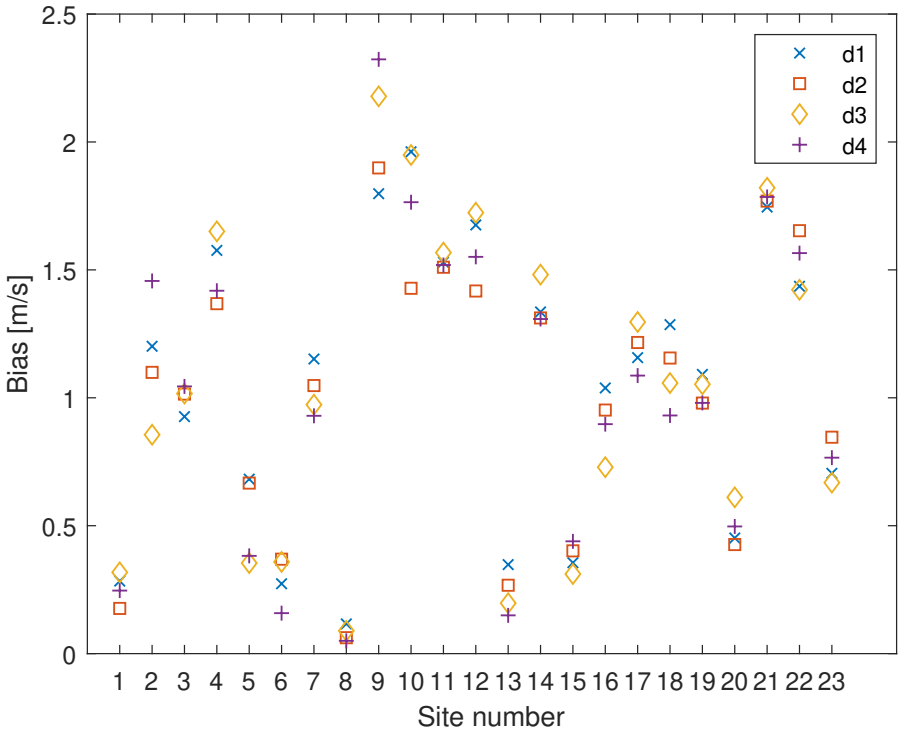
The results range from 0.73 % to 42 % from best to worst sites, with an average at 18.6 % ; and 0.04 m/s to 2.3 m/s for the bias, with an average of 1.02 m/s. It is noticeable that no reanalysis has a negative bias. However the standard deviation is quite high: at 12.52 % for the increase and 0.57 m/s for the bias, the results are scattered.

It is not always the nearest reanalysis, ie. d1 the blue cross, that would have been the best pick to estimate the wind. It was the case of 3 sites (3;9;21), the d2 was the best choice for 6 sites (1;4;10;11;12;20), d3 was the best for 6 sites (2;5;15;16;22;23) and d4 was the best for 8 sites (6;7;8;13;14;17;18;19).

In addition, the correlation coefficient between the EMD-WRF data and the measurements is calculated for all the reanalysis. The mean correlation coefficient is 0.79 with a standard deviation of 0.05 for the 23 sites.



(a) Percentage of increase



(b) Bias

Figure 5.1: Percentage of increase and bias of the reanalysis EMD-WRF on the 23 sites. d1-d2-d3-d4 represent the four reanalysis from nearest to furthest from the measurement mast respectively

5.2.2 Seasons

Figure 5.2 presents the average percentage of increase of all the sites for each month. It shows a clear variation throughout the seasons. The months January, February, October, November and December stand out as months with higher overestimation, higher than 20 %, and the other months are within the 15 % - 20 % range, with May that stands out more than the others with 19 % and July being the month with the smallest overestimation of 15 %. However the standard deviation of each month is still high, always being between 10.25 % and 13.94 %. No month stands out as being more precise than another.

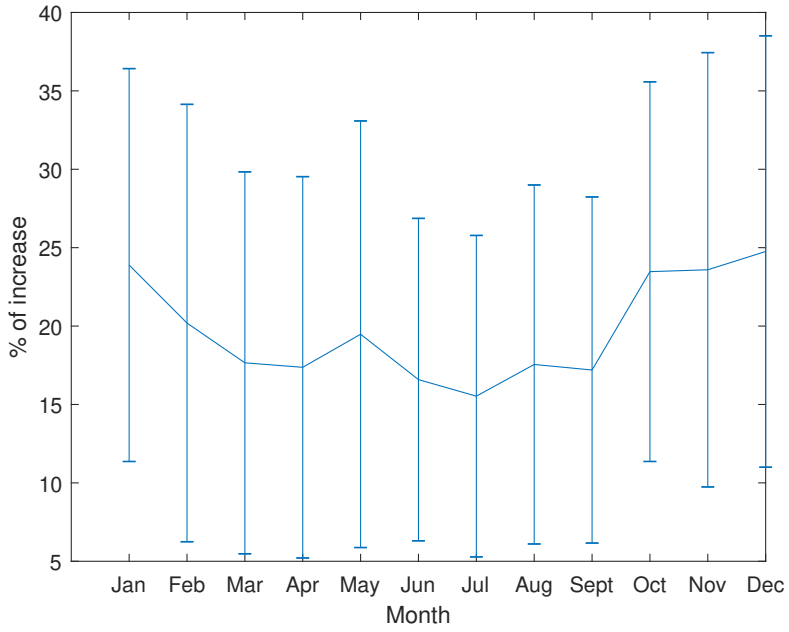


Figure 5.2: Monthly variation EMD-WRF - Real measurements. The bars represent the standard deviation for each months

5.2.3 Other parameters

The table 5.4 presents the value of correlation R_2 between the percentage of increase and the other parameters described in the method, ie. correlation coefficient R, altitude and height.

	Coefficient correlation R	**Altitude**	**Height**
R_2	-0.05	0.28	-0.08

Table 5.4: Relation between parameters

The correlation coefficient parameter R as well as the mast height does not show that there is a trend between those parameter and the overestimation, indeed they have a coefficient of correlation R_2 of -0.05 and -0.08 respectively which is low. The altitude has a better coefficient of correlation R_2 with the percentage of increase, 0.28. But that would show a little correlation between overestimation and altitude.

