Diese Arbeit wurde vorgelegt am Lehr- und Forschungsgebiet
Theorie der hybriden Systeme

Unsicherheitsquantifizierung von Windparkmodellen

Uncertainty quantification of wind farm models

Bachelorarbeit
Informatik

September 2017

Vorgelegt von Presented by	Roni Cakar Langerfelder Str. 113a, 42389 Wuppertal Matrikelnummer: 317104 roni.cakar@rwth-aachen.de
Erstprüfer First examiner	Prof. Dr. Erika Ábrahám Lehr- und Forschungsgebiet Theorie der hybriden Systeme RWTH Aachen University
Zweitprüfer First examiner	Prof. Dr. Martin Frank Lehrstuhl für Mathematik (MathCCES) RWTH Aachen University
Koreferent Co-supervisor	Dr. rer. nat. Pascal Richter Lehrstuhl für Mathematik (MathCCES) RWTH Aachen University

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich diese Bachelorarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Stellen meiner Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, habe ich in jedem Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht. Dasselbe gilt sinngemäß für Tabellen und Abbildungen. Diese Arbeit hat in dieser oder einer ähnlichen Form noch nicht im Rahmen einer anderen Prüfung vorgelegen.

Aachen, im September 2017

Roni Cakar

Contents

Nomenclature V

List of Figures VI

List of Tables IX

1 Introduction 1

2 Background 2

3 Offshore Wind Farm Modeling 3
- 3.1 Wind Model . . . 3
  - 3.1.1 Wind Data Processing . . . 3
  - 3.1.2 Weibull Distribution . . . 4
- 3.2 Wake Model . . . 5
- 3.3 Power Generation Model . . . 8
- 3.4 Gross annual Energy Production (AEP) . . . 9
- 3.5 Cost Model . . . 10
  - 3.5.1 Net annual Energy Production (AEP) . . . 11
  - 3.5.2 Levelized Cost of Electricity (LCOE) . . . 11
  - 3.5.3 Net Present Value (NPV) . . . 11
  - 3.5.4 Internal Rate of Return . . . 12

4 Offshore Wind Farm Modeling with Uncertainties 13
- 4.1 Uncertain Input Parameters . . . 13
- 4.2 Offshore Model with Uncertainties . . . 14
  - 4.2.1 Perturbed Wind data . . . 14
  - 4.2.2 Uncertain Wind Model . . . 16
  - 4.2.3 Uncertain Wake Model . . . 16
  - 4.2.4 Uncertain Power Generation Model . . . 17
  - 4.2.5 Uncertain gross Annual Energy Production (AEP) . . . 18
  - 4.2.6 Uncertain net Annual Energy Production (AEP) . . . 18
  - 4.2.7 Uncertain Levelized Cost of Electricity . . . 19
  - 4.2.8 Uncertain Net Present Value . . . 19
  - 4.2.9 Uncertain Internal Rate of Return . . . 19

5 Uncertainty Quantification (UQ) 20
- 5.1 Fundamentals of Probability and Statistics . . . 20
  - 5.1.1 Probability Basics . . . 20
  - 5.1.2 Statistics for Uncertainty Quantification . . . 22
  - 5.1.3 Boxplot . . . 23
- 5.2 Input Parameters . . . 24
- 5.3 Inverse Transform Method (ITM) . . . 24

5.4 Sampling Methods . . . 25
5.4.1 Monte Carlo Sampling . . . 25
5.4.2 Quasi-Monte Carlo Sampling . . . 26
5.4.3 Latin Hypercube Sampling . . . 28
5.5 Random Variable Generation . . . 29

6 Implementation **33**
6.1 Program Structure . . . 33
6.2 Accuracy of gross annual Energy Production (AEP) calculation . . . 34
6.3 Precision of Inverse Transform Method . . . 34

7 Results **38**
7.1 Horns Rev 1 . . . 38
7.1.1 Net Annual Energy Production (net AEP) . . . 40
7.1.2 Levelized Cost of Electricity . . . 42
7.1.3 Net Present Value . . . 44
7.1.4 Internal Rate of Return . . . 46
7.1.5 Summary of the Results . . . 47

8 Conclusion **49**

References **50**

Nomenclature

α_i	wind direction
β_κ	shadowing factor
δ_u	velocity deficit
κ	wake grow factor
λ	scale parameter weibull distribution
C_{capital}	total capital costs [€]
$C_{\text{o\&m}}$	annual operation & maintenance costs [%]
C_{PRV}	present residual value of the wind farm [€]
$C_t(u)$	thrust function
D	rotor diameter [m]
d	perturbation [%]
k	shape parameter weibull distribution
K_{energy}	price of energy on the market [€/kWh]
l	lifetime [years]
$P(u)$	power function in MW
p_{loss}	plant performance loss [%]
r_{rate}	discount rate [%]
u	wind speed [m/s]
u_{cutin}	cut-in speed [m/s]
u_{cutout}	cut-out speed [m/s]
z	hub height [m]
z_0	surface roughness [m]

List of Figures

1 Structure of the offshore model. Gray rectangles representing the four sub-models with their specific inputs (green ellipses) and their outputs (red ellipses). 3
2 Combined wind direction distribution at the FINO3 research platform over the years of July 2010 to June 2017, clustered into 12, 32 and 360 direction sectors. Source: Heiming[7] 4
3 Approximated wind speed distribution as a weibull distribution using the maximum likelihood estimation. The measured data is from the FINO3 research platform over the years of July 2010 to June 2017 for the wind direction sector $\alpha_i \in [225°, 255°)$. The measured wind speed distribution is illustrated with the histogram. 5
4 Turbine measurements: D represents the rotor diameter and z is standing for the hub height of the mast. Source: Heiming [7] 6
5 Visualization of the wake as described in PARK. The unperturbed free stream u_0 hits the turbine and its blades and generates the wake and the resulting wind speed u_w. The wake grows linear with the factor 2κ. D represents the rotor diameter of the turbine, D_w is the width of the generated wake and x stands for the distance from the turbine. Source: Heiming [7] 7
6 Intersection of wake effect on other turbine. Source: Heiming [7] 8
7 Thrust coefficent C_t and power production of the turbine Vestas V80 with cut-in speed of 4 m/s and cut-out speed of 25 m/s (dashed vertical lines). 9
8 Normal distribution $\mathcal{N}(0, 1)$ with coverage of generated values. 14
9 Fitted weibull distribution with maximum likelihood estimation (MLE) for wind speed distribution over the years of July 2010 to June 2017 for one wind direction sector $\alpha_i \in [225° - 255°)$ measured at the FINO3 research platform. Red and green plots representing the fitted weibull distribution with MLE after perturbing the wind speed data with a factor of ±6%, ±12% and ±18%. The random variable $\xi_{wind}(u)$ represents weibull distribution between the highest red plot (MLE fit -18%) and the lowest green plot (MLE fit +18%) with a probability of 99.73%. . . 17
10 Perturbed C_t and power curve of of the turbine Vestas V80 with cut-in speed of 4 m/s and cut-out speed of 25 m/s (dashed vertical lines). The red tube illustrates the C_t and power curve with a perturbation of ±10%. 18
11 Cumulative distribution function of normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$. 21
12 Probability density function of the normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$, also written as $\mathcal{N}(0, 1)$. 22
13 Boxplot representation with outliers, upper and lower whisker, upper and lower quartile and median. 24
14 One sample step of uncertainty quantification with mathematical model. 25

15 2D - C++ random number generator (Mersenne Twister Engine) with 1000 sample points. 26
16 2D - Sobol sequence with 1000 sample points. 27
17 2D - Latin Hypercube sampling with 1000 sample points. 28
18 Inverse Transform Method with C++ random number generator. In Figure (a) relation between probability density function of Uniform distribution and frequencies of C++ random number generator numbers. In Figure (b) result after transforming with $\texttt{erf}^{-1}(x)$ compared to probability density function of normal distribution with 4096 samples. . . . 31
19 Inverse Transform Method with Sobol Sequence. In Figure (a) relation between probability density function of Uniform distribution and frequencies of Sobol Sequence numbers. In Figure (b) result after transforming with $\texttt{erf}^{-1}(x)$ compared to probability density function of normal distribution with 4096 samples. 31
20 Inverse Transform Method with Latin Hypercube Sampling. In Figure (a) relation between probability density function of Uniform distribution and frequencies of Latin Hypercube Sampling numbers. In Figure (b) result after transforming with $\texttt{erf}^{-1}(x)$ compared to probability density function of normal distribution with 4096 samples. 32
21 UML class diagram 33
22 Accuracy of annual energy production calculation with the integration of two different weibull distributions (determined from the wind data of the FINO3 research platform) with one turbine and one direction with the trapezoidal quadrature rule with increasing number of quadrature points. The reference value of the AEP is the integration with $2^{17} = 131072$ quadrature points. 35
23 Accuracy of gross annual energy production calculation with increasing number of wind sectors. Reference value is annual energy production calculation with 4096 wind sectors and 64 quadrature points of the wind speed distribution. 36
24 Precision of inverse transform method according to normal distribution with pseudo-random and quasi-random numbers, measured with root-mean-squared error in dependency of sample size. The standard Monte Carlo method has a convergence rate of $\mathcal{O}(\frac{1}{\sqrt{N}})$ whereas the quasi-Monte Carlo method and the Latin Hypercube method have a rate of $\mathcal{O}(\frac{1}{N})$. . 37
25 Turbine positions and wake effect of wind farm Horns Rev 1. 39
26 Test with six uncertainty parameters wind speed, wake effect, C_t curve, surface roughness, power curve and power loss at the same time (all uncertainties simultaneously) compared to single perturbation of the uncertainty parameters computed with the quasi-Monte Carlo method with 4096 samples. 40

27 Test with nine uncertainty parameters wind speed, wake effect, C_t curve, surface roughness, power curve, power loss, capital costs, O&M costs and discount rate at the same time (simultaneously) compared to single effects of the uncertainty parameters computed with the quasi-Monte Carlo method with 4096 samples. 42
28 Test with ten uncertainty parameters wind speed, wake effect, C_t curve, surface roughness, power curve, power loss, capital costs, O&M costs, discount rate and energy price at the same time (simultaneously) compared to single effects of the uncertainty parameters computed with the quasi-Monte Carlo method with 4096 samples. 44
29 Test with nine uncertainty parameters wind speed, wake effect, C_t curve, surface roughness, power curve power loss, capital costs, o&m costs and energy price at the same time (simultaneously) compared to single effects of the uncertainty parameters computed with the quasi-Monte Carlo method with 4096 samples. 46

List of Tables

1 Ten uncertain input parameter which will be perturbed by a random variable ξ of its corresponding normal distribution $\mathcal{N}(\mu, \sigma)$. The parameter σ is obtained by the max deviation value d which is given in %. This value indicates that in the worst case the input parameter values will be perturbed by $\pm d\%$ which is achieved with the right determination of the normal distribution. A value of $d\%$ corresponds to the normal distribution parameter $\sigma = \frac{d}{300}$ so that 99.73% of the generated random variables are in a range of $\pm d\%$ if they are multiplied with the base case input value. 13

2 1D - Sobol sequence number generation. 27

3 Duration in seconds in dependency of generated 2-D Latin Hypercube sampling numbers and 2-D Sobol sequence numbers. For comparison: One calculation step of the annual energy production (AEP) averaged needs 0.726141 seconds but because of parallel computation the calculation depends linear on the number of threads. For example the calculation of the AEP with four threads effectively needs 0.181535 seconds. The used CPU is a 1,8 GHz Intel Core i5 of the second generation. . . 29

4 Precision measured with the root-mean-squared error in dependency of sample size. 36

5 Parameters for the Horns Rev wind farm. 39

6 Results of variances based sensitivity analysis of net annual energy production calculation with 4096 samples quasi-Monte Carlo Method. . . . 41

7 Results of variances based sensitivity analysis of levelized cost of energy (LCOE) calculation with 4096 samples quasi-Monte Carlo Method. . . 43

8 Results of variances based sensitivity analysis of net present value (NPV) calculation with 4096 samples quasi-Monte Carlo Method. 45

9 Results of variances based sensitivity analysis of IRR calculation with 4096 samples quasi-Monte Carlo Method. 47

1 Introduction

The climate change is one of the biggest environmental struggles at the time of humanity. The industrial revolution has left its traces and now it is up to us humans to counteract against climatic problems. We already have made big steps in reducing CO_2 emissions but it is still a long way until the whole world has switched to renewable so called green energy.
The sun and the wind are resources that will most likely be available longer as our species will live on this planet. That is the reason why solar and wind energy are the leading forms of renewable energy. Wind energy produced by wind farms are more practicable in regions of our western world because solar energy can only be efficiently used in certain parts of the world and also not during the night.
Therefore wind farms have the potential to become the worlds leading energy source. They are subdivided in onshore and offshore wind farms. The difference between them is that onshore wind farms are built up on the countryside whereas offshore wind farms are built up on the underwater landmass close to the coast. Both variants have their advantages and disadvantages.
Onshore wind farms are characterized by their lower capital and maintenance costs and easier access to the produced energy. Offshore wind farms on the other hand have higher capital and maintenance costs because they are built in the sea, which is more difficult to access. Also the produced energy has a bigger distance to the countryside where the energy is needed. But on longer view the offshore variant has some great advantages. On the sea, dominates a much stronger wind compared to the wind at the countryside which increases the total power output. Also there are no obstacles through mountains, hills etc., that interfere the wind which leads again to a higher energy production. There is also more space available on the sea so that bigger wind farms can be build. Another positive effect outsourcing the wind farm to the sea is that the wind farms have no bad visual impact or acoustic noise or even can not cause damages to human life.
In this thesis we will focus on offshore wind farms and we will investigate if and which uncertainties setting up such a wind farm can generate a miscalculating of the whole project. The investigation of uncertain model inputs will show us the impact of them regarding to specific quantities of interest and will tell us which of the uncertainties have to be focused on to plan future projects more precisely.

2 Background

Wind energy is an active area of research and very relevant today. Much of it is focused on solving the wind farm layout optimization problem. It deals with the wind farm design aiming to find the optimal positions of the turbines within the wind farm so that the power production is maximized.
The transformation of wind power into electrical power is done by wind turbines which are usually grouped into wind farms in order to have lower installation and maintenance costs. But the grouping of turbines leads to a reduction in the power production because of the presence of wake effects within the wind farm. When wind passes the turbine blades, the turbine blades are rotating and the turbine generates power from the wind. The sideeffect is that the rotating blades cause a wake of turbulence, that reduces the wind speed behind this turbine and therefore also reduces the power production by the turbines affected by the wake. That means that the wind speed behind a turbine is lower than the wind speed in the unperturbed wind stream. In large wind farms wake effects lead to significant power losses. That is why positioning the turbines to minimize the power loss and thus maximize the expected power output has an important economic purpose [20].
In a previous bachelor thesis a wake model named PARK was implemented which simulates the annual energy production (AEP) of an offshore wind farm [7]. Furthermore in a master thesis a first approach were done to make a uncertainty quantification on a few input parameter of the PARK wake model [25].
In this thesis the uncertainty quantification of the PARK wake model is enhanced by more sampling methods for the uncertainty quantification and more uncertainty parameters, to obtain more meaningful results by analyzing the significance of the uncertainties that arise in an offshore wind farm.

3 Offshore Wind Farm Modeling

In this section the modeling of an offshore wind farm will be presented which is divided in four sub-models. At first the wind model will be presented where is shown how we translate the raw wind data into a mathematical model. This representation of the wind will then be used by the next sub-model, the wake model. After that the power generation model will be presented which uses the wake model for the calculation of the gross annual energy production (AEP). Then the cost models will be introduced that consider the different quantity of interests of our offshore model. In Figure 1 the brief structure of the offshore model is shown which should help as an orientation and for an understanding of this section. The gray rectangles stand for the different sub-models. Each sub-model has it's own input parameters which are encircled with the green ellipses. The red ellipses are representing the outputs.

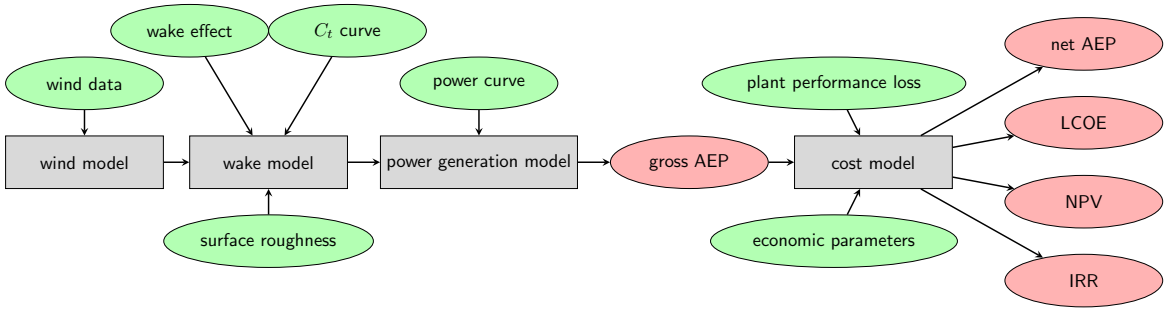

Figure 1: Structure of the offshore model. Gray rectangles representing the four sub-models with their specific inputs (green ellipses) and their outputs (red ellipses).

3.1 Wind Model

The wind data is important to calculate the generated power converted by the turbines. It is used as an input for the wind model and reflects the real worlds climate conditions for a specific area and time period.

3.1.1 Wind Data Processing

Typically the wind data consists of thousands measurements about the wind direction and the wind speed. Figure 2 shows a wind direction distribution over a time period of seven years at the FINO3 research platform which is located 45 sea miles (80 kilometres) west of Sylt in the North Sea. The wind directions are divided into 12, 32 and 360 sectors, standing for the cardinal points and representing the probability by the size of each sector.
The wind speed distribution is assumed as a weibull distribution as defined in equation (1) and is calculated by the raw wind data as described in Heiming [7] with the maximum likelihood estimation which is a method of estimating the parameters of a

statistical model with given observations, by finding the parameter values that represents the observations the most. In Figure 3 the wind speed distribution for the sector $\alpha \in [225°, 255°)$ for the same time period as for the wind direction distribution is shown.
For the wind representation the weibull distribution parameters λ and k are provided for each wind direction which means that each wind direction has its own probability and its own wind speed distribution. The probabilities for wind direction and wind speed are combined to a speed-direction probability for the computation by using the probability of the wind direction and multiplying it with the probability for the considered wind speed, so that the positioning of the turbines have a big impact to the accuracy of the calculated wake effect.

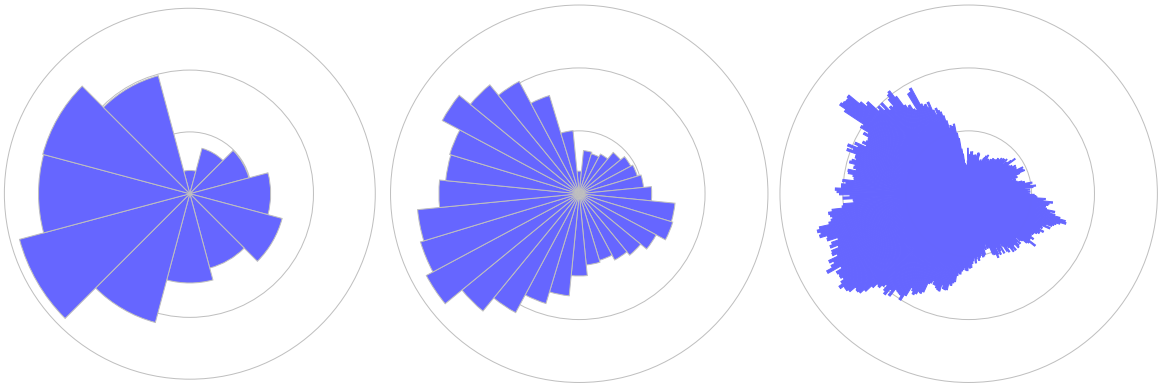

Figure 2: Combined wind direction distribution at the FINO3 research platform over the years of July 2010 to June 2017, clustered into 12, 32 and 360 direction sectors. Source: Heiming[7]

3.1.2 Weibull Distribution

The weibull distribution is a continuous probability distribution and has a lot of applications e.g. the approximation of the frequencies of different wind speeds as can be seen in Figure 3. The probability density function in general is defined as follows:

$$f(x;\lambda,k) = \begin{cases} \left(\frac{k}{\lambda}\right) \cdot \left(\frac{x}{\lambda}\right)^{k-1} \cdot \exp\left(-\left(x \cdot \lambda\right)^{k}\right) & x \geq 0, \\ 0 & x < 0 \end{cases}$$

The parameter $\lambda > 0$ and $k > 0$ are the scale- and shape parameter of the distribution. The advantage of the weibull distribution is that the probability distribution of the different wind speed levels of measured wind data can be approximated very well with these two parameters, higher speeds are less likely and that the probability of negative wind speeds is equal to zero [7].
In our model each wind direction sector α_i, $i \in \{1, 2, \ldots, N_{\text{directions}}\}$, has its own weibull distribution, which will be estimated with the maximum likelihood estimation

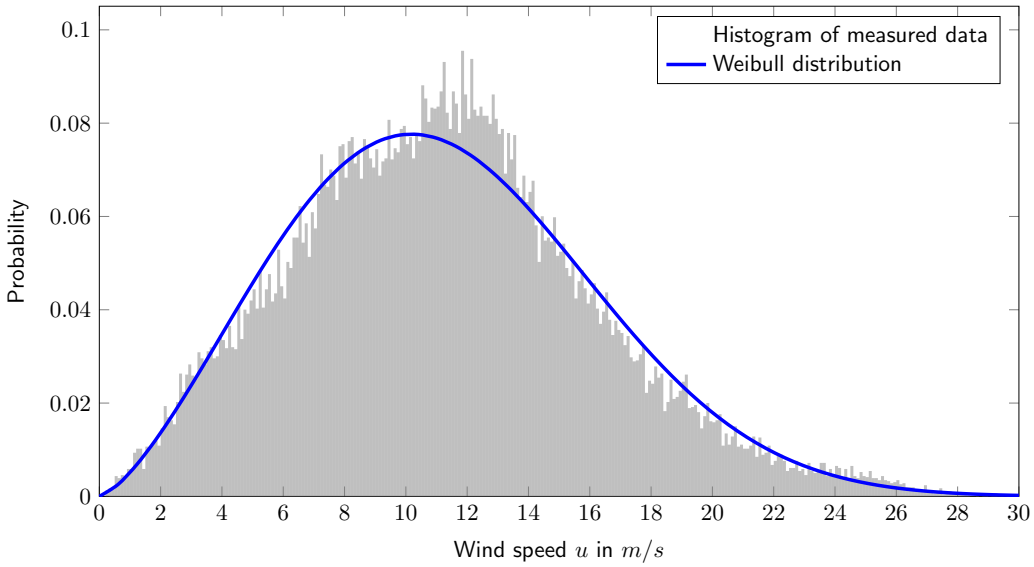

Figure 3: Approximated wind speed distribution as a weibull distribution using the maximum likelihood estimation. The measured data is from the FINO3 research platform over the years of July 2010 to June 2017 for the wind direction sector $\alpha_i \in [225°, 255°)$. The measured wind speed distribution is illustrated with the histogram.

[7] where the parameter λ and k will be determined, so that the probability $f_{\alpha_i}(u)$ for every wind speed u can be represented as:

$$f_{\alpha_i}(u) = \left(\frac{k}{\lambda}\right) \cdot \left(\frac{u}{\lambda}\right)^{k-1} \cdot \exp\left(-\left(u \cdot \lambda\right)^{k}\right) \tag{1}$$

It is important to note, that this wind model by itself is not deterministic as the wind source is modeled by a distribution. When we refer to the overall model as deterministic, this implies that the input parameters are not perturbed on purpose.

3.2 Wake Model

The wake model is important to calculate the velocity deficits within a wind farm caused by the mast and blades of the turbines that reduce the wind speed behind each turbine. In this thesis we focus on one wake model called the PARK wake model which was developed by Jensen [9] and Katic et al. [10] in 1986. Starting from that Samorani [20] has also worked on the PARK model. The presentation of the wake model is based on the work of Heiming [7].
The wake model does not consider the different wind speeds on different vertical levels so that the wind measurements from the FINO3 research platform are not adapted to the height of the turbines. The previous presented wind data are measured in a height of 100 m so that the discrepancy between the turbine's height and the measured

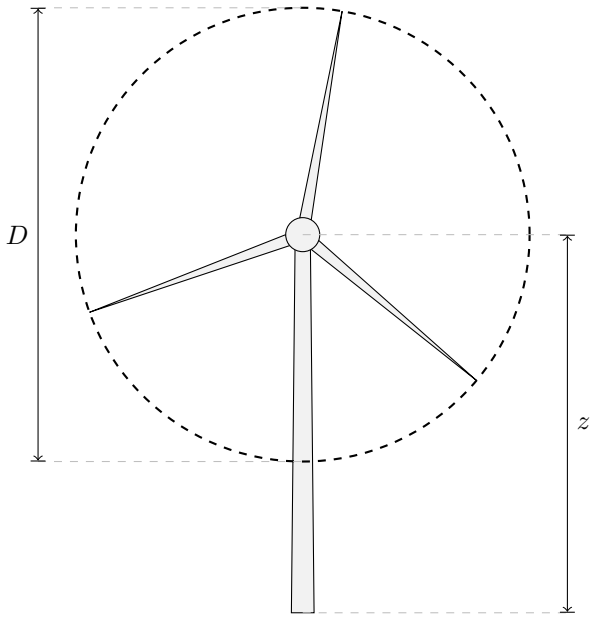

Figure 4: Turbine measurements: D represents the rotor diameter and z is standing for the hub height of the mast. Source: Heiming [7]

wind data is not that big. For example the height of the turbines at the Horns Rev1 wind farm is 70 m. This is an example for the need of considering uncertainties in our mathematical models which will be presented in Section 4. Also in the model the velocity deficit inside the wake only changes in stream direction of the wind, i.e. the velocity deficit only depends on the distance x from the perturbing turbine which is illustrated in Figure 5. Therefore, it is not suitable for the calculation of a flow field with an exact velocity distribution. The model is designed for the far wake case and is only valid for distances of the turbines of three rotor diameters or more. The wake radius grows with a factor κ so the wake diameter $D_{\mathrm{w}} = D_{\mathrm{w}}(x)$ grows linearly by 2κ (see Figure 5). The scalar κ is defined as follows [20]:

$$\kappa = \frac{0.5}{\ln \frac{z}{z_0}} \tag{2}$$

where z is the hub height of the turbine (see Figure 4) and z_0 is called surface roughness. The surface roughness depends on the site ground. It can be assumed as constant for most cases. But there is also a formula presented in [12] to compute the surface roughness depending on the sea state based on the the waves caused by the wind velocity. Heiming [7] assumed z_0 to be constant in his thesis and set it to a typical value of 0.03 m for offshore sites.
The equation for the velocity deficit at any point inside the wake of a turbine inside the free stream with velocity u_0 is

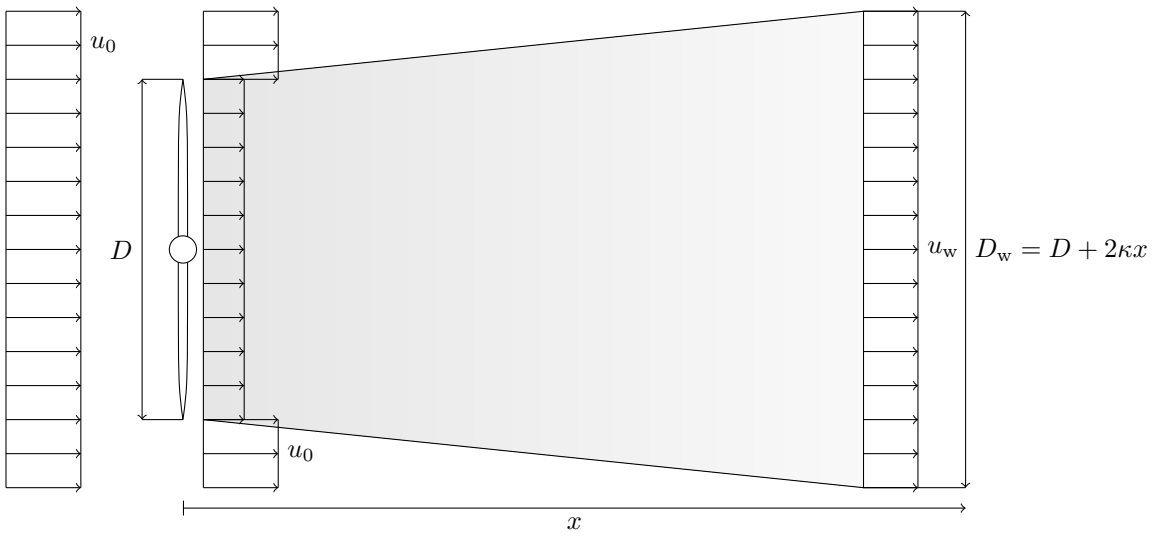

Figure 5: Visualization of the wake as described in PARK. The unperturbed free stream u_0 hits the turbine and its blades and generates the wake and the resulting wind speed u_w. The wake grows linear with the factor 2κ. D represents the rotor diameter of the turbine, D_w is the width of the generated wake and x stands for the distance from the turbine. Source: Heiming [7]

$$\delta_u(x) = 1 - \frac{u_w}{u_0} = \frac{1 - \sqrt{1 - C_t(u_0)}}{(1 + \frac{2\kappa x}{D})^2} \overset{(2)}{=} \frac{1 - \sqrt{1 - C_t(u_0)}}{(1 + \frac{x}{D \cdot \ln(z/z_0)})^2}, \tag{3}$$

where $C_t(u)$ is the thrust coefficient of the turbine (see Figure 7), D is the rotor diameter (see Figure 4) and u_w describes the velocity inside the wake. For detailed description see Heiming [7]. As the above derivation is only valid for the wake of a turbine in the free stream, we now have to generalize it to get the velocity deficit at a turbine inside the wake of another turbine that is not necessarily inside the free stream.

A few things have to be changed so the above requirements are met: We call the wake generating turbine i and the wake affected turbine j. The incident velocity of turbine i then is u_{inc_i}. If turbine j is fully affected by the wake, we can use the above formulation. However if turbine j is only partially affected, we need to introduce a shadowing factor $\beta_\kappa \in [0, 1]$ as presented by Choi and Shan [2]:

$$\beta_\kappa = \frac{A_{\text{Intersection}}}{A_{\text{Turbine}}},$$

where $A_{\text{Intersection}}$ is the circular intersection of the wake cross section with the turbine's circular area A_{Turbine} (see Figure 6). The equation for the velocity deficit has changed as follows:

$$1 - \frac{u_w}{u_{\text{inc}_i}} = \frac{\beta_\kappa \left(1 - \sqrt{1 - C_t(u_{\text{inc}_i})}\right)}{\left(1 + \frac{2\kappa x}{D}\right)^2}.$$

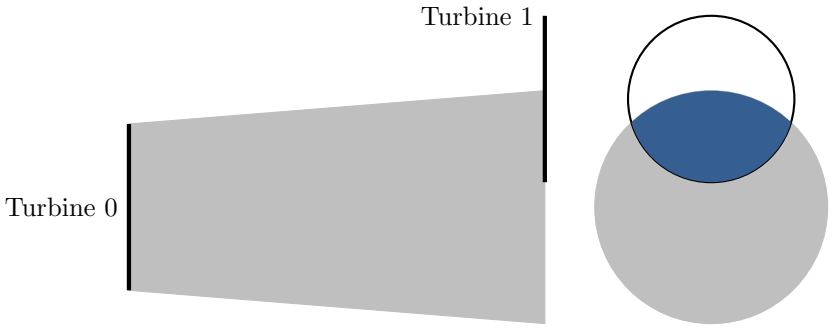

Figure 6: Intersection of wake effect on other turbine. Source: Heiming [7]

As this formulation depends on the incident velocity at turbine i, we need to transform it to make it dependent on the free stream velocity u_0:

$$\delta_{\mathrm{u}_{ij}} = 1 - \frac{u_{\mathrm{w},i}}{u_0} = 1 - \frac{u_{\mathrm{inc}_j}}{u_0} = \frac{u_0}{u_{\mathrm{inc}_i}} \left(\frac{\beta_\kappa \left(1 - \sqrt{1 - C_t(u_{\mathrm{inc}_i})}\right)}{\left(1 + \frac{2\kappa x}{D}\right)^2} \right).$$

For the case of two interacting wakes the velocity deficits are added as follows:

$$\begin{aligned} \delta^2_{\mathrm{u}_j} &= \delta^2_{\mathrm{u}_{1j}} + \delta^2_{\mathrm{u}_{2j}} \\ \Leftrightarrow \left(1 - \frac{u_{\mathrm{inc}_j}}{u_0}\right)^2 &= \left(1 - \frac{u_{\mathrm{w},1}}{u_0}\right)^2 + \left(1 - \frac{u_{\mathrm{w},2}}{u_0}\right)^2, \end{aligned}$$

with $u_{\mathrm{w},1}$ and $u_{\mathrm{w},2}$ representing the velocities inside the two wakes. In a more general case with an arbitrary number of interacting wakes we can compute

$$\delta_{\mathrm{u}_j} = \sqrt{\sum_{i=1}^{N} \delta^2_{\mathrm{u}_{ij}}},$$

where N is the number of interacting wakes and $\delta_{\mathrm{u}_{ij}}$ represents the velocity deficits of the single wakes i affecting turbine j.

3.3 Power Generation Model

The turbines are converting the wind's kinetic energy into electrical energy. When wind passes the turbine blades, the turbine blades are rotating and the turbine generates power from the wind. There are different types of wind turbines. The most common type is a horizontal axis wind turbine which has three blades and is installed on top of a tower, set against the wind direction. Horizontal axis wind turbines are characterized by their hub height z, which is the height from bottom to the top of the tower, where the turbine is placed and their rotor diameter D (see Figure 4). This type is also used in the wake model. A turbine has also a few more attributes. Important for our model are the cut-in speed u_{cutin} and the cut-out speed u_{cutout}. These values specify

the range of wind speed, in which the turbine is generating power. If the wind speed is lower than u_{cutin} there is not enough wind for efficient power production, and for wind speeds larger than u_{cutout} the turbine may become damaged. Typically the cut-in speed is about 3 to 4 m/s and the cut-out speed about 25 m/s. Furthermore, the turbine's performance can be described by curves which are specifying the power output with the power curve $P(u)$ and a C_t curve $C_t(u)$ specifying the thrust coefficient, both depending on the wind speed u (see Figure 7). There are often several curves for different air density values. In this work it is assumed that the density is constant at 1.225 kg/m^3, which is the air density at a temperature of 15 °C at sea level.

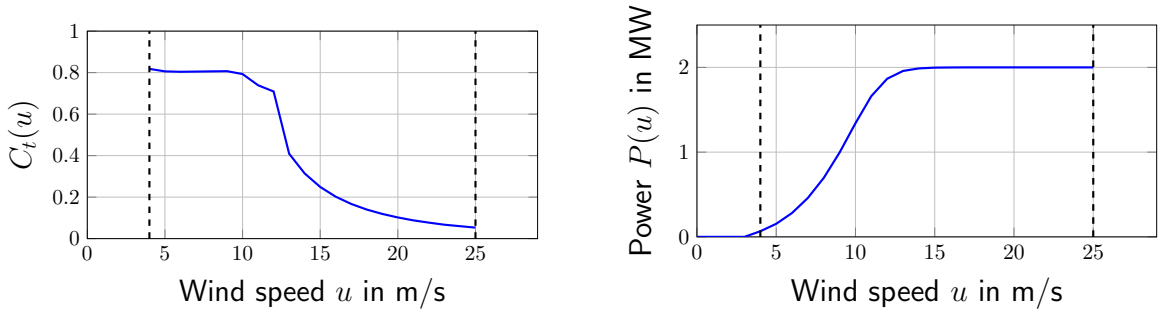

Figure 7: Thrust coefficent C_t and power production of the turbine Vestas V80 with cut-in speed of 4 m/s and cut-out speed of 25 m/s (dashed vertical lines).