5.3 Discussion

The result of the reanalysis comparison showed that reanalysis dataset EMD-WRF has a moderate accuracy regarding wind speed. On the overall 23 sites, the mean difference from real measurements is 18.6%, the mean standard deviation is 12.52 % and the bias is of 1.02 m/s. Those results are within the gap obtained by the study done on the previous reanalysis generation EMD-ConWx (Haxsen 2017), which found also an average bias of 1.02 m/s. Moreover the positive bias is in accord with the bias observed in France with different reanalyses (Benedict Jourdier 2015) and it indicates that wind power estimations are going to overestimate the wind power productions if reanalyses are used as inputs. However, the range of results is too broad, ie. the standard deviation is too high, to propose a general offset or modification of the reanalysis wind speed. Indeed the risk to underestimate the results seems too important.

The results obtained are however worse estimations than the Coburn 2019 article, that had bias between 0.127 m/s and 0.415 m/s. This could be explained by the low height of the measurement masts in this article, ie. 10 m agl. The wind is probably weaker at this height than for the measurements presented in this thesis. Therefore the bias will be smaller. Also, the study took place on another continent, as seen in Benedict Jourdier 2015 reanalysis bias are different on different geographical zones and Coburn 2019 work was in the Midwest, in the United States.

Some of the imprecision could come from the vertical interpolation, the power law was used and it is known to bear some imprecision as seen in Elkinton, Rogers, and McGowan 2006. But the interpolations done to get on the mast's height are relatively on a small scale. Indeed, the wind height is displaced by less than 25 m vertically, as the reanalysis dataset EMD-WRF proposes 50 m height steps between 10 m and 300 m. It would have been more in other reanalysis datasets that have only a few height values, eg. MERRA-2 has a maximum height of 50 m agl.

Furthermore, another source of overestimation could be from the horizontal extrapolation. One method was used to extrapolate in distance, the WAsP model and the downscaling method implemented in WindPRO. Another wind profile and wind flow modelling would probably have change the result. But with the tools available at the disposition and the time constraint it did not seem reasonable to investigate correctly this source of error. Nonetheless, the methods used are applied in the wind industry and usually bear good results when real wind data are used.

The month division of the data showed a variation of the overestimation values throughout the year. Winter has more deviation from the real wind speed while summer has relatively less. The high deviation in winter is to be put in relation with the Coburn 2019 study. There, winter also had higher bias but it was spring which bear the lowest bias for ERA-Interim and it was summer for MEARRA2. As said in section 3.2.4, the wind speed in winter is the highest of the year and the opposite happen in summer. Reanalyses are only hour-base datasets here and measurements are on a 10-minute step. So strong wind that occur on short time frame might be harder to model precisely.

The coefficient of correlation R between the datasets is 0.79 with a standard deviation of 0.05 for 23 sites. It is in the same range as the ERA-5 dataset, 0.78 (that is an input of the EMD-WRF reanalysis), and Haxsen 2017 study of EMD-ConWx, 0.813. The standard deviation is slightly lower, but it could be explained by the lesser number of analyzed sites. It is a sign of improvement of reanalysis generation after generation.

No trend was found between the coefficient of correlation R and the accuracy of the reanalysis. Meaning that a reanalysis data that possess a good correlation with the real wind data does not necessarily have a wind speed close to the wind speed of the real data. The same is true for the altitude, a mast at an high altitude does not have necessarily better or worse accuracy with the reanalysis. Finally, the height of the anemometer is also not a parameter that is correlated to the accuracy of the reanalysis.

5.4 Summary

It appeared that the reanalysis dataset EMD-WRF is prone to overestimate wind speed. For all the sites the bias was positive. However, the overestimation can be either high (more than 40 % for the site 9) or low (less than 2.5 % for the site 8). Due to the wide range of results, it was decided to look for parameters to reduce that range. The next chapter focuses on separating the sites depending on the terrain complexity to see if the imprecision is connected to it.

CHAPTER 6

Categorized data

In this chapter, the data are compared based on the terrain specificity. The terrain can be described as complex for either having a lot of forest or having a lot of hills. The method of categorizing the terrain is presented.

6.1 Method

In order to determine relations or correlations between data, other categories are set. The first one is depending on the geographic proximity of the sites. The second one is depending on the terrain complexity, it is sub-divided in three parts:

- Depending on the presence on forests in 20 % of the studied area
- If one fourth of the site has 20 % of forest
- If the terrain has hedges in the studied area

6.1.1 Geographic proximity

The first category aims to see if the accuracy of the reanalysis is geographically localized, fortunately some sites are near to each other, within a 13 km radius. 4 pairs of sites are found and a small cluster of 4 sites within a 11 km radius are also found. Their percentage of increase is calculated.

6.1.2 Terrain specificity - Orography

The second category concerns the terrain. As seen in section 3 the orography and the roughness of a terrain are important inputs in a wind estimation project. Due to the amount of data, it was arbitrarily decided to separate the orography scale in two categories. To assess the orography level, the RIX of the terrain is calculated in the WindPRO interface. It is assessed on a 1.5 km around all the four reanalysis surrounding the measurement mast, with a critical angle of 10 %. It was chosen to define a terrain as complex if at least one of the four RIX value showed that 10 % of the terrain had at least 10 % slope.