3.4 Gross annual Energy Production (AEP)

In this section the mathematical model for the calculation of the gross AEP will be presented. With the wind speed distribution described as the weibull distribution $f_{\alpha_i}(u)$ for each wind sector α_i, the calculated velocity deficits δ_u and the power curve $P(u)$ for each wind speed u the gross AEP can be calculated. This is the most important value for the cost models:

$$\begin{aligned} E_{\text{AEPgross}} &= (8760h + 6h) \cdot P \\ &\approx (8760h + 6h) \cdot \sum_{i=1}^{N_{\text{directions}}} w_{\alpha_i} \cdot \sum_{j=0}^{N_{\text{speeds}}} w_j \cdot f_{\alpha_i}(u_j) \cdot \sum_{k=1}^{N_{\text{turbines}}} P_{\alpha_i}(u_{\text{inc}_{j,k}}), \end{aligned} \tag{4}$$

where P describes the mean power for a given wind distribution in MW. This must be projected to the length of one year to get the AEP measured in MWh that is why P is multiplied by $(8760h + 6h)$, where 8760 is the number of hours in a year and the additional six hours is a correction to get the mean number of hours per year with a leap years respected. This complies to 365,25 days per year.
The expected power value P for **all** directions is

$$
\begin{aligned}
P &:= \int_0^{2\pi} P_\alpha d\alpha \\
&\approx \sum_{i=1}^{N_{\text{directions}}} w_{\alpha_i} \cdot P_{\alpha_i}
\end{aligned}
$$

with $N_{\text{directions}}$ describing the number of wind directions and w_{α_i} standing for the weight for direction α_i. P_{α_i} is the expected power value of a wind farm for all wind speeds u in **one** wind direction α_i which is given by

$$
\begin{aligned}
P_{\alpha_i} &:= E[P_{\alpha_i}(u)] \\
&= \int_0^{\infty} P_{\alpha_i}(u) \cdot f_{\alpha_i}(u) du \\
&= \int_{u_{\text{cutin}}}^{u_{\text{cutout}}} P_{\alpha_i}(u) \cdot f_{\alpha_i}(u) du \\
&\approx \sum_{j=1}^{N_{\text{speeds}}} w_j \cdot P_{\alpha_i}(u_j) \cdot f_{\alpha_i}(u_j)
\end{aligned}
$$

where $f_{\alpha_i}(u_j)$ describes the probability density function (see equation (1)) of the wind speed distribution, u_{cutin} and u_{cutout} are the turbine's cut-in and cut-out speed respectively, w_j is the weight for speed j and

$$
P_{\alpha_i}(u) = \sum_{k=1}^{N_{\text{turbines}}} P_{\alpha_i}(u_{\text{inc}_k})
$$

denotes the total power output of the wind farm at wind speed u and direction α. $P_{\alpha_i}(u)$ is the power production of one turbine at wind speed u and u_{inc_k} is the incident wind speed at turbine k at wind direction α_i and wind speed u_{inc}. The incident wind speed is computed by

$$
u_{\text{inc}} = (1 - \delta_u) u_0,
$$

where u_0 is the unperturbed free stream speed and δ_u is the velocity deficit at a turbine rotor with perturbed wind inside another turbine's wake. The calculation of the velocity deficit was described in section 3.2.

3.5 Cost Model

There are different quantity of interests for the evaluation of wind farms which can later be used as objective functions to optimize a wind farm by minimizing or maximizing the quantities of interest [22]. In this thesis the quantity of interest can be measured by cost models which calculate different indicators that point out the economic benefit of a wind farm. The quantities of interest of our model are illustrated with the red ellipses in Figure 1. In the following the net AEP calculation will be presented followed by the calculation of the levelized cost of energy, net present value and the internal rate of return.

3.5.1 Net annual Energy Production (AEP)

The net AEP results by reducing the gross AEP with a constant factor which includes availability losses as a result of turbine and grid downtime of the wind farm, electrical efficiency due to line losses and curtailment of grid and ramp-rate which can be summarized to the plant performance losses p_{loss}. The calculation for the net AEP can be formulated as:

$$E_{\text{AEPnet}} = E_{\text{AEPgross}} \cdot (1 - p_{\text{loss}}). \tag{5}$$

3.5.2 Levelized Cost of Electricity (LCOE)

The LCOE represents the price of electricity per kWh over the whole lifetime of a wind farm. The less the cost of electricity is the worthwhile gets the whole project. The formula of the LCOE by Lackner and Elkinton [11] can be represented as:

$$K_{\text{LCOE}} = \frac{C_{\text{capital}} \cdot \dfrac{(1 + r_{\text{rate}})^{\ell} \cdot r_{\text{rate}}}{(1 + r_{\text{rate}})^{\ell} - 1} + C_{\text{o\&m}}}{E_{\text{AEPnet}}}, \tag{6}$$

where C_{capital} is representing the total capital costs for turbines, cabling, substation, decommission etc., $C_{\text{o\&m}}$ are the annual operation and maintenance costs and discount rate r_{rate} including debt, taxes and insurance over the lifetime ℓ in years.

3.5.3 Net Present Value (NPV)

The NPV by Gonzales et al. [6] is a indicator for the today's monetary value of a project. The higher the value the better is the project. In general a negative value is a bad indicator for the calculated project. A positive value is better but have not a big power of expression without comparing the value with another project or alternative settings of the considered project. The NPV can be calculated as follows:

$$C_{\text{NPV}} = C_{\text{PRV}} - C_{\text{capital}} + \sum_{t=1}^{\ell} \frac{E_{\text{AEPnet}} \cdot K_{\text{energy}} - C_{\text{o\&m}}}{(1 + r_{\text{rate}})^{t}}, \tag{7}$$

where C_{PRV} is the present residual value of the wind farm after the lifetime l in years, C_{capital} includes the total capital costs for turbines, cabling, substation, decomission etc., $C_{\text{o\&m}}$ are the annual operation and maintenance costs with respect to the discount rate r_{rate} and K_{energy} represents the price of energy on the market. Usually E_{AEPnet}, K_{energy} and $C_{\text{o\&m}}$ are time-dependent, but because we are working with averaged values (e.g. the wind data is aggregated from several years), we assume them to be constant in time.

3.5.4 Internal Rate of Return

The internal rate of return r_{IRR} is also a measure for the profitability of an investment for a project. The r_{IRR} is the discount rate r_{rate} in (see equation (7)) which makes the net present value C_{NPV} of all cash flows of the considered project equal to zero. If the r_{IRR} is greater than the real discount rate r_{rate} plus risk deficits it can be said that the considered project is worth to be done in an economical view. The r_{IRR} can be calculated with a lot of numerical methods e.g. with the interpolation method where a rate r_{IRR} have to be estimated first and then have to be modified with certain rules until the required accuracy of r_{IRR} is reached.

4 Offshore Wind Farm Modeling with Uncertainties

In this section the offshore model presented in Section 3.2 will be modified to a stochastic offshore model considering uncertainties regarding the input parameters. The presented formulas and functions of our four sub-models have to be transformed for the uncertainty quantification. How this is achieved will be presented in the next subsections.

4.1 Uncertain Input Parameters

As I will be mentioning in Section 5.2 we have to examine the uncertain input parameters of our offshore model. They include measured data which can be affected with a certain error. To examine these parameters we have to perturb them with random variables [21].

For this investigation we have listed some uncertainties with their predicted bounds

Uncertainty	Max deviation 3σ	Analogical normal distribution
1 Wind speed	5%	$\xi_{\text{wind}} \sim \mathcal{N}(1, 0.01\bar{6})$
2 Wake Effect	5%	$\xi_{\text{wake}} \sim \mathcal{N}(1, 0.01\bar{6})$
C_t Curve	5%	$\xi_{\text{ct}} \sim \mathcal{N}(1, 0.01\bar{6})$
Surface roughness	5%	$\xi_{\text{rough}} \sim \mathcal{N}(1, 0.01\bar{6})$
3 Power Curve	5%	$\xi_{\text{power}} \sim \mathcal{N}(1, 0.01\bar{6})$
4 Plant performance losses	5%	$\xi_{\text{loss}} \sim \mathcal{N}(1, 0.01\bar{6})$
Capital costs	5%	$\xi_{\text{capital}} \sim \mathcal{N}(1, 0.01\bar{6})$
Annual O&M costs	5%	$\xi_{\text{o\&m}} \sim \mathcal{N}(1, 0.01\bar{6})$
Discount rate	5%	$\xi_{\text{rate}} \sim \mathcal{N}(1, 0.01\bar{6})$
Energy price on market	5%	$\xi_{\text{energy}} \sim \mathcal{N}(1, 0.01\bar{6})$

Table 1: Ten uncertain input parameter which will be perturbed by a random variable ξ of its corresponding normal distribution $\mathcal{N}(\mu, \sigma)$. The parameter σ is obtained by the max deviation value d which is given in %. This value indicates that in the worst case the input parameter values will be perturbed by $\pm d\%$ which is achieved with the right determination of the normal distribution. A value of $d\%$ corresponds to the normal distribution parameter $\sigma = \frac{d}{300}$ so that 99.73% of the generated random variables are in a range of $\pm d\%$ if they are multiplied with the base case input value.

of inaccuracy. In Table 1 the parameters are listed and are divided into four categories respectively standing for the four sub-models of our offshore model. The parameters are derived from the uncertainties from [3]. The values of the maximum deviation for the parameters lead to a perturbation of the base case values of the input parameters which are the unperturbed values that are measured and estimated for the deterministic models. For example the error affection of 5% for a specific parameter value means that in the worst case the uncertain parameter value vary about $\pm 5\%$ to the original

base case value. That means that for the investigation the base case parameter values will be perturbed with random variables ξ_i of their corresponding normal distribution $\mathcal{N}(\mu, \sigma)$. A value of $d\%$ corresponds to a normal distribution with mean $\mu = 1$ and a sigma σ with the value of $\frac{d}{300}$. That leads to a perturbation of the input value of $\pm d\%$ if the base case input value is multiplied with the generated random variable. The determination of σ due to the fact that a value of 3σ covers $99,73\%$ of all possible occurring random variables of a specific distribution (see Figure 8). With the choice of normal distributed random variables, the base case values of the input parameters are with the highest probability while larger deviations get more unlikely as we assume that values closer to the base case values are more likely than larger deviations from the base case values.

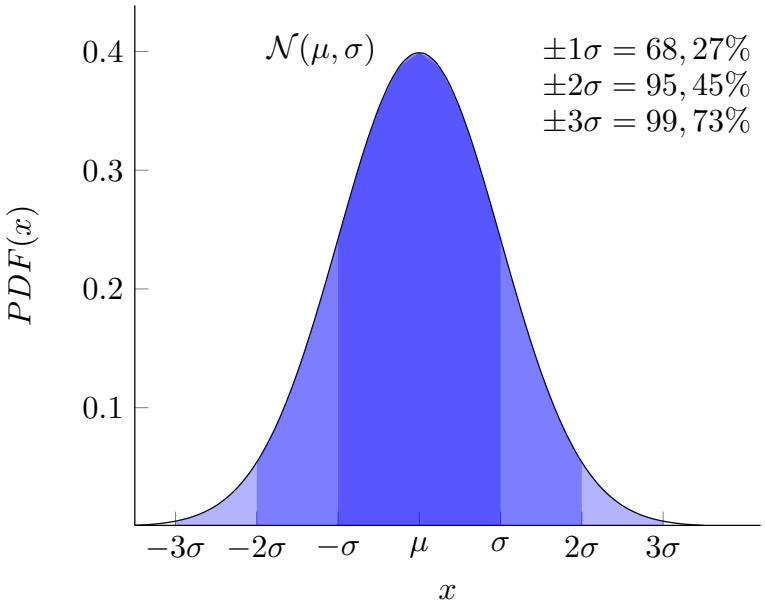

Figure 8: Normal distribution $\mathcal{N}(0, 1)$ with coverage of generated values.

4.2 Offshore Model with Uncertainties

The above defined random variables are now included into the formulas and functions of our offshore model.

4.2.1 Perturbed Wind data

One of the main input of our offshore model is the wind data. In Section 3.1 the wind data processing and the approximation of the wind data to a specific wind speed distribution is shown. Our input for the wake model consists information about the weibull distribution parameters λ and k and the probability for every considered direction. Lets assume that the measurements of our raw wind data is error affected by a fixed

factor. Therefore that the wind data will be processed beforehand to the specified input which includes weibull distribution parameters and probabilities for each wind direction we have to find out how to perturb the weibull distribution parameters to purpose a error on the measured raw wind data.
Lets assume that we processed our wind data to a weibull distribution with $\lambda = \lambda_0$ and $k = k_0$. When we perturb our raw wind data with a perturbation factor d, we can show that this leads to a new weibull distribution with perturbed weibull parameters $\tilde{\lambda} = d \cdot \lambda_0$ and $\tilde{k} = k_0$ (see Figure 9).

Proof. We consider the maximum likelihood estimation (see Heiming [7]) for the parameter λ and k and raw wind data $x_1, x_2, \ldots, x_n$:

$$k^{-1} = \frac{\sum_{i=1}^{n} x_i^k \ln x_i}{\sum_{i=1}^{n} x_i^k} - \frac{1}{n} \sum_{i=1}^{n} \ln x_i \tag{8}$$

$$\lambda = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^k \right)^{1/k} \tag{9}$$

Now we perturb our raw wind data with the factor d which leads to perturbed raw wind data $\tilde{x}_i = d \cdot x_i$ and show that this only influences the parameter λ:

$$\begin{aligned}
\tilde{k}^{-1} &= \frac{\sum_{i=1}^{n} \tilde{x}_i^k \ln \tilde{x}_i}{\sum_{i=1}^{n} \tilde{x}_i^k} - \frac{1}{n} \sum_{i=1}^{n} \ln \tilde{x}_i \\
&= \frac{\sum_{i=1}^{n} (d \cdot x_i)^k \ln(d \cdot x_i)}{\sum_{i=1}^{n} (d \cdot x_i)^k} - \frac{1}{n} \sum_{i=1}^{n} \ln(d \cdot x_i) \\
&= \frac{\sum_{i=1}^{n} d^k x_i^k (\ln d + \ln x_i)}{\sum_{i=1}^{n} d^k x_i^k} - \frac{1}{n} \sum_{i=1}^{n} (\ln d + \ln x_i) \\
&= \frac{\sum_{i=1}^{n} x_i^k \ln d + x_i^k \ln x_i}{\sum_{i=1}^{n} x_i^k} - \frac{1}{n} \left(\sum_{i=1}^{n} (\ln x_i) + n \ln d \right) \\
&= \frac{\sum_{i=1}^{n} x_i^k \ln x_i}{\sum_{i=1}^{n} x_i^k} + \frac{\sum_{i=1}^{n} x_i^k \ln d}{\sum_{i=1}^{n} x_i^k} - \ln d - \frac{1}{n} \sum_{i=1}^{n} \ln x_i \\
&= \frac{\sum_{i=1}^{n} x_i^k \ln x_i}{\sum_{i=1}^{n} x_i^k} - \frac{1}{n} \sum_{i=1}^{n} \ln x_i \\
&= k^{-1}
\end{aligned} \tag{10}$$

Such that the parameter $\tilde{k} = k$.

$$
\begin{aligned}
\tilde{\lambda} &= \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i^k\right)^{1/k} \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}(d\cdot x_i)^k\right)^{1/k} \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}d^k x_i^k\right)^{1/k} \\
&= d\left(\frac{1}{n}\sum_{i=1}^{n}x_i^k\right)^{1/k} \\
&= d\cdot\lambda
\end{aligned}
\tag{11}
$$

Finally, the parameter $\tilde{\lambda}$ results from λ multiplied by the factor d. Same results for estimated weibull parameter of the perturbed wind data can be found in [23].

□

4.2.2 Uncertain Wind Model

Due to inaccurate measurements, imprecise long-term predictions, interannual variability, and further interferences [3], the distribution of the wind speed u is a highly uncertain parameter. Therefore we perturb the raw data of the wind speed with a normally distributed random variable ξ_{wind} such that perturbed probability density functions of the weibull distribution are obtained, see Figure 9. In Section 4.2.1 is shown that a perturbation d of the wind speed corresponds to the perturbation d of the weibull parameter λ, such that the resulting probability for each wind speed u can be formulated as random variable, compare with equation (1):

$$
\widetilde{f}_{\alpha_i}\left(u,\xi_{\text{wind}}\right) = \left(\frac{k}{\lambda\cdot\xi_{\text{wind}}}\right)\cdot\left(\frac{u}{\lambda\cdot\xi_{\text{wind}}}\right)^{k-1}\cdot\exp\left(-\left(u\cdot\lambda\cdot\xi_{\text{wind}}\right)^k\right) \tag{12}
$$

where $\lambda > 0$ is the scale parameter and $k > 0$ is the shape parameter of the weibull distribution which are determined with the maximum likelihood estimation using the unperturbed wind speed data.

4.2.3 Uncertain Wake Model

The predicted wake effect is perturbed due to uncertainty in the model inputs (including wind direction), and that associated with the wake model performance. Furthermore we need to include uncertainty concerning any proposed neighboring sites (layout, turbine type) [3]. Therefore we perturb the velocity deficit of a wake with a normally distributed random variable ξ_{wake}. Furthermore, the wake model depends on the C_t curve and the surface roughness z_0. Due to imprecise measurements of the C_t curve, we perturb this parameter with the a normally distributed random variable ξ_{ct},

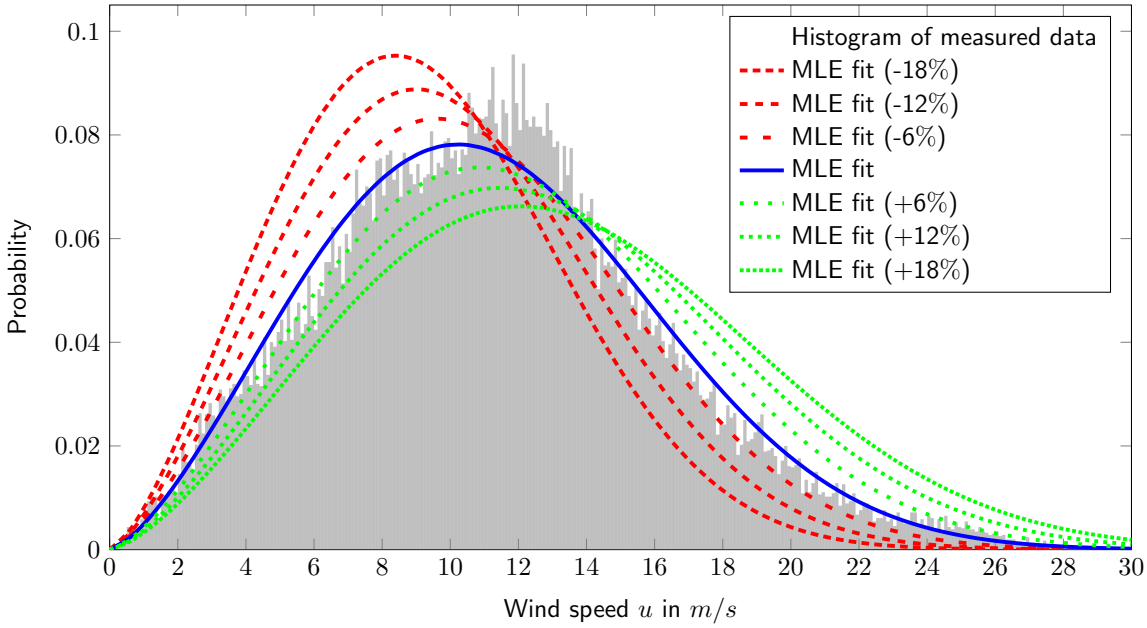

Figure 9: Fitted weibull distribution with maximum likelihood estimation (MLE) for wind speed distribution over the years of July 2010 to June 2017 for one wind direction sector $\alpha_i \in [225° - 255°)$ measured at the FINO3 research platform. Red and green plots representing the fitted weibull distribution with MLE after perturbing the wind speed data with a factor of ±6%, ±12% and ±18%. The random variable $\xi_{wind}(u)$ represents weibull distribution between the highest red plot (MLE fit -18%) and the lowest green plot (MLE fit +18%) with a probability of 99.73%.

see Figure 10. The value of $C_t(u_0) \cdot \xi_{\text{ct}}$ must not be grater than or equal to one because in formula (13) can be seen that this would lead to an undefined value (root from negative number) for the velocity deficit so that the perturbation of the uncertain C_t curve is truncated so that it can not be equal to one. The surface roughness depends on the topography and flora [24], i.e. for offshore wind farms it depends on the wave field, wind speed, upstream fetch and water depth [13]. Thus, also this parameter z_0 should be stochastic and is therefore perturbed with a normally distributed random variable ξ_{rough}. Altogether, the velocity deficit behind a turbine at any point x from equation (3) changes as follows:

$$\widetilde{\delta_u}(x) = \frac{1 - \sqrt{1 - C_t(u_0) \cdot \xi_{\text{ct}}}}{(1 + \frac{x}{D \cdot \ln(z/z_0 \cdot \xi_{\text{rough}})})^2} \cdot \xi_{\text{wake}}, \tag{13}$$

4.2.4 Uncertain Power Generation Model

The turbine performance is perturbed due to material fatigue which leads to uncertainty in the power curve. Furthermore, there is uncertainty on the performance under

site conditions for which the power curve might not be valid. This also includes the impact of atmospheric stability and uncertainty associated with icing losses as well as other environmental losses, e.g. blade soiling, blade degradation, weather effects. Altogether we perturb the power curve by the normally distributed random variable ξ_{power} (see Figure 10) such that the power curve changes as follows:

$$\widetilde{P}\left(u, \xi_{\text{power}}\right) = P(u) \cdot \xi_{\text{power}}. \tag{14}$$

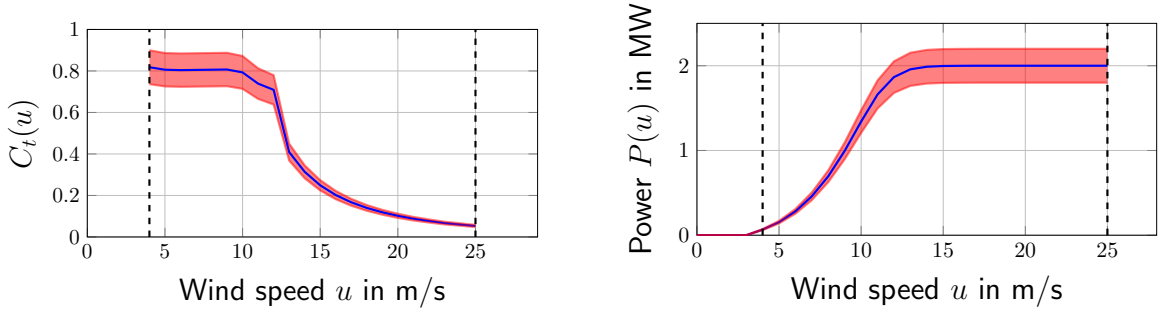

Figure 10: Perturbed C_t and power curve of of the turbine Vestas V80 with cut-in speed of 4 m/s and cut-out speed of 25 m/s (dashed vertical lines). The red tube illustrates the C_t and power curve with a perturbation of $\pm 10\%$.

4.2.5 Uncertain gross Annual Energy Production (AEP)

The uncertain gross AEP depends on the probability density function (12) for the wind distribution, on the uncertain velocity deficit function (13) and on the uncertain power function (14) so that formula (4) for the gross AEP changes as follows:

$$\begin{aligned}\widetilde{E}_{\text{AEPgross}} &= (8760h + 6h) \cdot P \\ &\approx (8766h) \cdot \sum_{i=1}^{N_{\text{directions}}} w_{\alpha_i} \cdot \sum_{j=0}^{N_{\text{speeds}}} w_j \cdot \widetilde{f}_{\alpha_i}(u_j, \xi_{\text{wind}}) \\ &\cdot \sum_{k=1}^{N_{\text{turbines}}} \widetilde{P}_{\alpha_i}(u_{\text{inc}_{j,k}, \xi_{\text{power}}, \xi_{\text{ct}}, \xi_{\text{rough}}, \xi_{\text{wake}}})\end{aligned} \tag{15}$$

4.2.6 Uncertain net Annual Energy Production (AEP)

The net AEP depends from the availability losses p_{loss} which consider the grid and turbine downtime. Because these influences highly depend on external circumstances this parameter should be modeled stochastic. Thus, we perturb the availability losses from (5) with the normally distributed random variable ξ_{loss}, which results into the following formula:

$$\widetilde{E}_{\text{AEPnet}} = \widetilde{E}_{\text{AEPgross}} \cdot (1 - p_{\text{loss}} \cdot \xi_{\text{loss}}). \tag{16}$$

Note, that the perturbation of the random variable ξ_{loss} is chosen with 2.18 % as the same value as the availability losses p_{loss}, compare Table 5 and 1. This means, that for a deviation of 3σ the losses vanish and thus the net annual energy production equals to the gross annual energy production. Of course we need to ensure, that a value of higher than 3σ can never be reached. Thus, finally ξ_{loss} is modeled as a *truncated* normally distributed random variable because it should not be possible to calculate a higher AEP with a uncertain power loss than it would be without considering the power loss.

4.2.7 Uncertain Levelized Cost of Electricity

The capital costs C_{capital} mainly depend on the price of steel. But because of the long planning stage of several years for a wind farm the calculation depends on long-term predictions for the steel price which is very volatile. Therefore we perturb the capital costs with a normally distributed random variable ξ_{capital}. The same argument holds for the discount rate r_{rate}, which is in an early planning stage very unsure such that this parameter is perturbed with a normally distributed random variable ξ_{rate}. The costs for annual operation and maintenance $C_{\text{o\&m}}$ also underlie the volatile behavior of the price of steel (for the material), and other political decisions like payroll taxes. Therefore also this parameter is perturbed with a normally distributed random variable $\xi_{\text{o\&m}}$. Altogether, the the levelized cost of electricity from (6) changes as follows:

$$\widetilde{K}_{\text{LCOE}} = \frac{C_{\text{capital}} \cdot \xi_{\text{capital}} \cdot \dfrac{(1+\tilde{r}_{\text{rate}})^{\ell} \cdot \tilde{r}_{\text{rate}}}{(1+\tilde{r}_{\text{rate}})^{\ell} - 1} + C_{\text{o\&m}} \cdot \xi_{\text{o\&m}}}{\widetilde{E}_{\text{AEPnet}}}, \tag{17}$$

with perturbed discount rate $\tilde{r}_{\text{rate}} = r_{\text{rate}} \cdot \xi_{\text{rate}}$.

4.2.8 Uncertain Net Present Value

The net present value depends on the selling price of energy per kWh K_{energy} which is in many countries a value which is defined by political laws. In Germany, for example, this value changes every quarter of a year. In an early planning stage this parameter is very unsure such that it is perturbed with a normally distributed random variable ξ_{energy}. This changes the calculation of the net present value from (7) as follows:

$$\widetilde{C}_{\text{NPV}} = C_{\text{PRV}} - C_{\text{capital}} \cdot \xi_{\text{capital}} + \sum_{t=1}^{\ell} \frac{\widetilde{E}_{\text{AEPnet}} \cdot K_{\text{energy}} \cdot \xi_{\text{energy}} - C_{\text{o\&m}} \cdot \xi_{\text{o\&m}}}{(1+\tilde{r}_{\text{rate}})^{t}}. \tag{18}$$

4.2.9 Uncertain Internal Rate of Return

The perturbed internal rate of return $\tilde{r}_{\text{IRR}}$ is the discount rate r which makes the uncertain net present value (18) equal to zero. This value should be greater than the discount rate r_{rate} plus risk deficits for a profitable project.

5 Uncertainty Quantification (UQ)

In this section furthermore some techniques from the field of UQ are presented. The random variable generation is explained which is important for the sampling methods in Section 5.4.
Uncertainty quantification is the combination of probability theory and statistical practice with the 'real world'. It involves one or more mathematical models which represents scenarios from the real world and investigates uncertainties of the models itself or model inputs for those models. It is a method to delimit risks beforehand in order to prevent losses because of miscalculations in e.g. financial planning of a project [8].
Like mentioned before there are different sources of uncertainties. One source are the model errors or discrepancies itself which occur because of the approximation of physical, biological, economic or social processes. Components of those models are simplified such that the power of representation of those models are just representative up to a certain degree. Another source of risks are the input uncertainties. These uncertainties occur when a model has nonphysical parameters whose values must be determined using measured data where measuring inaccuracy can appear. Numerical errors are also a source of uncertainties which arise in computer algorithms using numerical approximations. They include round off, discretization or approximation errors, bugs or coding errors and other unpredictable scenarios [21].
In this thesis the input parameters of our offshore model will be considered and examined to make an uncertainty quantification on our quantity of interest. So we will perturb the input parameters used in our offshore model to make statements about risks in planning an offshore wind farm.

5.1 Fundamentals of Probability and Statistics

Before we come to the operative part of uncertainty quantification we have to define some probability and statistic basics.

5.1.1 Probability Basics

Probability Space A probability space (Ω, Σ, P) is made of three components:

Ω : sample space is the set of all possible results from an experiment

Σ : σ-field of subset of Ω that contains all events of interest

P : $\Sigma \to [0,1]$: measure that gives the events probability values

(i) $P(\emptyset) = 0$

(ii) $P(\Omega) = 1$

(iii) if $A_i \in \Sigma$ and $A_i \cup A_j = \emptyset$ and $i \neq j$, then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

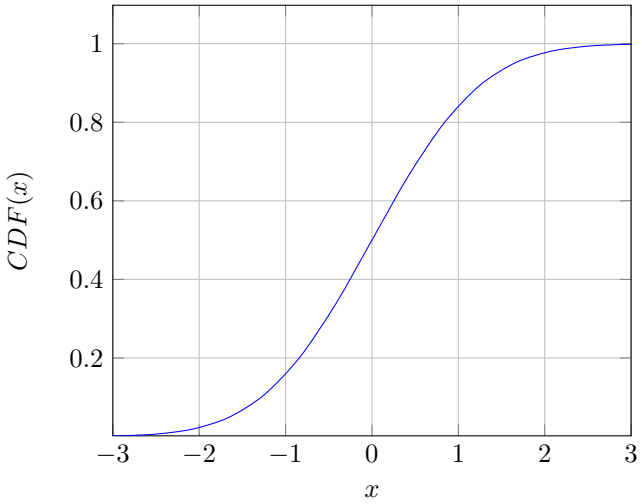

Figure 11: Cumulative distribution function of normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$.

Random Variable A random variable X is a function $X : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega | X(\omega) \leq x\} \in \Sigma$ for each $x \in \mathbb{R}$. The value $x = X(\omega)$, for an event $\omega \in \Omega$ is called realization of X.

Cumulative Distribution Function The cumulative distribution function (see Figure 11) $F_X : \mathbb{R} \to [0, 1]$ describes the probability of a random variable X given by

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega | X(\omega) \leq x\}). \tag{19}$$

Probability Density Function The probability density function is the distribution of a random variable X (see Figure 12). If the cumulative distribution function of X is continuous it can be expressed as

$$F_X(x) = \int_{-\infty}^{\infty} f_X(s)ds, \quad x \in \mathbb{R}, \tag{20}$$

where the derivative $f_X = \frac{dF_X}{dx}$, mapping $\mathbb{R}$ to $[0, \infty)$ is called the probability density function of X.

Law of Large Numbers and Central Limit Theorem The law of large numbers and central limit theorem are fundamental in the probability theory. The main idea of them is that for a large number of attempts of a random experiment, the average defined by

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{21}$$

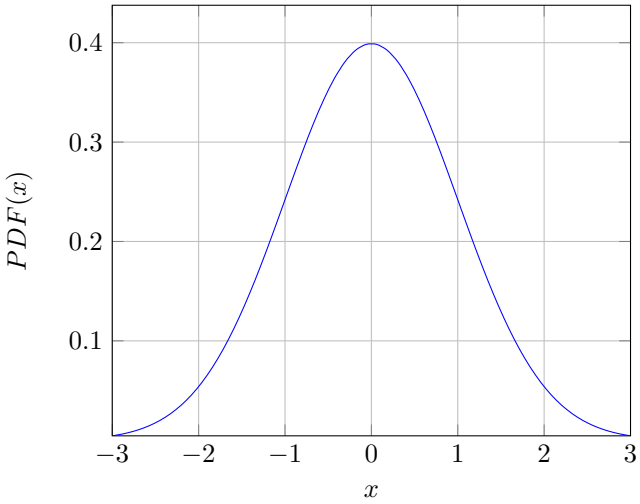

Figure 12: Probability density function of the normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$, also written as $\mathcal{N}(0, 1)$.

where X_i are random variables, converges to the mean μ. So it can be said that $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$, for sufficiently large n. For detailed explanation look up Smith [21, p.86].

5.1.2 Statistics for Uncertainty Quantification

To evaluate our results later on we have to use some statistics. A data set with n entries of a quantity of interest where every entry is represented as x_i for $1 \leq i \leq n$ can be evaluated with the following formulas which make statements about the data set:

Mean μ

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{22}$$

The mean μ describes the expected value of the data set.

Standard Deviation σ

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2} \tag{23}$$

The standard deviation is a measure for the spread of the entries of the data set. A big value indicates that most of the entries are far away from the mean μ.

Variance σ^2

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2 \tag{24}$$

The variance σ^2 is the square of the standard deviation σ and states different to the standard deviation not the wideness of the spread but the strength of the spread referred to the data set.

Minimum x_{min} and Maximum x_{max}

$$\begin{aligned} x_{min} &= \min\{x_i | 1 \leq i \leq n\} \\ x_{max} &= \max\{x_i | 1 \leq i \leq n\} \end{aligned} \tag{25}$$

The minimum x_{min} and maximum x_{max} represents the smallest and the biggest value of the data set.

Range R

$$R = x_{max} - x_{min} \tag{26}$$

The range R describes the distance between the smallest and biggest value of the data set.
All of these statistical measures are referred from [18].

5.1.3 Boxplot

A boxplot is a graphical representation of data which visualize statistical characteristics of a considered data set. In Figure 13 a boxplot can be seen where statistical characteristics are shown. The boxplot is made of a box, two antennas or arms and points outside of the antennas. The box's width is determined by the lower and upper quartile and stands for the inter quartile range (IQR) where 50% of the data in the data set are located. The lower or 0.25 quartile $q_{0.25}$ indicates the bound where 25% of the data is below or equal to the quartile value and the upper or 0.75 quartile $q_{0.75}$ indicates the bound where 75% of the data is equal or over to the quartile value. The vertical line inside the box stands for the median of the data set which means that 50% of the data is smaller or equal to this value. The fact that 50% of the data is located inside the box means that the rest of the data is located outside the box. The range of the data is represented with the limitation of the lower and upper whisker w_{lower} and w_{upper} or if outliers are shown between the first and last outlier. The bounds w_{lower} and w_{upper} are calculated as follows:

$$\begin{aligned} w_{\text{lower}} &= q_{0.25} - 1.5 \cdot \text{IQR} \\ w_{\text{upper}} &= q_{0.75} + 1.5 \cdot \text{IQR} \end{aligned} \tag{27}$$

With this representation the distribution of the data set is nicely visualized and can be easily compared with other distributions on the same plot.

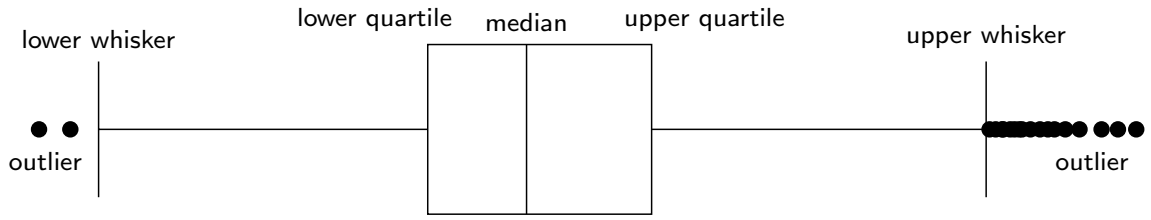

Figure 13: Boxplot representation with outliers, upper and lower whisker, upper and lower quartile and median.