6.1.3 Terrain specificity - Roughness

As no way of categorizing the roughness of a terrain was found in the literature, the roughness is chosen to be compared in three different ways. The percentage of forest in the area surrounding the mast is calculated. To do so, the software *ImageJ* is used, it allows to calculate surfaces with satellite view of sites. A screenshot is presented on figure 6.1. Each of the four squares surrounding the mast define what is called later an ”area”. For each of those area, the percentage of trees is calculated by dividing the surface of the trees by the total surface, as shown in equation 6.1.

$$\text{Percentage of tree} = \frac{\text{Surface of trees in the area}}{\text{Total surface of the area}} \tag{6.1}$$

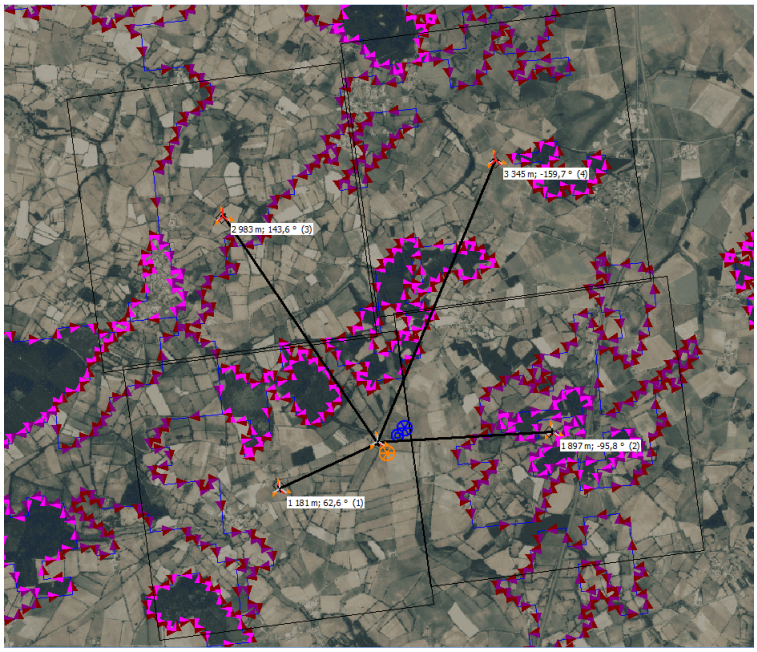


Figure 6.1: Screenshot of the image used in *ImageJ*. The violet lines represent the Corine land cover data (roughness). The black squares represent the area of which the density of trees are calculated, they are centered on a reanalysis location

It was decided to test three subdivisions to define the roughness complexity. Those are presented below.

First it was decided to define the roughness as complex if on average the four areas surrounding the mast have a percentage of trees above 20 %, this is called the **first alternative - Mean**. The terrain is described as presented in table 6.1.

	Complex orography	**Simple Orography**
Complex roughness	Forest area >20 % in the site RIX >10 % **Case 1**	Forest area >20 % in the site RIX <10 % **Case 3**
Simple roughness	Forest area <20 % in the site RIX >10 % **Case 2**	Forest area <20 % in the site RIX <10 % **Case 4**

Table 6.1: Terrain definition for the first alternative - Mean

Then it was deemed interesting to also separate the sites if one of the four areas has more than 20 % of its area composed of forest, it is called the **second alternative - Single**.

	Complex orography	**Simple Orography**
Complex roughness	1 of the 4 areas surrounding the mast has more than 20 % of forest RIX >10 % **Case 1**	1 of the 4 areas surrounding the mast has more than 20 % of forest RIX <10 % **Case 3**
Simple roughness	No area surrounding the mast has more than 20 % of forest RIX >10 % **Case 2**	No area surrounding the mast has more than 20 % of forest RIX <10 % **Case 4**

Table 6.2: Terrain definition for the second alternative - Single

A final separation for roughness is depending of the presence of hedges on the terrain. Hedges are rows of trees separated, an example can be seen on figure 6.2. The separation is done from satellite views if the presence of hedges is important or not. It is called the **third alternative - Hedges**.

	Complex orography	**Simple orography**
Complex roughness	Presence of hedges RIX >10 % **Case 1**	Presence of hedges RIX <10 % **Case 3**
Simple roughness	Absence of hedges RIX >10 % **Case 2**	Absence of hedges RIX <10 % **Case 4**

Table 6.3: Terrain definition for the third alternative - Hedges

In addition, for those three alternatives the same three parameters investigated in section 5.2.3 are looked into, ie. the coefficient correlation R_2 between the overestimation and the altitude of the mast, the height of the mast and the correlation coefficient R.

Figure 6.2: Example of hedges of trees on a site

6.2 Results

This section displays the findings made by separating sites depending on certain criteria. Its goal is to see if the group accuracy is less scattered than the overall accuracy found in section 5.2.1.

6.2.1 Geographic proximity

Table 6.4 represents the 4 pairs of sites and the distance between each pairs, as well as the cluster of 4 sites within a 11 km radius, the site's number is also indicated . For the pairs 1, 2 and 4 there is more than a difference of 10 % between the two sites. Only the pair 3 has close percentage of overestimation as well as two sites of the cluster (14 and 21). Therefore, the results do not indicate that the distance between sites is correlated with the accuracy of the reanalysis data.

	Distance	Percentage of increase
Pair 1 (6-7)	5 km	5.9 % - 19.37 %
Pair 2 (13-23)	13 km	2.7 % - 13.38 %
Pair 3(17-19)	8 km	19.9 % - 16.53 %
Pair 4 (3-22)	3 km	17.9 % - 34.25 %
Cluster (11-14-16-21)	11km	27 % - 32 % - 14 % - 35 %

Table 6.4: Impact of site proximity regarding overestimation

6.2.2 Terrain category

This section presents the results for the alternative of terrain categorization. The first alternative depends on the presence of 20 % of forest around the mast (on average), the second alternative depends on having one of four areas around the mast having more than 20 % of forest. The third and last alternative depends on the presence of hedges on the studied site.

6.2.2.1 First alternative - Mean

Table 6.5 presents which and how many sites are in each case of terrain complexity. The third and fourth columns are the mean of percentage of increase (overestimation) for each cases and the mean standard deviation (STD). The last three columns show the coefficient correlation R_2 between the percentage of increase and the three parameters defined in the section 5.1.2, correlation R, altitude of the site and height of the measurement mast.