5.2 Input Parameters

An investigation of a deterministic model, which does not consider variations of the input parameters, is often termed base case scenario, since the values of these input parameters are assumed or measured values. An effective model should consider risks in these input parameters. In the most cases scientists investigate the base case, the best case, and the worst case scenario for the values of the input parameters. But this approach has some disadvantages, e.g. it can be difficult to find the best and worst case scenarios for each input parameter. Also the input parameters may not be at their best or worst levels at the same time which also can lead to unpredictable outputs. One have to increase the number of possible cases, but it is not practical to go through all possible values of each input parameter [18]. So sample methods will be used to manage this (see Section 5.4) by filling up the probability space to cover as much as possible cases to see which influence the uncertain model inputs have regarding to the quantities of interest.

5.3 Inverse Transform Method (ITM)

To perturb the input parameter of our offshore model we need random numbers to change the input according to a specified distribution. The random number generation is usually split into two stages:

1. Generation of uniform distributed random variables in the interval of $[0, 1]$.
2. Conversation of uniform distributed random variables into random variables of a specific probability distribution.

The inverse transform method can use independent and identically distributed random variables to generate random variables of any other desired distribution. To generate a random variable X with a cumulative distribution function F the inverse of F have to be evaluated with a random variable Y of the uniform distribution $\mathcal{U}(0, 1)$[14]. The generalized inverse of F, F^{-1} is defined by:

$$X = F^{-1}(Y) := \min\{x | F(x) \geq Y\}. \tag{28}$$

Proof.

$$F(x) = P(X \leq x) = P(F^{-1}(Y) \leq x) = P(Y \leq F(x)) = F(x)$$

□

In Figure 18, 19 and 20 the inverse transform method from independent and identically distributed random variables to the normally distributed random variables is shown.

5.4 Sampling Methods

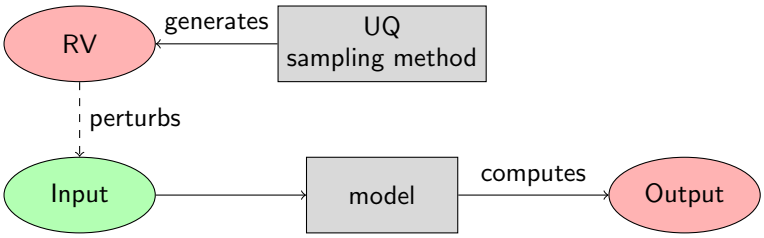

Figure 14: One sample step of uncertainty quantification with mathematical model.

To make an uncertainty quantification on the input parameters of the PARK wake model we have to perturb them like mentioned in Section 5.3. Then we use the perturbed input parameters in our simulation model to get an individual outcome for every perturbation.
With the use of the law of large numbers and the central limit theorem we can then predict the impact of each input parameter on the quantities of interest. Therefore it is necessary to use sampling methods which generate random variables from probability distributions which perturb the input parameters in every sample step according to the random variable which was generated as described in Figure 14. Every input parameter has its own specific distribution, for example the normal distribution (see Figure 12). The size of samples highly affect the accuracy of the results so that with a larger number of samples more accurate results will be computed.
The sampling methods which will be presented in the following are the Monte Carlo sampling method, the Quasi Monte Carlo sampling method and the Latin Hypercube sampling method. Each method have their advantages and disadvantages according to the generation of random numbers.

5.4.1 Monte Carlo Sampling

The Monte Carlo sampling method is based on repeated random sampling and statistical analysis of the results [18]. This sampling method generates the random variables with the help of a random number generator. These random values then will be transformed to random variables of a probability distribution. By perturbing the input parameters in every sample step according to their specific distributions we get, after

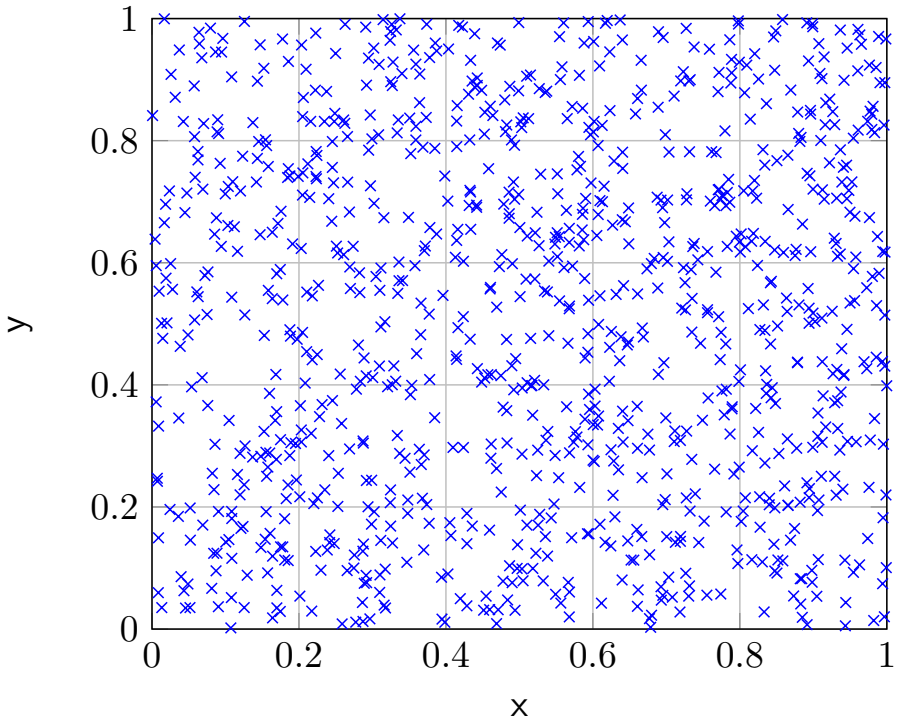

Figure 15: 2D - C++ random number generator (Mersenne Twister Engine) with 1000 sample points.

a sufficient number of samples, a resulting distribution which can be analyzed (see Section 5.1.2). If one input parameter is considered it is easy to fill the whole probability space of the distribution. But for the multidimensional case it gets more difficult to fill the whole multidimensional room as illustrated in Figure 15. In the Figure are 1000 random sample points and there are still areas which are not considered. The samples are not distributed very evenly and therefore an even larger number of samples is required to also consider the empty spaces. Since Monte Carlo sampling method is very simple there are improved sampling methods which try to reduce the size of samples.

5.4.2 Quasi-Monte Carlo Sampling

One of the improved sampling methods is the quasi-Monte Carlo sampling method. It relies on the foundation of the Monte Carlo sampling method which generates the random variables with the help of a low discrepancy sequences. These sequence values then will be transformed to random variables of a probability distribution e.g. with the inverse transform method described in section 5.3. A random number generator, as for the traditional Monte Carlo method, produces independent and identically distributed random numbers. In Figure 18(a) a normalized histogram of the generated numbers shows a significant discrepancy between the generated numbers and underlines that the generated numbers are not distributed very evenly.

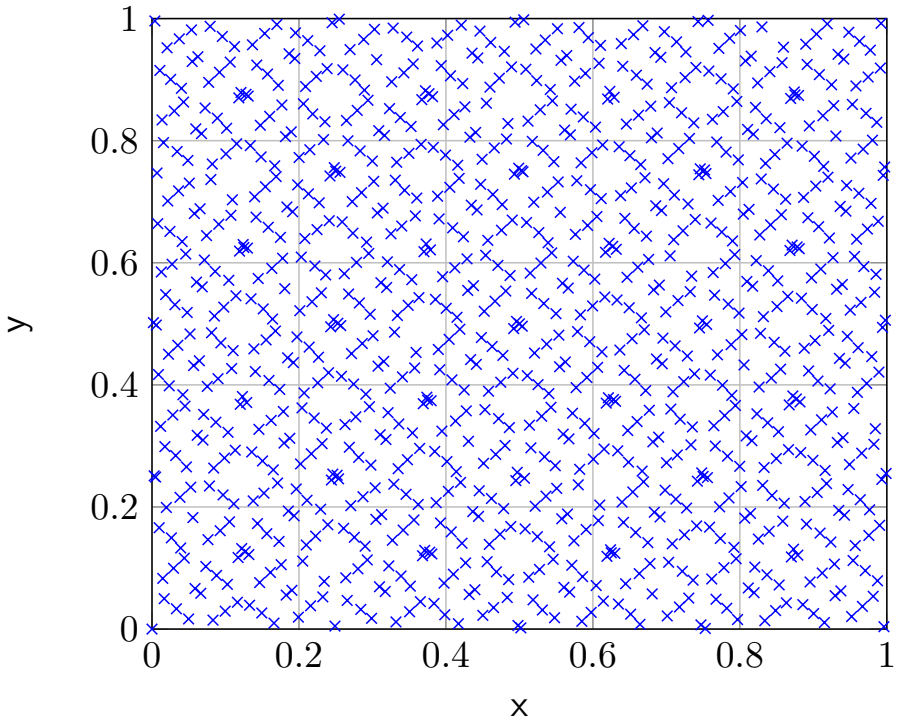

Figure 16: 2D - Sobol sequence with 1000 sample points.

step	0	1	2	3	4	5	...
value	0	0.5	0.75	0.25	0.875	0.375	...

Table 2: 1D - Sobol sequence number generation.

The aim of low discrepancy sequences for the quasi-Monte Carlo method is to reduce the discrepancy of the generated numbers to reduce in this way the size of samples. Morokoff and Caflisch [15] have investigated different approaches of low discrepancy sequences which can be used as the needed numbers for the transformation into the specific random variables.
These low discrepancy sequences have the characteristic that they are an infinite sequence of values such that for every number of generated values the sequence stays uniformly distributed. Morokoff and Caflisch examined three different low discrepancy sequences. The Halton, Sobol and Faure sequence. The result was that Halton sequences are the best for up to around six dimensions and for higher dimensions the Sobol sequence was better. The Faure sequence was for the most part of validation outperformed by the other two techniques. We want to consider the Sobol sequence in this thesis with the reason that we want to be able to perturb more than six input parameters.
The Sobol sequence can briefly be explained as an infinite sequence of uniformly distributed values in $[0, 1)$. The sequence generates new values by filling the empty spaces

as shown e.g. for the one dimensional case in Table 2 or for the two dimensional case in Figure 16. It can be seen that the 1000 points are spread more evenly and with a certain scheme, so that there are less and smaller areas that are not filled compared to the traditional Monte Carlo method. For a detailed explanation of the generation of the sequence see Bratley and Fox [1].

5.4.3 Latin Hypercube Sampling

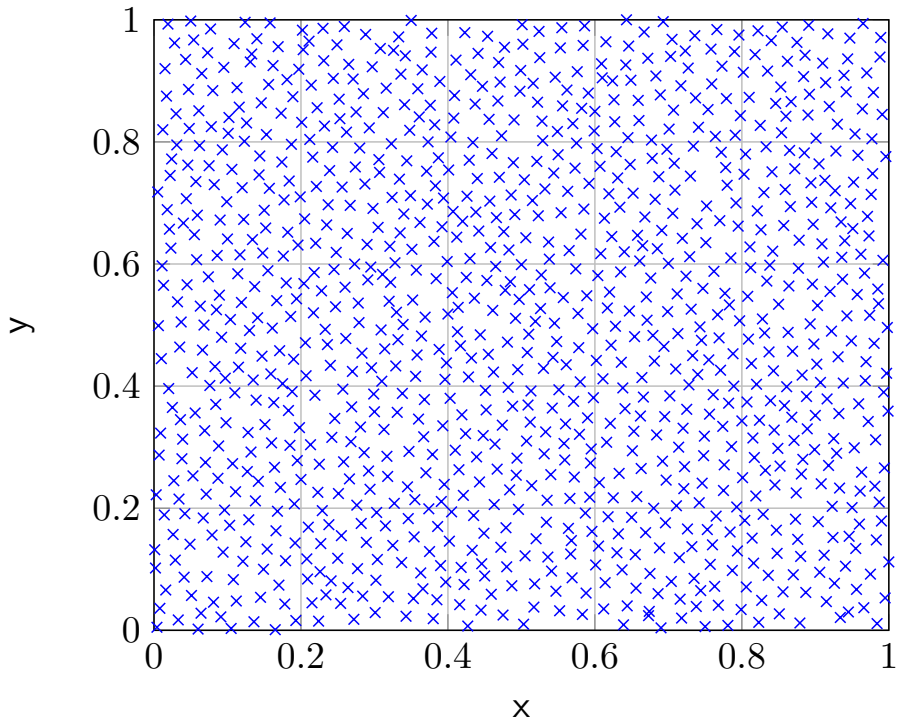

Figure 17: 2D - Latin Hypercube sampling with 1000 sample points.

The Latin Hypercube sampling method is also an improved sampling method with the same approach as the quasi-Monte Carlo sampling method. The aim here is also to have a low discrepancy between the values generated by random number generator. The difference to the quasi-Monte Carlo sampling method is that the Latin Hypercube sampling method has to know the number of samples beforehand and calculates the generated numbers based on the sample size. This has the advantage, that the generated numbers can be more evenly distributed than with the Sobol sequence as illustrated in Figure 17. It has to be mentioned that the Latin Hypercube sampling method has many implementations. In this thesis the Improved Distributed Hypercube sampling is used because of it's evenly distributed numbers in even more than one dimension. But there are also disadvantages because the numbers for the Latin Hypercube sampling method have to be computed beforehand and have to be stored which increases the the computational time before it comes to the actual simulation.

	seconds	
Generated numbers	Latin Hypercube sampling	Sobol sequence
8	0.000 012	0.000 001
16	0.000 050	0.000 001
32	0.000 309	0.000 001
64	0.002 102	0.000 002
128	0.015 182	0.000 005
256	0.115 596	0.000 012
512	0.893 315	0.000 022
1024	7.039 090	0.000 042
2048	44.882 800	0.000 090
4096	265.766 000	0.000 157
8192	2441.780 000	0.000 307

Table 3: Duration in seconds in dependency of generated 2-D Latin Hypercube sampling numbers and 2-D Sobol sequence numbers. For comparison: One calculation step of the annual energy production (AEP) averaged needs 0.726141 seconds but because of parallel computation the calculation depends linear on the number of threads. For example the calculation of the AEP with four threads effectively needs 0.181535 seconds. The used CPU is a 1,8 GHz Intel Core i5 of the second generation.

For example in Table 3 the computation time for the two dimensional case compared to the Sobol sequence numbers is shown. The duration of two dimensional Latin Hypercube sampling method numbers is quadratic compared to a almost linear computation time of two dimensional Sobol sequence numbers. In [17] the Improved Distributed Hypercube sampling is presented and it is said that the algorithm requires an array of the dimension $D \cdot N \cdot N$ with N for the number of Latin Hypercube sampling method numbers and D for the duplication factor which is set to the value of one in our implementation. Moreover, the program is inherently quadratic in N in the computation time. The computation of the i-th point requires a consideration of the value of the coordinates used up by the previous points, and the distances from each of those points to the candidates for the next point. In Section 6.3 the three sample methods will be compared with regard to the random variable generation.

5.5 Random Variable Generation

To generate the random variables for the UQ part it is necessary to use random numbers as discribed in Section 5.3. As the offshore model's simulation program is programmed in C++ there are standard implementations for that. C++ provides random number generators and also methods and classes to use them for probability distributions. They has been used for the Monte Carlo sampling method as it uses random numbers. The use of truly random numbers in a computer program is impossible because of its precise and deterministic behavior. Random numbers generated by computers are

called pseudo random numbers which are computer generated sequences that are imitating the randomness in a acceptable and not bad way [16]. They are implemented in a way that the outcome is simulating independent and identically distributed random numbers (see Figure 18(a)). To generate random variables of a specific distribution the standard implementation of C++ uses either the inverse transform method or other methods as the Box Muller method or the rejection method. In the case of normal distributed random variables C++ uses by default the Box Muller method where two uniform distributed random variables will be transformed to two standard normal distributed random variables which is described in detail in [16].
The quasi-Monte Carlo sampling method and Latin Hypercube sampling method as described in Section 5.4 have to use quasi random numbers. These numbers are characterized by their truly even distributed numbers illustrated in Figure 19(a) and in Figure 20(a). They have a very low discrepancy according to their spacing.
In my investigation I found out that the random variable generation for normal distributed random variables (which are in our case the most important random variables for UQ) with the use of quasi random numbers with the Box Muller and the rejection method leads to the damage of the low discrepancy of the random variables and the result was that is was not better as with the use of pseudo random numbers (see Figure 18(b)). The solution for this problem is the use of the inverse transform method. But out of that another problem occurred because the inverse transform method needs the inverse of the cumulative distribution function of the normal distribution (see Section 5.3). The problem here is that there is no analytically determinable inverse function F^{-1} for the cumulative distribution function of the normal distribution. The solution was provided by the inverse error function $\texttt{erf}^{-1}(x)$. Hence that the cumulative distribution function F of the normal distribution and the error function $\texttt{erf}(x)$ have a similarity, F of the standard normal distribution can be represented as

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} \mathrm{d}t = \frac{1}{2} \left(1 + \texttt{erf} \left(\frac{x}{\sqrt{2}} \right) \right) \tag{29}$$

that yields

$$F^{-1}(x) = \sqrt{2}\ \texttt{erf}^{-1}(2x - 1). \tag{30}$$

If x is an independent and identically distributed random variable then $y = F^{-1}(x)$ is a standard normal distributed random variable, with zero mean and unit variance [4]. To get a normal distributed random variable $\hat{y}$ with different mean μ and variance σ^2 the standard normal distributed random variable y has to be used as follows:

$$\hat{y} = \mu + y \cdot \sigma. \tag{31}$$

In Figure 19(b) for the quasi-Monte Carlo sampling method with Sobol sequences or in Figure 20(b) for the Latin Hypercube sampling method can be seen that the solution is correct as the $\texttt{erf}^{-1}(x)$ preserves the beneficial properties of the low-discrepancy sequences which is explained in detail in [5].

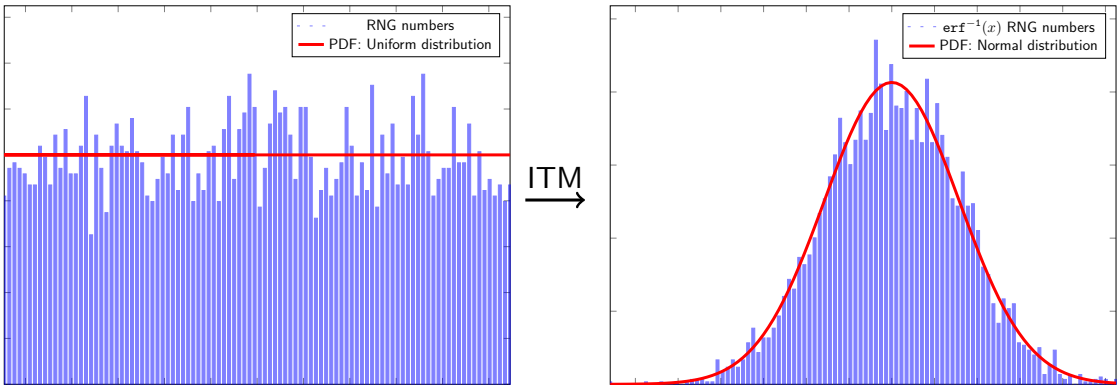

Figure 18: Inverse Transform Method with C++ random number generator. In Figure (a) relation between probability density function of Uniform distribution and frequencies of C++ random number generator numbers. In Figure (b) result after transforming with $\texttt{erf}^{-1}(x)$ compared to probability density function of normal distribution with 4096 samples.

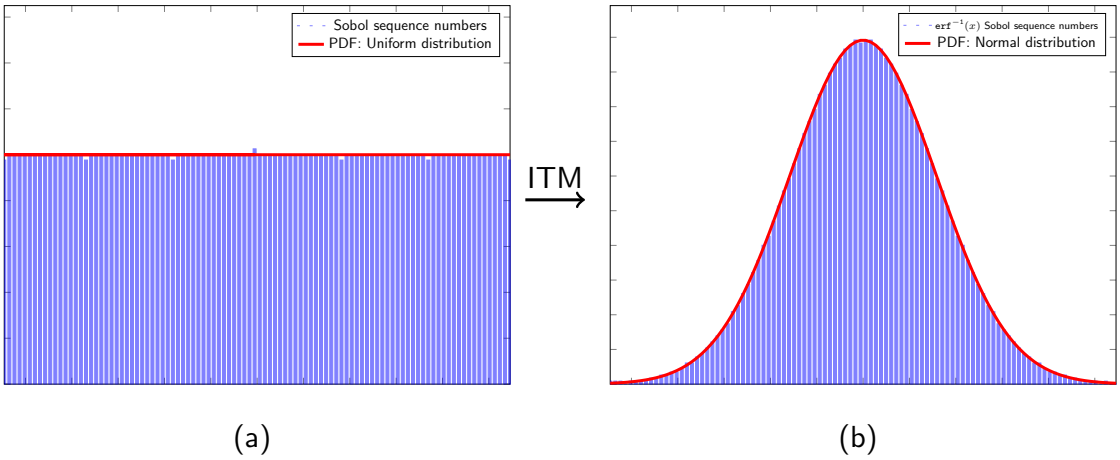

Figure 19: Inverse Transform Method with Sobol Sequence. In Figure (a) relation between probability density function of Uniform distribution and frequencies of Sobol Sequence numbers. In Figure (b) result after transforming with $\texttt{erf}^{-1}(x)$ compared to probability density function of normal distribution with 4096 samples.

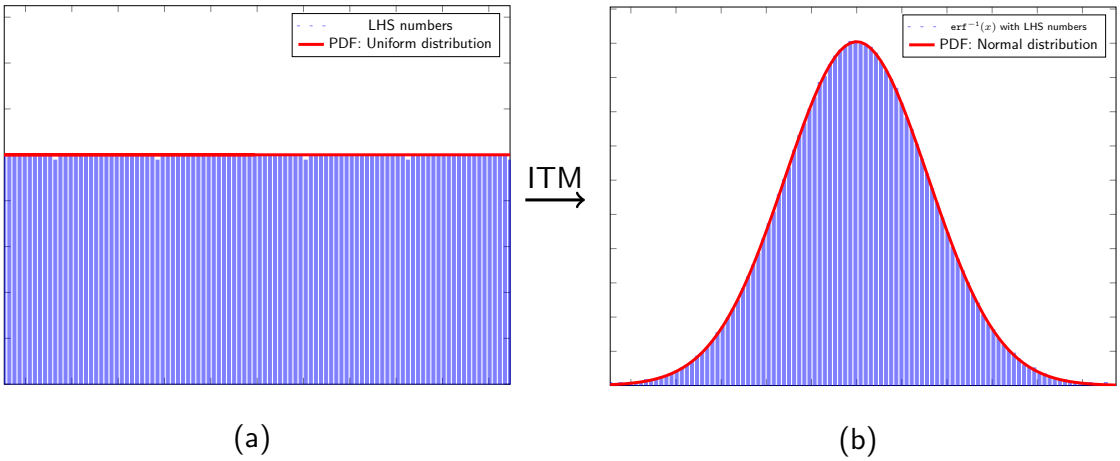

Figure 20: Inverse Transform Method with Latin Hypercube Sampling. In Figure (a) relation between probability density function of Uniform distribution and frequencies of Latin Hypercube Sampling numbers. In Figure (b) result after transforming with $\mathtt{erf}^{-1}(x)$ compared to probability density function of normal distribution with 4096 samples.

6 Implementation

In this section a brief description about the implementation is presented which includes basic information about the code and the program structure. The implementation is based on the simulation program of Heiming [7] and Wolters [25].
The code is implemented in C++ and uses JSON files as the input for the variable properties of the simulation program. The JSON input files can be generated with standalone programs which convert raw data into JSON files. The input files are separated into a turbine file which contains the needed turbine parameter of the wind farm, a position file which provides all positions of the respective turbines, a wind data file which describes the distribution of the wind speed in the form of the weibull distribution for all wind directions and the probability for the specific wind direction, a file which sets up the distribution for the uncertain parameters and which distribution for each parameter should be used and a settings file for site and program settings.

6.1 Program Structure

In Figure 21 the program structure of the offshore model simulation is illustrated. The simulation program is made of four (or five including the optimizer class) main classes:

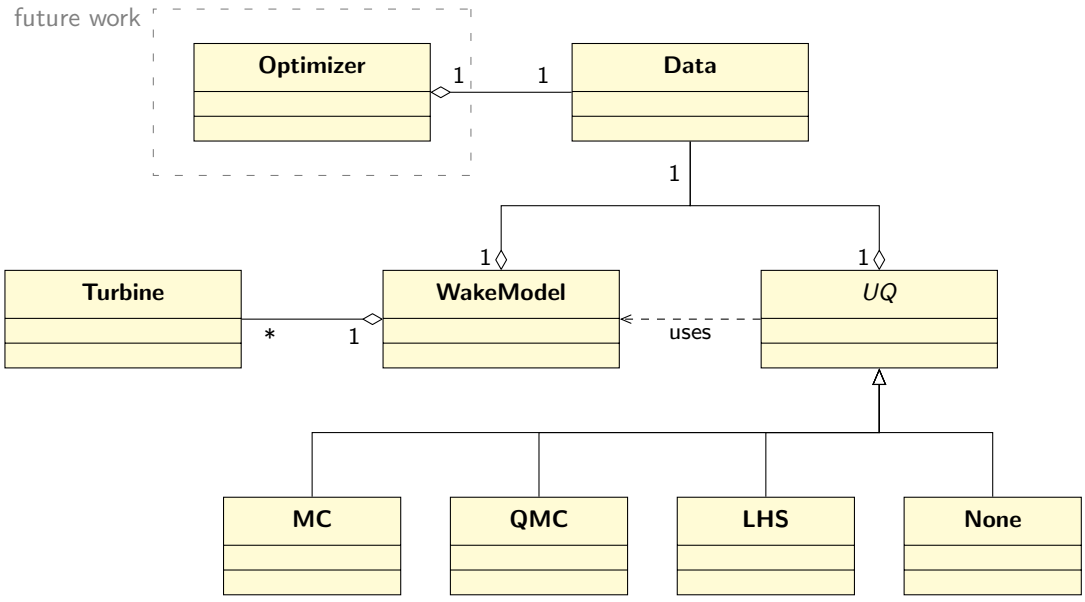

Figure 21: UML class diagram

- The Data class reads the input parameters from the JSON files and provides them for the other classes. The data is accessible with a data pointer for the other classes so that the input parameters are not take up so much memory with redundant data.

- The Turbine class is providing the attributes and curves of the turbine.
- The WakeModel class is using the referenced Data object to generate the Turbine objects which are used for the calculation of the quantities of interest which are also implemented in this class.
- The uncertainty quantification (UQ) class is an abstract class which provides the fundamental attributes and methods for the sample methods. It also has a reference to the Data object. In the class diagram can be seen that the UQ class has an association to the WakeModel class. The reason for that is that the sample method is creating in each sample iteration a WakeModel object initialized with a perturbed Data object, which was copied before from the referenced Data object per copy constructor, to calculate the result of the selected quantity of interest. After that the WakeModel and copied Data object will be deleted and in the next iteration this process will be repeated.
- Hence this work is a preparation for the last step of the wind farm layout optimization problem: An Optimizer class has to be added which perturbs the positions of the turbines to find the optimal wind farm layout in relation to the selected quantity of interest.

6.2 Accuracy of gross annual Energy Production (AEP) calculation

In Figure 22 the impact on the accuracy of the number of speed classes N_{Speeds} are illustrated. The more speed classes will be used for the numerical integration of the wind speed distribution $f_{\alpha_i}(u)$ the higher the accuracy for the gross AEP calculation becomes. We consider the gross AEP because it is used by all quantities of interest of the cost model. The accuracy of the integration with 16 quadrature points leads to a accuracy of at least 99%. For the integration of the wind speed's weibull distribution we use the trapezoidal quadrature rule which generally can be formulated as follows:

$$\int_a^b f(x)dx \approx \frac{b-a}{2N} \sum_{k=1}^{N} \left(f(x_{k+1}) + f(x_k)\right)$$

In Figure 23 we can see the accuracy regarding to the calculation of the AEP in dependency to the number of wind sectors. It can be seen that with more wind direction sectors more precise results can be achieved. The number of wind sectors can have a big influence to the computation time because the integration of the wind direction is one of the outer integration loops in the simulation.

6.3 Precision of Inverse Transform Method

The precision of the generation of normal distributed random variables with the inverse transform method (see Section 5.3) depends on the size of samples and can be read off in Table 4 and is illustrated in Figure 24. The precision is measured with the root-mean-squared-error (RMSE) which is a statistical measure for the difference of measured and

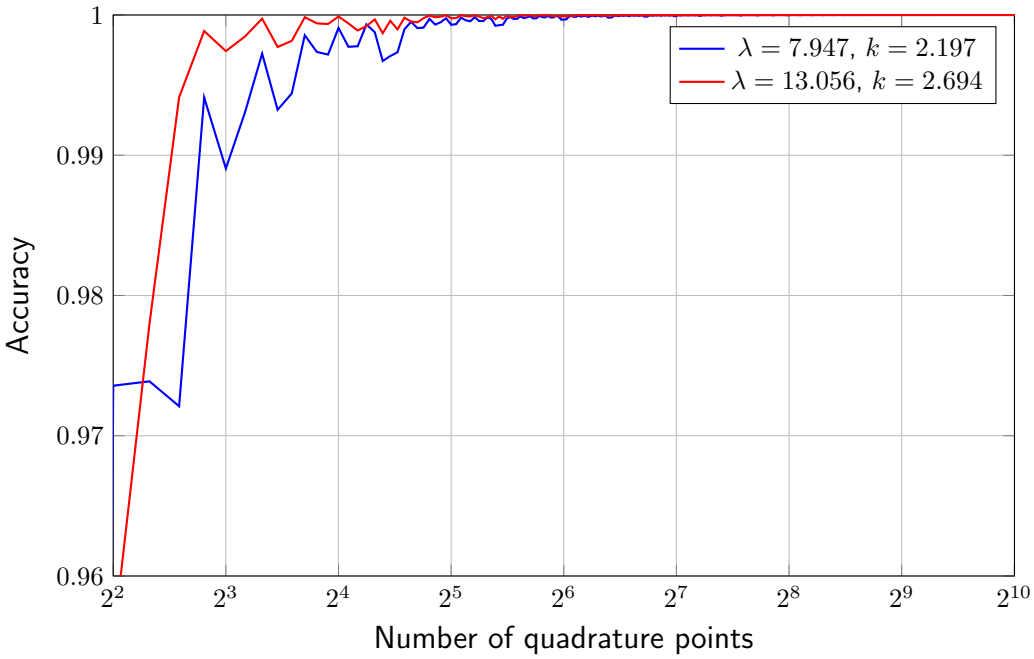

Figure 22: Accuracy of annual energy production calculation with the integration of two different weibull distributions (determined from the wind data of the FINO3 research platform) with one turbine and one direction with the trapezoidal quadrature rule with increasing number of quadrature points. The reference value of the AEP is the integration with $2^{17} = 131072$ quadrature points.

reference values. In our case the measured values are the relative frequencies of the generated random variables compared to the values of the probability density function of the normal distribution. The root-mean-squared-error is given as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{\text{reference},i} - x_{\text{mesured},i})^2} \tag{32}$$

It can be seen in Table 4 that the random variable generation with the Sobol sequence and the Latin Hypercube sampling are very similar. Moreover these methods have a better precision with 512 samples regarding to the generation of normal distributed random variables than pseudo random numbers with the random number generator of C++ with 8192 samples. It also can be seen that the convergence rate with the random number generator leads to a convergence of about $c \cdot \frac{1}{\sqrt{N}}$ with $c \approx 7$ and the convergence rate of the other two methods leads to $\hat{c} \cdot \frac{1}{N}$ with $\hat{c} \approx 30$. That means the standard Monte Carlo method has a convergence rate of $\mathcal{O}(\frac{1}{\sqrt{N}})$ whereas the quasi-Monte Carlo method and the Latin Hypercube method have a rate of $\mathcal{O}(\frac{1}{N})$.

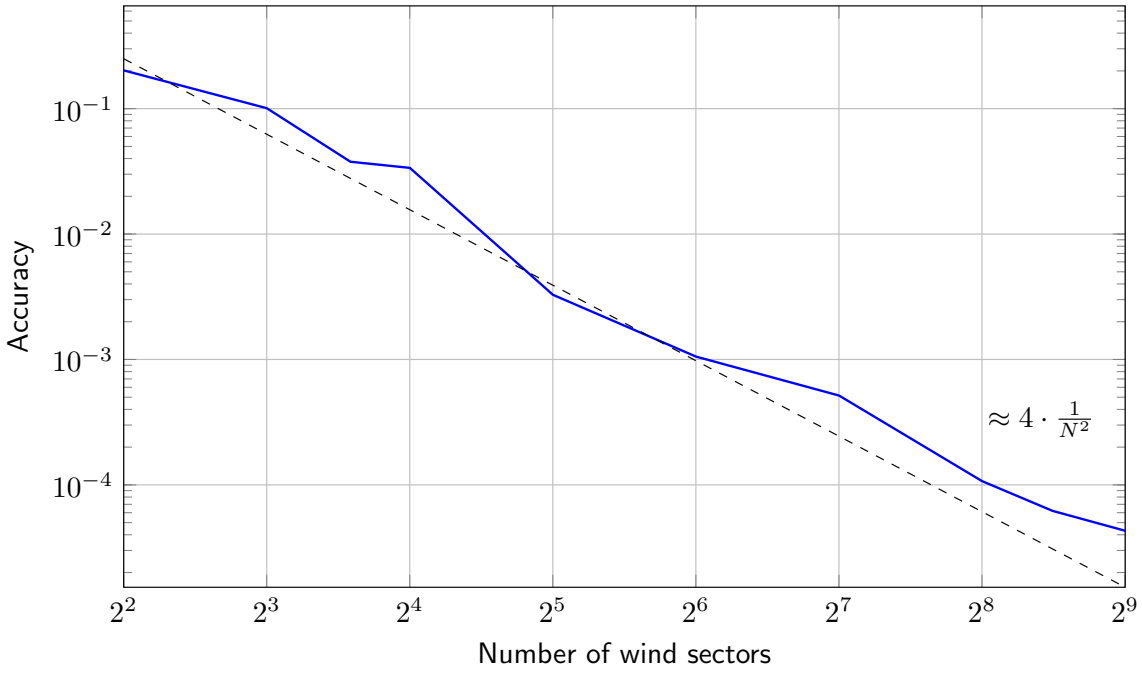

Figure 23: Accuracy of gross annual energy production calculation with increasing number of wind sectors. Reference value is annual energy production calculation with 4096 wind sectors and 64 quadrature points of the wind speed distribution.

Samples	Sobol sequence numbers	Random number generator numbers	Latin Hypercube sampling numbers
128	0.285738	0.646411	0.277768
256	0.128956	0.465927	0.127106
512	0.065717	0.363601	0.0611908
1024	0.0307469	0.237796	0.0299242
2048	0.0152109	0.141674	0.0157462
4096	0.00679491	0.0983468	0.00610846
8192	0.00347264	0.0729508	0.00335724

Table 4: Precision measured with the root-mean-squared error in dependency of sample size.