				Correlation R_2		
	Site number	**Mean (%)**	**STD (%)**	**Correlation R**	**Altitude**	**Height**
Case 1	5;10;16;18	14.32	10.47	0.09	-0.47	0.26
Case 2	2;3;4;11 17;21;22	24.73	8.7	-0.31	0.06	-0.44
Case 3	12;15;20	18.2	6.95	0.51	0.96	-0.78
Case 4	1;6;7;8;9;12 14;19;23	12.62	12.39	-0.31	0.70	0.05

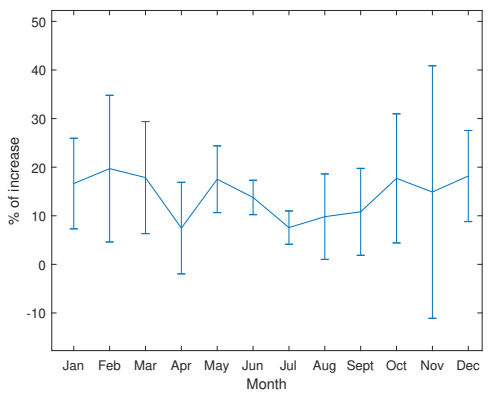
Table 6.5: Alternative 1 - Results

Figure 6.3 represents the percentage of increase for the four Cases of terrain defined in table 6.1. For each month the mean of each cases is displayed as well as the standard deviation represented by the vertical lines. Case 3 has months without standard deviation, eg. January or February, because the sites used do not have recorded data

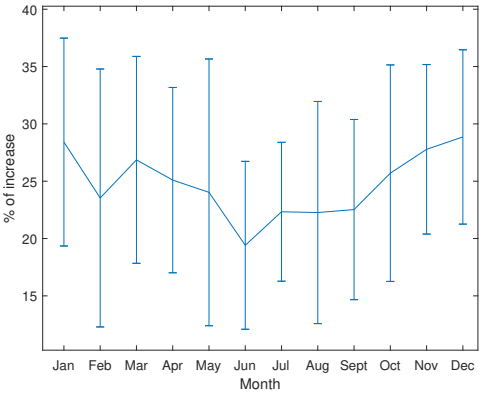
on those month. It also explain the gaps of value between February and March for the overestimation in this Case. Overall, the standard deviation is equal or inferior to the global value obtained in section 5.2.1, ie. 12.52 %.

Case 1 and 4 have the least overestimation, 14.32 % and 12.62 % respectively, but Case 1 only have 4 sites while Case 4 has 9 sites. Case 1 presents small standard deviations for each month except for November. Case 2 has the highest mean percentage of increase of 24.73 %.

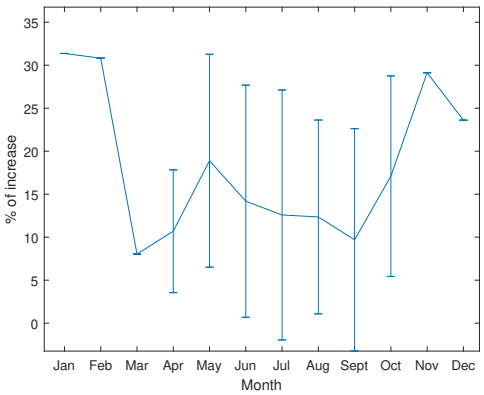
Regarding the correlation R_2 between the percentage of increase and the three studied parameters, no clear trend is seen, the values are sometimes positive and sometimes negative.



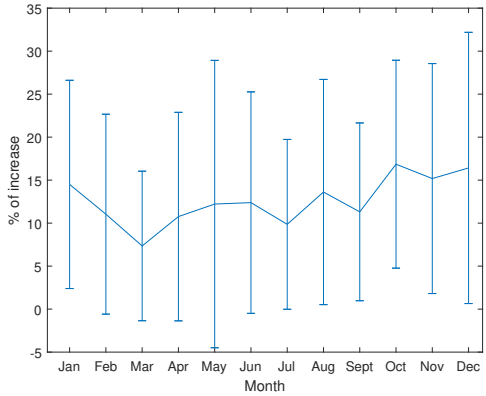
(a) Case 1: Complex (orography) complex (roughness)



(b) Case 2: Complex (orography) Simple (roughness)



(c) Case 3: Simple (orography) complex (roughness)



(d) Case 4: Simple (orography) Simple (roughness)

Figure 6.3: Percentage of increase on monthly variation for the 4 categories

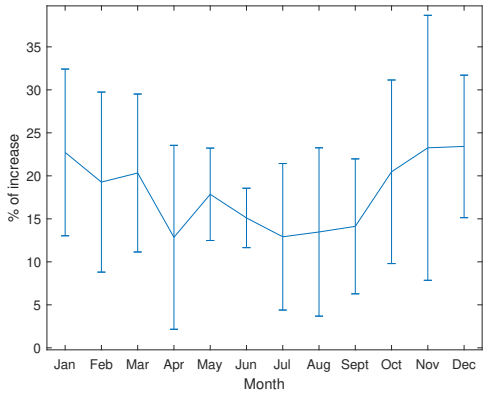
6.2.2.2 Second alternative - Single

The same values are calculated for the second alternative. The standard deviation is also lesser or equal to the global value. Case 2 still has the highest overestimation (26.22 %) and Case 4 the lowest (10.58 %).

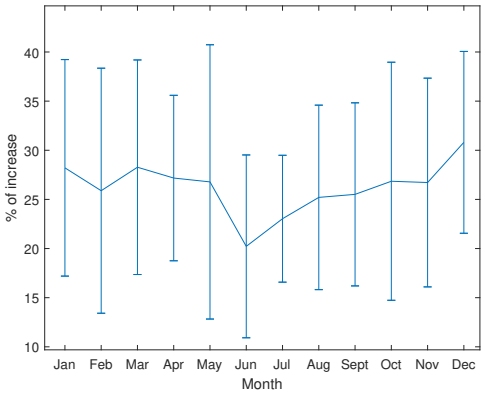
The three parameters studied do not show any particular trend compared to the previous alternative.