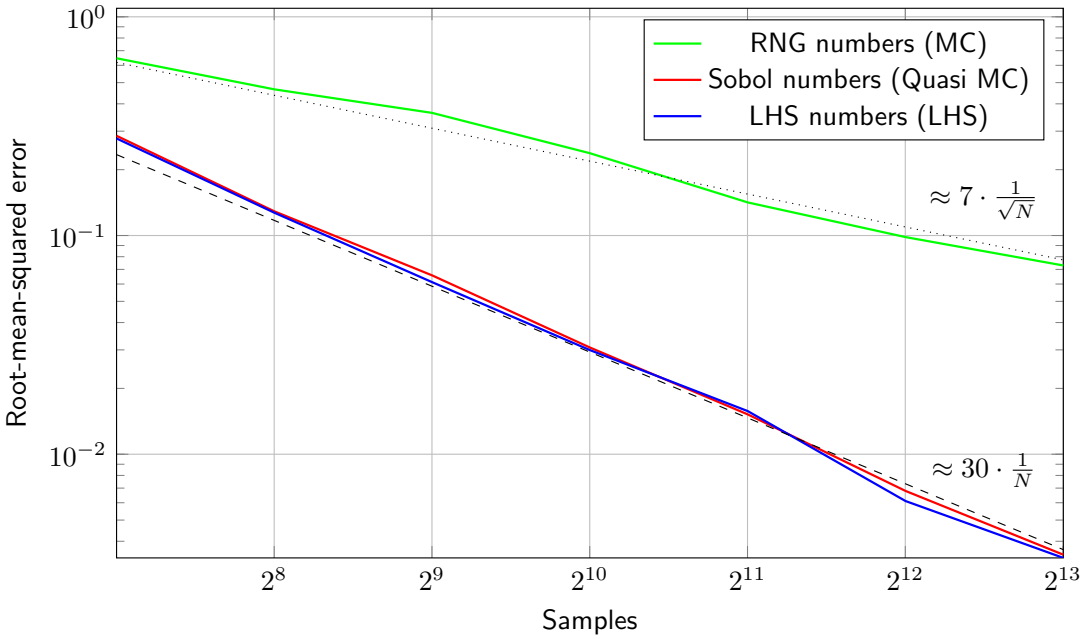

Figure 24: Precision of inverse transform method according to normal distribution with pseudo-random and quasi-random numbers, measured with root-mean-squared error in dependency of sample size. The standard Monte Carlo method has a convergence rate of $\mathcal{O}(\frac{1}{\sqrt{N}})$ whereas the quasi-Monte Carlo method and the Latin Hypercube method have a rate of $\mathcal{O}(\frac{1}{N})$.

7 Results

In this section we will examine the sensitivity of our uncertain input parameters concerning our uncertain quantities of interest described in Section 4 using the techniques from Section 5. The results of all four outputs will be analyzed and the influence of the input variables on the computed net annual energy production, levelized cost of electricity, net present value and internal rate of return will be determined.
Before we can analyze the uncertainties, it is necessary to introduce some of the notation used by the variance based sensitivity analysis described by Saltelli et al. [19]. This notation is especially designed to express how the uncertainty in the model output can be linked to uncertainties in the inputs.
The previously derived stochastic model will in the following be resembled by Y, while the model inputs will be referred to as ξ_i. The uncertainty propagation of ξ_i through Y is best described in the context of this notation by the sensitivity measure:

$$S_i = \frac{\mathrm{var}_{\xi_i}(E_{\xi_{\sim i}}(\mathrm{Y}|\xi_i))}{\mathrm{var}(\mathrm{Y})}, \tag{33}$$

which is technically the first order sensitivity coefficient that measures e.g. the additive effect of ξ_i on the model output. S_i can also be interpreted in terms of expected reduction of variance. This interpretation allows an easier understanding of the factors involed in the computation of S_i:

- $\mathrm{var}(\mathrm{Y})$: variance of the output with all inputs modeled as random variables
- $\mathrm{var}_{\xi_i}(E_{\xi_{\sim i}}(\mathrm{Y}|\xi_i))$: expected reduction in variance that would be obtained if ξ_i could be fixed

7.1 Horns Rev 1

We consider the wind farm Horns Rev 1 with it's 80 Vestas V80 offshore wind turbines. The positioning of the turbines can be seen in Figure 25. The wind data is provided from the FINO3 research platform over the years of July 2010 to June 2017 and in Table 5 our used base case input parameters for the simulation can be seen. It has to be mentioned that the values in the upper section of the table are real properties of the Horns Rev 1 wind farm but the economical parameters in the second section of the table are roughly estimated because they contain sensitive information which we can not provide yet.
We perturbed the input parameters as presented in Table 1 in Section 4. That means that every base case input parameter will be perturbed every sample step by a normally distributed random variable with mean $\mu = 1$ and standard deviation $\sigma = 0.01\bar{6}$ to obtain a perturbation of $\pm 5\%$ for each uncertain parameter. The perturbation of $\pm 5\%$ may not be realistic in real world because every uncertainty may vary with its real perturbation, but since we do not have the real information of the perturbation values it makes sense to make the uncertainty quantification with the same perturbation value

for all input parameter to see a normalized influence on the model.
In the next subsections we will present the results and how significant the influence of each uncertainty is regarding to the quantities of interest. In the simulation we used the quasi-Monte Carlo sampling with 4096 samples, twelve wind direction sectors and 64 quadrature points for the integration of the wind speed distribution. For each quantity of interest we prepared a boxplot representation with the results and a table which is presenting the values of the variance based sensitivity analysis for each used uncertainty.

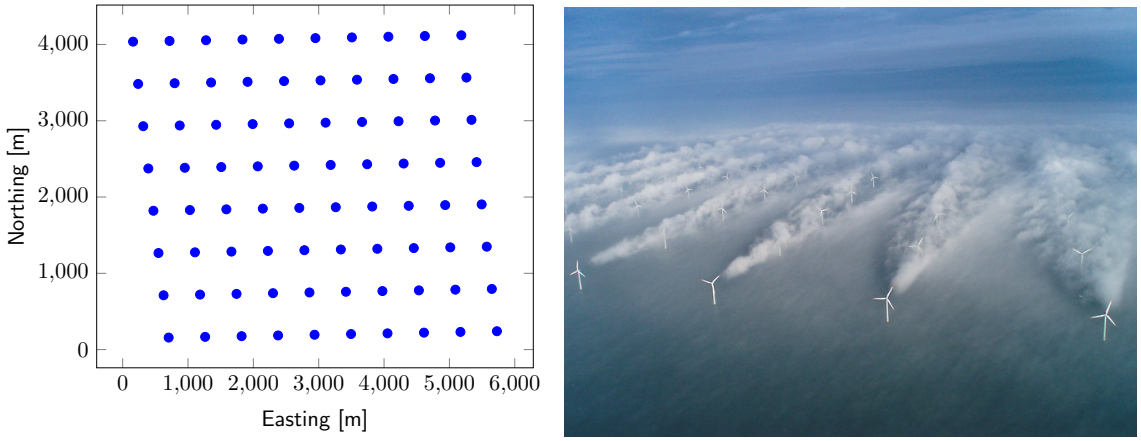

Figure 25: Turbine positions and wake effect of wind farm Horns Rev 1.

Parameter	**Value**
Rotor diameter D	80 m
Hub height z	70 m
Cut-in speed u_{cutin}	4 m/s
Cut-out speed u_{cutout}	25 m/s
Surface roughness z_0	$0.2 \cdot 10^{-3}$ m
Power curve $P(u)$	see Figure 7
C_t curve $C_t(u)$	see Figure 7
Positions	see Figure 25
Plant performance loss p_{loss}	7.00 %
Total capital costs C_{capital}	278 000 000 €
Annual operation and maintenance costs $C_{\text{O\&M}}$	24 000 000 €
Discount rate r_{rate}	2.75 %
Project lifetime ℓ	20 years
Present residual value C_{PRV}	0 €
Selling energy price K_{energy}	0.10 €/kWh

Table 5: Parameters for the Horns Rev wind farm.

7.1.1 Net Annual Energy Production (net AEP)

In Figure 26 the results regarding to the net annual energy production for all influencing uncertainty parameters, i.e. wind speed, wake effect, C_t curve, surface roughness, power curve and power loss are shown, since the other four uncertainties do not influence the net AEP calculation. The result for each uncertain parameter is shown in a separate boxplot. It can be seen that the wind speed has the largest impact on the net annual energy production in comparison to the other uncertainties. The topmost boxplot represents the model output for the case that the six uncertainties are perturbed at the same time. The result can be compared with the boxplots for the case of the perturbation of the single uncertainties. We can see that the simultaneous perturbation of the parameters leads to a bigger spread of the results comparing to the perturbation of a single parameter.

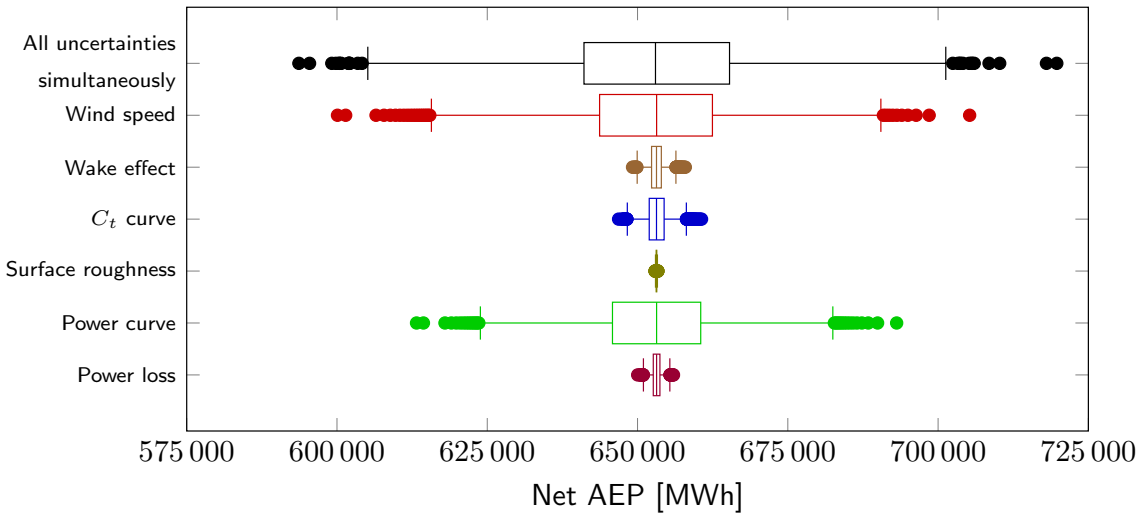

Figure 26: Test with six uncertainty parameters wind speed, wake effect, C_t curve, surface roughness, power curve and power loss at the same time (all uncertainties simultaneously) compared to single perturbation of the uncertainty parameters computed with the quasi-Monte Carlo method with 4096 samples.

Table 6 shows the results in terms of the variance based sensitivity analysis. As already seen in the boxplot, the results clearly indicate that ξ_{wind} has the biggest influence on the variance of the model output with a sensitivity coefficient of $S_{\text{wind}} = 61.15\%$. The second highest sensitivity is linked to the uncertain power curve with $S_{\text{power}} = 37.39\%$. All remaining coefficients are below 2% and thus have a rather negligible influence on the variance of the model output. The uncertain parameter ξ_{rough} representing the uncertain surface roughness has, at least with a perturbation of $\pm 5\%$, almost no influence on the variance of the output. Furthermore the uncertain wake effect ξ_{wake} only has a sensitivity of $S_{\text{wake}} = 00.48\%$ and the uncertain C_t curve has with the same perturbation a higher sensitivity with $S_{\text{ct}} = 01.19\%$ although the

C_t curve is used to calculate the wake effect.

$\mathrm{var}(\mathrm{Y}) = 3.1712 \cdot 10^8$		
Uncertainty	$\mathrm{var}_{\xi_i}(E_{\xi_{\sim i}}(\mathrm{Y}\|\xi_i))$	S_i
ξ_{wind}	$1.939 \cdot 10^8$	**61.15%**
ξ_{wake}	$1.5299 \cdot 10^6$	**00.48%**
ξ_{ct}	$3.7875 \cdot 10^6$	**01.19%**
ξ_{rough}	$1.5781 \cdot 10^4$	**00.01%**
ξ_{power}	$1.1856 \cdot 10^8$	**37.39%**
ξ_{loss}	$6.8761 \cdot 10^5$	**00.21%**

Table 6: Results of variances based sensitivity analysis of net annual energy production calculation with 4096 samples quasi-Monte Carlo Method.

7.1.2 Levelized Cost of Electricity

In Figure 27 the resulting boxplots for the levelized cost of electricity for all influencing uncertainty parameters, i.e. wind speed, wake effect, C_t curve, surface roughness, power curve, power loss capital costs, operation and maintenance costs and discount rate are shown, since the the energy price do not influence the levelized cost of electricity calculation. Also here can be seen that for the case that all uncertainties are perturbed simultaneously leads to the biggest spread and that for the single case again the wind speed has the biggest deviation. After that the power curve have the biggest influence on the result. As for this quantity of interest there are also three economical parameters involved, it can be seen that the annual operation and maintenance costs have the biggest sensitivity of them.

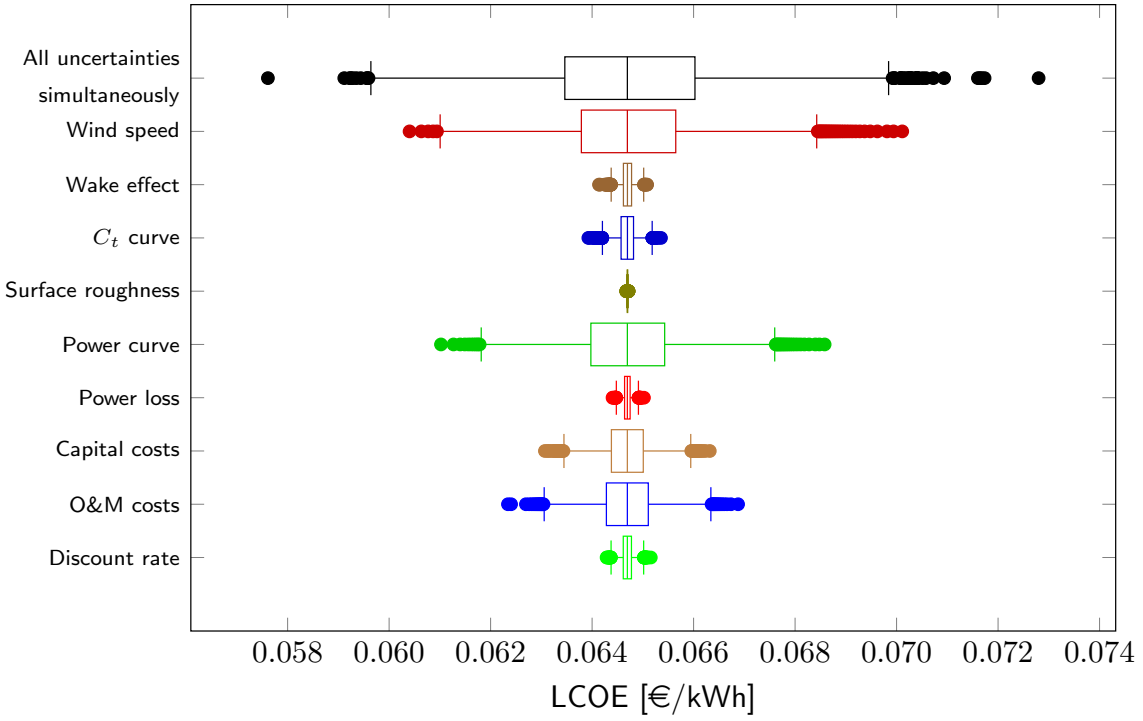

Figure 27: Test with nine uncertainty parameters wind speed, wake effect, C_t curve, surface roughness, power curve, power loss, capital costs, O&M costs and discount rate at the same time (simultaneously) compared to single effects of the uncertainty parameters computed with the quasi-Monte Carlo method with 4096 samples.

In Table 7 the results in terms of the variance based sensitivity analysis support our statements from above and this time it can be seen that the biggest sensitivities are from the uncertainties ξ_{wind} with sensitivity $S_{\text{wind}} = 51.13\%$, followed by ξ_{power} with sensitivity $S_{\text{power}} = 31.05\%$, $\xi_{\text{o\&m}}$ with sensitivity $S_{\text{o\&m}} = 10.11\%$ and ξ_{capital} with sensitivity $S_{\text{capital}} = 05.88\%$. The other uncertainties have a negligibly low sensitivity in comparison to the four mentioned uncertainties because they are all below 2%. As

for the uncertain net annual energy production, the wake effect ξ_{wake} has a very low influence concerning this quantity of interest with a sensitivity of $S_{\text{wake}} = 00.51\%$ whereas the uncertain C_t curve has a sensitivity of $S_{\text{ct}} = 01.11\%$.

$\text{var}(\text{Y}) = 3.7517 \cdot 10^{-6}$		
Uncertainty	$\text{var}_{\xi_i}(E_{\xi_{\sim i}}(\text{Y}\|\xi_i))$	S_i
ξ_{wind}	$1.9184 \cdot 10^{-6}$	**51.13%**
ξ_{wake}	$1.9239 \cdot 10^{-8}$	**00.51%**
ξ_{ct}	$4.1659 \cdot 10^{-8}$	**01.11%**
ξ_{rough}	$3.5824 \cdot 10^{-9}$	**00.10%**
ξ_{power}	$1.1649 \cdot 10^{-6}$	**31.05%**
ξ_{loss}	$1.2366 \cdot 10^{-6}$	**00.28%**
ξ_{capital}	$2.2074 \cdot 10^{-7}$	**05.88%**
$\xi_{\text{o\&m}}$	$3.7909 \cdot 10^{-7}$	**10.11%**
ξ_{rate}	$1.8181 \cdot 10^{-8}$	**00.48%**

Table 7: Results of variances based sensitivity analysis of levelized cost of energy (LCOE) calculation with 4096 samples quasi-Monte Carlo Method.

7.1.3 Net Present Value

In Figure 28 the results regarding to the net present value for all influencing uncertainty parameters, i.e. wind speed, wake effect, C_t curve, surface roughness, power curve, power loss capital costs, operation and maintenance costs, discount rate and energy price are shown. The result for each uncertain parameter is shown in a separate boxplot. It can be seen that the wind speed has the largest impact on the net present value in comparison to the other single uncertainties. The topmost boxplot represents the model output for the case that the ten uncertainties are perturbed at the same time. The result can be compared with the boxplots for the case of the perturbation of the single uncertainties. Again, we can see that the simultaneous perturbation of the parameters leads to a bigger spread of the results comparing to the perturbation of a single parameter. The parameter energy price and power curve are behave almost similar because both are multiplied in the same term for the net present value calculation.