				Correlation R_2		
	Site number	**Mean (%)**	**STD (%)**	**Correlation R**	**Altitude**	**Height**
Case 1	3;4;5;10;11;16;18	17.98	9.12	0.27	-0.62	0.23
Case 2	2;17;21;22	26.22	10.27	-0.47	0.25	-0.34
Case 3	1;9;12;15;20;23	16.44	12.94	0.22	0.76	-0.28
Case 4	6;7;8;13;14;19	10.58	10.89	-0.52	0.74	-0.14

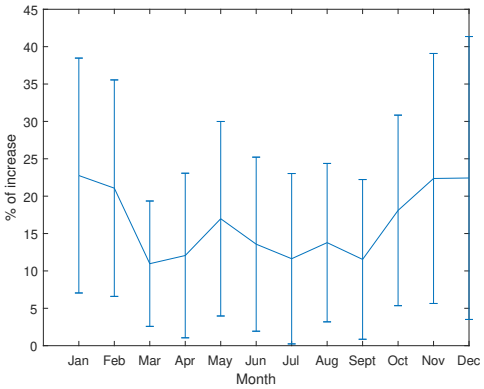
Table 6.6: Alternative 2 - Results



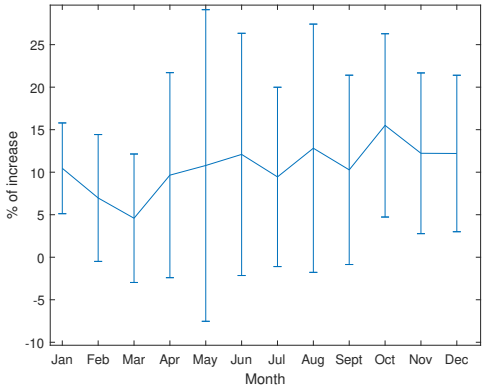
(a) Case 1: Complex (orography) complex (roughness)



(b) Case 2: Complex (orography) Simple (roughness)



(c) Case 3: Simple (orography) complex (roughness)



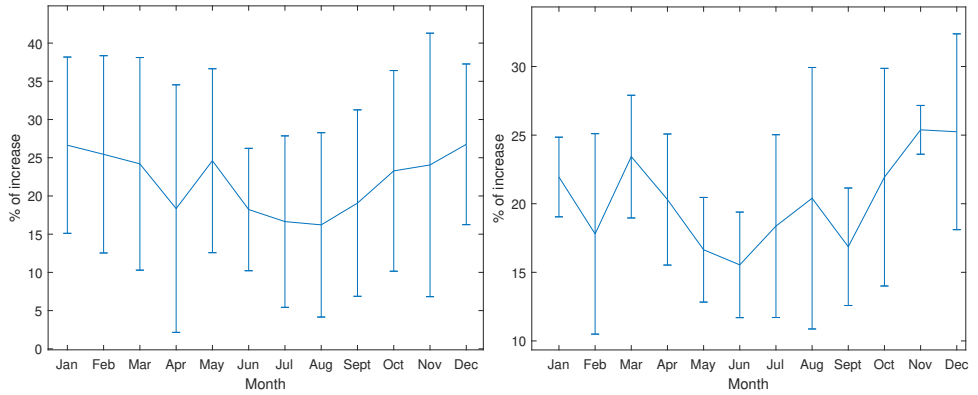
(d) Case 4: Simple (orography) Simple (roughness)

Figure 6.4: Percentage of increase on monthly variation for the 4 categories

6.2.2.3 Third alternative - Hedges

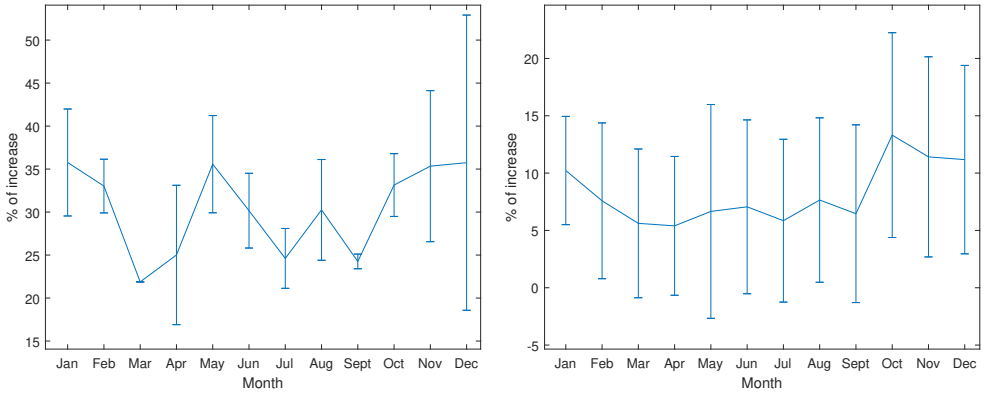
For the last alternative, the same calculations are also made. A noticeable improvement is present for the standard deviation values. They are lesser or equal to the global result and have good results for the cases 2, 3 and 4 with numbers between 7.4 % and 5.37 %. The least overestimation is for the case 4, with a minimum of 8.20 % and the strongest is for the sites with hedges and simple orography (Case 3) with a maximum of 30.40 %.

Identically as the first two alternatives, the three parameters displayed in the last three columns do not show a trend.



(a) Case 1: Complex (orography) complex (roughness)

(b) Case 2: Complex (orography) Simple (roughness)



(c) Case 3: Simple (orography) complex (roughness)

(d) Case 4: Simple (orography) Simple (roughness)

Figure 6.5: Percentage of increase on monthly variation for the 4 categories

				Correlation R_2		
	Site number	**Mean (%)**	**STD (%)**	**Correlation R**	**Altitude**	**Height**
Case 1	2;4;5;10;16;21;22	21.96	12.58	0.40	-0.54	0.06
Case 2	3;11;17;18	20.32	5.37	-0.49	-0.008	0.04
Case 3	9;12;14	30.4	5.6	-0.11	-0.96	0.01
Case 4	1;6;7;8;13 15;19;20;23	8.2	7.4	0.44	0.04	-0.04

Table 6.7: Alternative 3 - Results

6.3 Discussion

6.3.1 Geographic proximity

Unfortunately, geographical proximity did not present a correlation with overestimation. Indeed, close sites did not present the same range of overestimation. Therefore reanalysis bias does not seem to be closely related between two geographical locations.

6.3.2 Terrain specificity

The three different terrain categorization allowed to see a reduction of the standard deviation going from 12.52 % to values below 10 %. This means that the uncertainties regarding the overestimation of the reanalysis decreased. The simplest case, ie. simple orography and simple roughness or hedges density, displays a noticeably lower mean of percentage of increase than the other cases. For example, the site number 8, which is the one with the lowest overestimation measured in the previous chapter, is part of the Case 4 in the three alternatives and has RIX value of 0.8 % and less than 7 % of forests in its surroundings. Consequently, sites with at least one complex terrain parameter (roughness or orography) are more likely to be more overestimated.