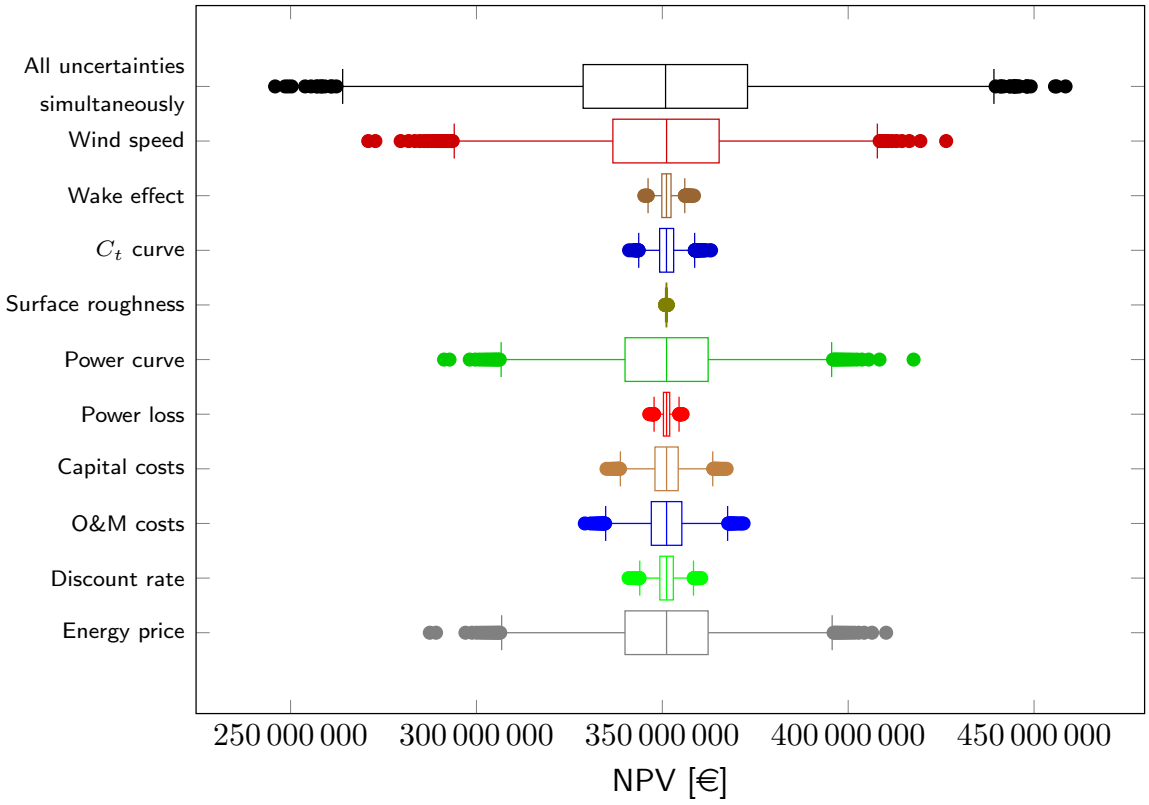

Figure 28: Test with ten uncertainty parameters wind speed, wake effect, C_t curve, surface roughness, power curve, power loss, capital costs, O&M costs, discount rate and energy price at the same time (simultaneously) compared to single effects of the uncertainty parameters computed with the quasi-Monte Carlo method with 4096 samples.

Table 8 shows the results in terms of the variance based sensitivity analysis. As

already seen in the boxplot, the results clearly indicate that ξ_{wind} has the biggest influence on the variance of the model output with a sensitivity coefficient of $S_{\text{wind}} = 41.59\%$. The second highest sensitivity is linked to the two uncertainties of the power curve with $S_{\text{power}} = 25.46\%$ and the energy price with $S_{\text{energy}} = 25.45\%$. This time the uncertainty $\xi_{\text{o\&m}}$ has a lower influence on the output with $S_{\text{o\&m}} = 03.43\%$ and the uncertain capital costs lead to a sensitivity $S_{\text{capital}} = 01.99\%$ All remaining coefficients are below 1% and thus have a rather negligible influence on the variance of the model output. As for the uncertain net AEP calculation the uncertain parameter ξ_{rough} representing the uncertain surface roughness has, at least with a perturbation of $\pm 5\%$, almost no influence on the variance of the output. As for the other two analysis before, the wake effect ξ_{wake} have a very low influence concerning this quantity of interest with a sensitivity of $S_{\text{wake}} = 00.33\%$ whereas the uncertain C_t curve, which is used in the calculation for the wake effect, has a sensitivity of $S_{\text{ct}} = 00.81\%$.

$\text{var}(\text{Y}) = 1.0807 \cdot 10^{15}$		
Uncertainty	$\text{var}_{\xi_i}(E_{\xi_{\sim i}}(\text{Y}\|\xi_i))$	S_i
ξ_{wind}	$4.4940 \cdot 10^{14}$	**41.59%**
ξ_{wake}	$3.5298 \cdot 10^{12}$	**00.33%**
ξ_{ct}	$8.7789 \cdot 10^{12}$	**00.81%**
ξ_{rough}	$3.3726 \cdot 10^{10}$	**00.00%**
ξ_{power}	$2.7508 \cdot 10^{14}$	**25.46%**
ξ_{loss}	$1.5913 \cdot 10^{12}$	**00.15%**
ξ_{capital}	$2.1487 \cdot 10^{13}$	**01.99%**
$\xi_{\text{o\&m}}$	$3.7106 \cdot 10^{13}$	**03.43%**
ξ_{rate}	$7.2907 \cdot 10^{12}$	**00.68%**
ξ_{energy}	$2.7501 \cdot 10^{14}$	**25.45%**

Table 8: Results of variances based sensitivity analysis of net present value (NPV) calculation with 4096 samples quasi-Monte Carlo Method.

7.1.4 Internal Rate of Return

Figure 29 represents the results regarding to the internal rate of return for all influencing uncertainty parameters, i.e. wind speed, wake effect, C_t curve, surface roughness, power curve, power loss capital costs, operation and maintenance costs and energy price, since the discount rate do not influence the internal rate of return mainly because the discount rate will be determined by the algorithm of this quantity of interest which makes the net present value equal to zero. The result for each uncertain parameter is shown in a separate boxplot. It can be seen that the wind speed has the largest impact on the net annual energy production in comparison to the other single uncertainties. The topmost boxplot represents the model output for the case that the ten uncertainties are perturbed at the same time. The result can be compared with the boxplots for the case of the perturbation of the single uncertainties. We can see that the simultaneous perturbation of the parameters leads to a bigger spread of the results comparing to the perturbation of a single parameter, as expected.

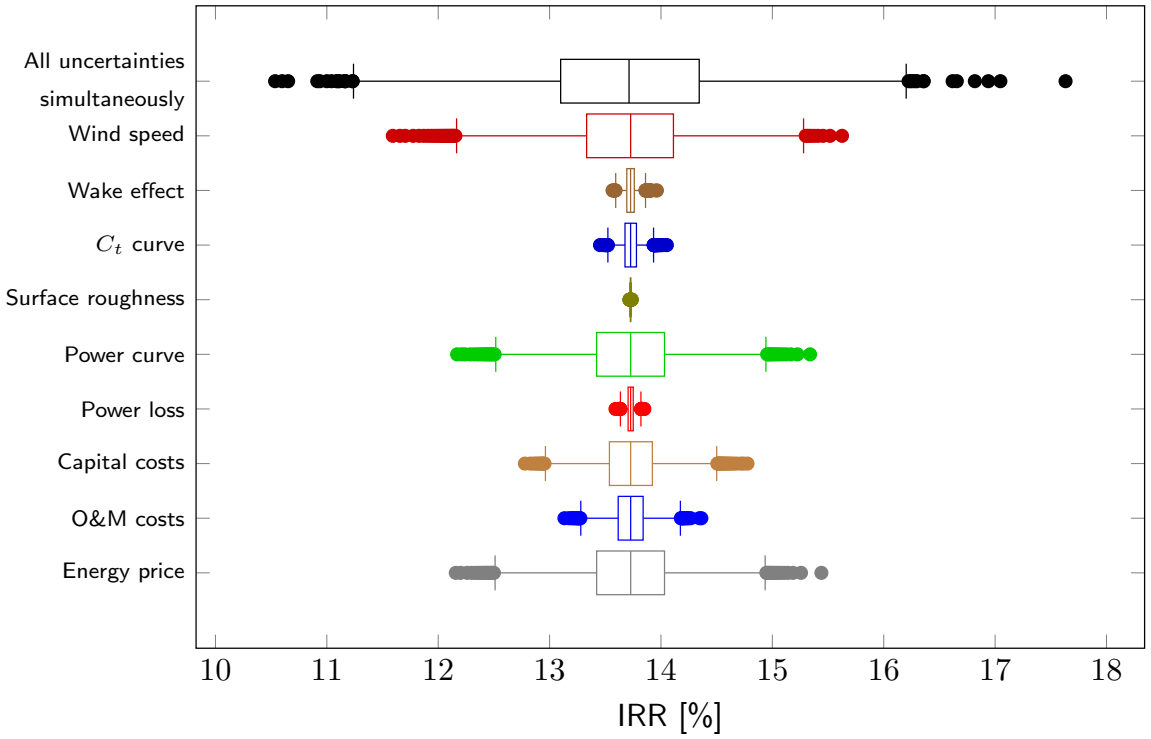

Figure 29: Test with nine uncertainty parameters wind speed, wake effect, C_t curve, surface roughness, power curve power loss, capital costs, o&m costs and energy price at the same time (simultaneously) compared to single effects of the uncertainty parameters computed with the quasi-Monte Carlo method with 4096 samples.

Table 9 shows the results in terms of the variance based sensitivity analysis. As already seen in the boxplot, the results clearly indicate that ξ_{wind} has the biggest

influence on the variance of the model output with a sensitivity coefficient of $S_{\text{wind}} = 38.77\%$. The second highest sensitivity is linked to the uncertain power curve with $S_{\text{power}} = 23.66\%$ and furthermore to the uncertain parameter of the energy price with almost the same result with $S_{\text{energy}} = 23.68\%$. This time the uncertainty ξ_{capital} has a higher influence on the output with $S_{\text{capital}} = 09.51\%$. The the uncertain annual operation and maintenance costs have a sensitivity of $S_{\text{o\&m}} = 03.20\%$. All remaining coefficients are below 1% and thus have a rather negligible influence on the variance of the model output. As for the uncertain net AEP calculation the uncertain parameter ξ_{rough} representing the uncertain surface roughness has, at least with a perturbation of $\pm 5\%$, almost no influence on the variance of the output. As for all quantities of interest, the wake effect ξ_{wake} have a very low sensitivity of $S_{\text{wake}} = 00.31\%$ and again the uncertain C_t curve have a higher sensitivity of $S_{\text{ct}} = 00.76\%$.

$\text{var}(\text{Y}) = 8.6264 \cdot 10^{-1}$		
Uncertainty	$\text{var}_{\xi_i}(E_{\xi_{\sim i}}(\text{Y}\|\xi_i))$	S_i
ξ_{wind}	$3.3445 \cdot 10^{-1}$	**38.77%**
ξ_{wake}	$2.6961 \cdot 10^{-3}$	**00.31%**
ξ_{ct}	$6.5709 \cdot 10^{-3}$	**00.76%**
ξ_{rough}	$5.3411 \cdot 10^{-5}$	**00.01%**
ξ_{power}	$2.0413 \cdot 10^{-1}$	**23.66%**
ξ_{loss}	$1.2298 \cdot 10^{-3}$	**00.14%**
ξ_{capital}	$8.2068 \cdot 10^{-2}$	**09.51%**
$\xi_{\text{o\&m}}$	$2.7643 \cdot 10^{-2}$	**03.20%**
ξ_{energy}	$2.0425 \cdot 10^{-1}$	**23.68%**

Table 9: Results of variances based sensitivity analysis of IRR calculation with 4096 samples quasi-Monte Carlo Method.

7.1.5 Summary of the Results

In this subsection we summarize the most relevant results:

- The perturbation of all uncertain input parameters simultaneously leads always to the biggest spread of outcomes, which was expected.
- For the single case the uncertain wind speed ξ_{wind} has the biggest sensitivity concerning all quantities of interest.
- For the single case the uncertain power curve ξ_{power} and if considered the uncertain energy price ξ_{energy} have also big sensitivities regarding to all quantities of interest.

- The uncertainties which are affecting the wake effect, i.e. the uncertain C_t curve ξ_{ct} and the uncertain surface roughness ξ_{rough} and the uncertainty of the wake effect ξ_{wake} itself have a very low sensitivity regarding to the quantities of interest.
- The uncertain surface roughness ξ_{rough} has almost no influence on the quantities of interest with a sensitivity of $S_{\text{rough}} \approx 0\%$ for the net annual energy production, net present value and internal rate of return.

8 Conclusion

In this thesis the uncertainty quantification of offshore wind farm models was investigated. At first the wind model was presented and its modeling from the measured raw wind data into wind direction sectors and wind speed distributions as weibull distributions. After that the wake model was explained and the turbine properties were introduced to calculate the gross annual energy production. With these models it was possible for us to compute results for our quantity of interest, which are a measure to make statements for a specific constellation of an offshore wind farm.
From there we declared the uncertain input parameter of the offshore wind farm model and achieve an offshore wind farm model which is considering uncertainties. The models and formulas were transformed by the declared uncertainties and the deterministic model became a stochastic model in that way that the input parameter get perturbed by normally distributed random variables.
To make an uncertainty quantification on the model's input parameters we used the technique of Monte Carlo sampling. This method was developed to cover an n-dimensional space with the use of random values to obtain numerical results. In every sample step of this thesis, random variables were used as input parameters so that within a certain number of samples the whole needed outcome space of the random variables was covered. We presented this technique but because of bad performance we introduced two modifications of this technique to obtain a better performance by decreasing the number of samples for meaningful results. Also we showed how to use the sample methods to transform them to normally distributed random variables which is one of the most common distributions in uncertainty quantification.
In the results section we showed how to use our stochastic model in combination with the uncertainty quantification techniques and plotted the results in a comprehensible boxplot format. Furthermore we introduced the so called variance based sensitivity analysis which made it easy to compute the influence of each uncertainty regarding to the quantities of interest. It turned out that the most significant uncertainty is the wind speed because in every considered quantity of interest it was the parameter with the biggest sensitivity. It also turned out that the uncertainties which are affecting the wake effect, i.e. the C_t curve and the surface roughness and the uncertainty of the wake effect itself have a very low sensitivity regarding to the quantities of interest.
As an outlook this means that the research has to be focused more on a precise prediction of the wind conditions and the climate conditions in future as the wind speed has the biggest influence on the quantities of interest. Also it would be interesting to make the investigation of this thesis with real world values as the economic parameter are roughly estimated and all the uncertainties are all perturbed with the same value which may not be realistic in real world because every uncertainty may vary with its real perturbation.

References

[1] Paul Bratley and Bennett L. Fox. Algorithm 659: Implementing sobol's quasirandom sequence generator. *ACM Trans. Math. Softw.*, 14(1):88–100, March 1988.

[2] Jungchul Choi and Martin Shan. Advancement of jensen (park) wake model. *Journal of Physics: Conference Series*, pages 1–8, 2013.

[3] DNV. Framework for the categorisation of losses and uncertainty for wind energy assessments. Technical report, KEMA, 2013.

[4] Mike Giles. Approximating the erfinv function. *GPU Gems*, 2010.

[5] Paul Glasserman. *Monte Carlo methods in financial engineering*. Applications of mathematics. Springer, New York, 2004. Première parution en édition brochée 2010.

[6] J. S. Gonzalez, A. G. G. Rodriguez, J. C. Mora, J. R. Santos, and M. B. Payan. A new tool for wind farm optimal design. In *2009 IEEE Bucharest PowerTech*, pages 1–7, June 2009.

[7] Gregor Heiming. Modeling and simulation of offshore wind farms. Bachelor thesis, RWTH Aachen University, Aachen, Germany, 2015.

[8] T. J. Sullivan. *Introduction to Uncertainty Quantification*. Springer, Heidelberg, New York, Dordrecht, 2015.

[9] Niels Otto Jensen. *A note on wind generator interaction*. Number 2411 in Risø-M. Risø National Laboratory for Sustainable Energy, 1983.

[10] Katić I. Jensen N.O., Højstrup J. A simple model for cluster efficiency. In *European Wind Energy Association Conference and Exhibition*, pages 407–410, Rome - Italy, 7-9 Octorber 1986. European Wind Energy Association.

[11] Matthew A. Lackner and Christopher N. Elkinton. An analytical framework for offshore wind farm layout optimization. *Wind Engineering*, 31(1):17–31, 2007.

[12] Bernhard Lange, Hans-Peter Waldl, Algert Gil Guerrero, Detlev Heinemann, and Rebecca J. Barthelmie. Modelling of offshore wind turbine wakes with the wind farm program flap. *Wind Energy*, 6(1):87–104, 2003.

[13] Bernhard Lange, Søren Larsen, Jørgen Højstrup, and Rebecca Barthelmie. Importance of thermal effects and sea surface roughness for offshore wind resource assessment. *Journal of wind engineering and industrial aerodynamics*, 92(11): 959–988, 2004.

[14] Pierre L'Ecuyer. *Random number generation and quasi-Monte Carlo*. John Wiley, Chichester, UK, June 2015.

[15] William J. Morokoff and Russel E. Caflisch. Quasi-monte carlo integration. *Journal of Computational Physics*, 122(2):218 – 230, 1995.

[16] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C (2Nd Ed.): The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 1992.

[17] Brian K. Beachkofski Ramana V. Grandhi. Improved distributed hypercube sampling. In *43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference / Denver, CO*, 4 2002.

[18] Samik Raychaudhuri. Introduction to monte carlo simulation. In *Proceedings of the 40th Conference on Winter Simulation*, WSC '08, pages 91–100. Winter Simulation Conference, 2008.

[19] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010.

[20] Michele Samorani. The wind farm layout optimization problem. In *Handbook of wind power systems*, pages 21–38. Springer, 2013.

[21] Ralph C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2013.

[22] Angelo Tesauro, Pierre-Elouan Réthoré, and Gunner Chr. Larsen. *State of the art of wind farm optimization*, chapter 0, page 11. European Wind Energy Association (EWEA), 2012.

[23] A. Tuzuner and Zuwei Yu. A theoretical analysis on parameter estimation for the weibull wind speed distribution. In *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–6, July 2008.

[24] Jon Wiernga. Representative roughness parameters for homogeneous terrain. *Boundary-Layer Meteorology*, 63(4):323–363, 1993.

[25] Jannick Wolters. Uncertainty quantification of wind farm models. Master's thesis, RWTH Aachen University, Aachen, Germany, 2016.