Regarding the third alternative, the overestimation seems more important in the presence of hedges on sites. No explanation was found in the literature. The hedges possibly slow down the wind but because their surfaces are low they are not well modelled on a mesoscale reanalysis.

No clear correlation was found between the overestimation and correlation coefficient R. The correlation coefficient R_2 between those dataset is negative or positive, meaning that sometimes an increase of the correlation coefficient means an increase of the overestimation (if they are positive) and sometimes it leads to the opposite (if they are negative). It can be noted that R_2 for the Case 4 (simple terrain for orography and roughness) for the first two alternatives are negative, which would mean that for simple terrain a better coefficient of correlation R implies a better accuracy of the reanalysis data. It is possible that simple terrains have clearer trend as they would have less differences with the mesoscale terrain data.
For the parameter of the altitude of the site, the coefficient of correlation R_2 is also positive and negative. So no clear trend can be deducted from it. But, in the simple terrains, ie. case 4, the correlation coefficient R_2 is around 0.7 meaning that when the altitude of the site increase the overestimation also increases, which would be coherent with Sharp et al. 2015.
For the parameter of the height of the anemometer, the coefficient of correlation R_2 is also positive and negative in different cases. No clear trend can be deducted from it. For the simple terrains, the correlation coefficient R_2 is close to zero, which would imply that the height of the anemometer and the overestimation are not correlated

which would be different from the findings made by Haxsen 2017.

Overall, the choice of the terrain categorizing had a part of arbitrary decision as not a lot of literature was found on the topic. More precise results could have been made with different criteria for simple or complex terrain.

CHAPTER 7

Wind power calculation

This chapter showcases the accuracy of the EMD-WRF in a wind power calculation for five sites that provided their real production values. The reanalysis is used first as a raw data and then is corrected by the results found in the chapter 6.

7.1 Method

The comparison between the real-life production and the calculated one is done by first selecting a way to correct the reanalysis data in order to have it closer to the real-life wind speed. Then, the losses in the calculated production are presented to match more accurately the real-life production.

7.1.1 Sites presentation

Five sites have provided their annual production. The sites have between 3 and 9 wind turbines installed. The models and the emplacements are anonymized, but the sites are situated in the same area of France as the measurement masts studied.

Each site is categorized into a case using the second alternative method's and are presented in table 7.1. This alternative was chosen because it was the only one that has at least four sites for each categories, making it more representative. The overestimations per month are presented in tables 7.2, 7.3, 7.4 and 7.5. They correspond to the percentage of increase per month for the four cases obtained in the second alternative - Single, in figure 6.4, which formula is presented in equation 5.1.

Sites	**A**	**B**	**C**	**D**	**E**
Category	Case 3	Case 3	Case 2	Case 1	Case 4

Table 7.1: Wind farm site categories

Month	Jan	Feb	Mar	Apr	May	Jun
Overestimation	22%	19%	20%	13%	18%	15%
Month	**Jul**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
Overestimation	12%	13%	14%	20%	23%	23%

Table 7.2: Overestimation for Case 1

Month	Jan	Feb	Mar	Apr	May	Jun
Overestimation	28%	25%	28%	27%	26%	20%
Month	**Jul**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
Overestimation	23%	25%	25%	27%	27%	31%

Table 7.3: Overestimation for Case 2

Month	Jan	Feb	Mar	Apr	May	Jun
Overestimation	23%	21%	11%	12%	17%	13%
Month	**Jul**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
Overestimation	12%	14%	11%	18%	22%	22%

Table 7.4: Overestimation for Case 3

Month	Jan	Feb	Mar	Apr	May	Jun
Overestimation	10%	7%	4%	9%	11%	12%
Month	**Jul**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
Overestimation	9%	13%	10%	15%	12%	12%

Table 7.5: Overestimation for Case 4

On each of the five sites available, reanalysis data is selected and its wind speed is corrected according to the terrain category. For example, for the case 1 in January, the overestimation is of 22 % so:

$$\text{Percentage of increase} = 22\% = \frac{V_{EMD-WRF} - V_{corrected}}{V_{corrected}} \tag{7.1}$$

$$V_{corrected} = \frac{V_{EMD-WRF}}{1 + 22\%} \tag{7.2}$$

Then, with WindPRO module PARK, the electricity production of the farm is estimated using the raw data of the reanalysis and the corrected one.

7.1.2 Losses

The data given by the developers contain already some of the losses, so they need to be estimated for each project. The wake effect is calculated directly by PARK module in WindPRO. For the other losses, it was chosen to take the values described in section 2.3:

- Unavailability losses: arbitrary 2 %;
- Grid unavailability: arbitrary 0.75 %;
- Electricity losses: arbitrary 0.5 %;
- Park consumption: arbitrary 1.5 %;

No curtailment losses were accounted as the developers did not disclose the information about their wind farms.

7.2 Result

Table 7.6 presents in its third column the percentage of increase between the real production and the one calculated with the reanalysis wind data as input. As expected, the reanalysis alone overestimate the production a lot. The values range between 21.1 % and 50.2 %. The fourth column presents the percentage of increase with the corrected reanalysis data. The results are improving. The sites A, D and E have a production estimation very similar to the real one, within 4 %. The site B went from 21.1 % to -21.1 % and the site C went from 48.8 % to -19.9 %.

Site	Category	Difference from reality (with *raw* reanalysis data)	Difference from reality (with corrected reanalysis data)
A	Case 3	48.8%	-0.9%
B	Case 3	21.1%	-21.1%
C	Case 2	48.8%	-19.9%
D	Case 1	50.2%	3.9 %
E	Case 4	28.3%	3.9 %

Table 7.6: Power production difference

7.3 Uncertainties

For those five projects, it is possible to estimate what would have been the uncertainties defined in section 2.4. Those are :

- 20 % if the wind data is from a reanalysis and around 3 % if real wind data is used
- No interannual variability. The period of the wind data chosen correspond exactly to the period of the production.
- 1 % for each 10 m between the reanalysis data height and the wind turbine height
- 1 % for each kilometers between the reanalysis location and the wind turbine the furthest

Then the formula 2.8, which give the total uncertainty, is applied and the results are displayed in table 7.7. The total uncertainty is around 21 %.

Uncertainties	Site A	Site B	Site C	Site D	Site E
Wind data	20%	20%	20%	20%	20%
Vertical extrapolation	0.1%	1.0%	0.5%	0.5%	0.8%
Horizontal extrapolation	1.3%	1.7%	3.7%	1.4%	1.0%
Total	**20.04%**	**20.10%**	**21.35%**	**20.06%**	**20.04%**

Table 7.7: Uncertainties for the 5 studied sites

In comparison, if real wind data were used the uncertainties would be lower, that case is presented in table 7.8.

Uncertainties	Site A	Site B	Site C	Site D	Site E
Wind data	3.0%	3.0%	3.0%	3.0%	3.0%
Vertical extrapolation	0.1%	1.0%	0.5%	0.5%	0.8%
Horizontal extrapolation	1.3%	1.7%	3.7%	1.4%	1.0%
Total	**3.3%**	**3.6%**	**4.8%**	**3.3%**	**3.3%**

Table 7.8: Uncertainties if real wind data were used

The WindPRO software uses a normal distribution to assess the probability of a wind estimation to be within a certain range. A general rule is to say that there is a probability of 95 % to have the real production in the following range (Kazmier 2009):

$$P_{estimation} \pm 2 \cdot \text{total uncertainty} \tag{7.3}$$

The results obtained in the third column of table 7.6 show that, without the offset, the estimations of sites B and E are within the estimated range of the uncertainty advised for the reanalysis data, which is approximately 40 % for a probability of 95 %. But the sites A, C and D are above this limit and are too imprecise.

With the offset applied, the five estimations are well within this range, even for the worse estimations, ie. Site B and C. It is noticeable that the sites A, D and E are even within the range of uncertainties in the case where real wind data is used, approximately 6 % - 10 % for a probability of 95 %.

7.4 Discussion

Reanalyses data without any modification were not accurate inputs to assess wind production. It led to an overestimation between 21.1 % and 50.2 % for the sites studied. By taking into account the uncertainties of the method used, three sites out of five are deemed very imprecise with a difference of around 50 % with the real production. This range of overestimation is expected as it was seen in section 2.2.1 the power production is dependent to the wind speed to the power of three. For example, an overestimation of 18 % of the wind speed by the reanalysis data is going to add an overestimation of 64 % of the wind production.

With the modification proposed, the result were much closer to the reality. Sites A, D and E were within 4 % of the real production value. Site C went from 48.8 % overestimation to 19.9 % underestimation. In this case, the accuracy of the reanalysis data improved but the correction led to have an underestimation of the power production which is important. It was the same for Site B which went from an overestimation of 21.1 % to an underestimation of 21 %.

To draw a firm conclusion on the modification used on the reanalysis, more sites should be tested. So far the Case 1 and Case 4 have avoided to underestimate the wind production of the site. Finally, a more conservative position could be to only apply the offset of the Case 4 as it is the weaker offset proposed. It would avoid underestimating wind potential of a site.

In the context of a wind power estimation that is supposed to assess the viability of a site to be profitable, the estimations made here could not be judged acceptable by a wind power company yet as the accuracy of reanalysis data is varying. Based on real wind power estimations, the uncertainties on the results should not be more than 10 %. However, for wind power estimations happening during the prospecting phase of the wind development timeline, those estimations could possibly be judged acceptable as their goal is to give a rough and fast estimation of the potential of the prospected site.

CHAPTER 8
Conclusion and future work

Wind estimations are a central part of wind development process. Depending on their results, the future of the wind farms can change. They require costly measurement masts to be set up even if some data, eg. wind maps or reanalysis, are available to have a rough estimate of the site's potential. Nowadays softwares offer different methods to extrapolate a wind measurement, for example with the logarithmic law or the power law. However the precise methods used by the developers are kept secret. But in general, it would save time and money to be able to assess accurately the energy potential on one site only according to reanalysis data.

Unfortunately, even if the reanalysis dataset, global and mesoscale, are getting more and more accurate and with better correlation coefficient with the real measurements, they still are not able to estimate the potential of a site with a good accuracy. The EMD-WRF reanalysis seems to be one of the best reanalysis analyzed in this thesis, as it has equal or better accuracy compared to them according to the literature found.

The thesis aimed at finding parameters that cause more important overestimation in wind power estimation. Terrain complexity showed good promise to have a role in the reanalysis overall accuracy. On the other hand, correlation coefficient, altitude, height of the mast did not show strong signs of being linked to the reanalysis accuracy except for simple terrains. Also the geographical proximity of sites did not seem to be link to the reanalysis accuracy of those sites.

The proposed solution to reduce the wind speed according to the terrain complexity showed promising results in the form of a better accuracy for wind power estimation.

Here are some areas which may be of interest for future research around the subject of the thesis:

- A possible link between different input parameters of a wind power estimation would be interesting. A general formula depending on exterior parameters allowing better estimation was considered but was not done due to the lack of time available and the few parameters showing links with overestimation.

- To go deeper into the root of the inaccuracy, one could look into the input parameters used in the modelling of the reanalysis as well of the model used and

their limits with a given terrain for example. Unfortunately the models used by reanalysis providers are often kept secret.

Bibliography

REN21 (2019). *Renewables 2019 Global Status Report*, page 336. ISBN: ISBN 978-3-9818911-7-1. URL: https://wedocs.unep.org/bitstream/handle/20.500.11822/28496/REN2019.pdf?sequence=1&isAllowed=y%0Ahttp://www.ren21.net/cities/wp-content/uploads/2019/05/REC-GSR-Low-Res.pdf.

ADEME (2017). *ÉTUDE SUR LA FILIÈRE ÉOLIENNE FRANÇAISE*. Technical report. ADEME, page 324. URL: https://www.ademe.fr/sites/default/files/assets/documents/filiere_eolienne_francaise_2017-rapport.pdf.

Wind in France: power installed (2018). URL: https://bilan-electrique-2018.rte-france.com/eolien/#.

Manwell, J. F., J. G. McGowan, and A. L. Rogers (2009). *Wind Energy Explained*. Second, page 705. DOI: 10.1002/9781119994367.

Carson, Rachel et al. (2002). *Silent Spring*. Volume 5384, page 378. ISBN: 0-618-24906-x.

Annual electricity generation in Germany (2019). URL: https://www.energy-charts.de/energy.htm?source=all-sources&period=annual&year=all.

Benedict Jourdier (2015). "Ressource éolienne en France métropolitaine : méthodes d'évaluation du potentiel , variabilité et tendances". PhD thesis, pages 1–229.

Poyri (2016). "Observatoire Des Coûts De L'Éolien Terrestre". In:

Troen, I. and E. Lundtang Petersen (1989). *European wind atlas*, page 656. ISBN: 8755014828. URL: http://orbit.dtu.dk/en/publications/european-wind-atlas(335e86f2-6d21-4191-8304-0b0a105089be).html.

DTU Wind Energy and World Bank Group (2018). *Global Wind Atlas*. DOI: 10.1093/mnras/sts501. URL: https://globalwindatlas.info/.

WindProspecting (2019). URL: https://www.windprospecting.com/.

Meteo France (2012). URL: https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=92&id_rubrique=37.

MEASNET (2016). *MEASNET PROCEDURE: Evaluation of Site-Specific Wind Conditions. Version 2*. Technical report.

WASP (2019). *Wake Effect Model*. URL: https://www.wasp.dk/wasp#details__wakeeffectmodel.

Enercon (2007). *Enercon wind turbines*. URL: http://www.tut.fi/eee/research/adine/materiaalit/Active%20network/System%20integration/Enercon/ENERCON_Technology+Service.pdf.

General Electric (2020). *Proven performance, reliability, and availability*. URL: https://www.ge.com/renewableenergy/wind-energy/onshore-wind/2mw-platform.

GoldWind (2020). *Advanced PMDD Technology*. URL: https://www.goldwindamericas.com/technology.

Nordex (2018). "Nordex group annual report". In:

SIEMENS GAMESA (2019). *Service Wind*. URL: https://www.siemensgamesa.com/en-int/products-and-services/service-wind.

Nielsen, John (2012). "Solutions and Services – Our second revenue stream". In:

Martin, Rebecca et al. (2016). "Sensitivity analysis of offshore wind farm operation and maintenance cost and availability". In: *Renewable Energy* 85.July, pages 1226–1236. ISSN: 18790682. DOI: 10.1016/j.renene.2015.07.078. URL: http://dx.doi.org/10.1016/j.renene.2015.07.078.

EMD International A/S (2013). *An Introduction to the MCP Facilities in WindPRO*. Technical report. EMD.

NTNU (2016). *Course : The Wind*. URL: https://slideplayer.com/slide/7788151/.

Danish Wind Industry Association (2003). *The Roughness Rose*. URL: http://xn--drmstrre-64ad.dk/wp-content/wind/miller/windpower%20web/en/tour/wres/rrose.htm.

Mortensen, N G, A ; Tindal, and L Landberg (2008). *General rights Field validation of the RIX performance indicator for flow in complex terrain FIELD VALIDATION OF THE ΔRIX PERFORMANCE INDICATOR FOR FLOW IN COMPLEX TERRAIN*. Technical report.

Elkinton, Melissa R., A. L. Rogers, and J. G. McGowan (2006). "An investigation of wind-shear models and experimental data trends for different terrains". In: *Wind Engineering* 30.4, pages 341–350. ISSN: 0309524X. DOI: 10.1260/030952406779295417.

Bayon-Barrachina, Arnau et al. (2014). "Using downscaled NCEP/NCAR reanalysis data for wind resource mapping". In: *International Journal of Energy & Environment* 5.3, pages 305–316. URL: http://www.ijee.ieefoundation.org/vol5/issue3/IJEE_03_v5n3.pdf.

Rathmann, Ole et al. (2006). *Turbine Wake Model for Wind Resource Software*. Technical report. URL: https://www.researchgate.net/publication/268300272.

EMD International A/S (2019). *WindPRO user manual*. Technical report, page 193.

Jothiprakasam, Venkatesh Duraisamy (2015). "Descente en échelle de la ressource en énergie éolienne de la mésoéchelle à l'échelle locale par imbrication et assimilation de données à l'aide d'un modèle de CFD". In:

Landberg, Lars et al. (2003). *Wind resource estimation - An overview*. DOI: 10.1002/we.94.

Coburn, Jacob J. (2019). "Assessing wind data from reanalyses for the upper Midwest". In: *Journal of Applied Meteorology and Climatology* 58.3, pages 429–446. ISSN: 15588432. DOI: 10.1175/JAMC-D-18-0164.1.

Bosilovich, Michael et al. (2015). "MERRA-2 : Initial Evaluation of the Climate". In: *NASA Technical Report Series on Global Modeling and Data Assimilation* 43.September, page 139. DOI: NASA/TM-2015-104606/Vol.43.

ECMWF (2019). *What is ERA Interim ?* URL: https://confluence.ecmwf.int/display/CKB/What+is+ERA-Interim.

Haxsen, Sören (2017). "Suitability of the EMD-ConWx europe mesoscale data for wind resource assessments". PhD thesis. Uppsala University, page 73.

Sharp, Ed et al. (2015). "Evaluating the accuracy of CFSR reanalysis hourly wind speed forecasts for the UK, using in situ measurements and geographical information". In: *Renewable Energy* 77, pages 527–538. ISSN: 18790682. DOI: 10.1016/j.renene.2014.12.025. URL: http://dx.doi.org/10.1016/j.renene.2014.12.025.

EMD (2019). *EMD-ConWx Meso Data Europe*. URL: http://help.emd.dk/mediawiki/index.php?title=EMD-ConWx_Meso_Data_Europe.

Kazmier, Leonard (2009). "Schaum's Outline of Business Statistics, Fourth Edition". In: URL: https://books.google.nl/books/about/Schaum_s_Outline_of_Business_Statistics.html?id=dkL-hJAYzSwC&pgis=1.