



Wind Farm Layout Optimisation Under Uncertainty

Using Surrogate Models

J.S. Chuang

Master of Science Thesis

Wind Farm Layout Optimisation Under Uncertainty Using Surrogate Models

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Aerospace Engineering
at Delft University of Technology

J.S. Chuang

November 24, 2020

Student number: 4790634

Supervisor: Dr.ir. M.B. Zaayer TU Delft - Wind Energy

Committee: Prof.dr.ir. S.J. Watson TU Delft - Wind Energy

Dr.ir. R.P. Dwight TU Delft - Aerodynamics

An electronic version of this thesis is available at <http://repository.tudelft.nl/>
Cover picture: Thorntonbank Wind Farm, by © Hans Hillewaert / CC BY-SA 4.0

Wind Energy Group · Faculty of Aerospace Engineering
Delft University of Technology



Copyright ©
All rights reserved.

Abstract

Wind Energy has experienced rapid growth in recent decades. As the size of offshore wind farms also increases due to the great demand for the renewable energy, the layout of the wind farm requires optimisation to maximise the energy yield over the lifetime of the wind farm. It has been shown that the inter-year variation could lead to a significant risk for the wind farm developer. This research aims to develop a method to create a wind farm layout that is robust against the uncertainty source, the inter-year variation of Weibull parameters and wind direction sector probabilities. A wind farm layout optimisation problem under uncertainty corresponds to optimisation under uncertainty (OUU), which is computationally expensive. It is, therefore, proposed to solve the computational issue by applying surrogate modelling. The resulting layout should have a consistent performance throughout its wind farm lifetime period.

The research considers and compares two different kinds of surrogate models, polynomial chaos expansion (PCE) and Kriging, on two different applications. First, the surrogate model is built for the wind farm power model to emulate the medium-fidelity model, which requires to evaluate the wake effect in every wind speed and wind direction. From the results, it is concluded that both surrogate techniques provide estimation with an error within 2% compared to the medium-fidelity model, and the PCE model is 25% faster to construct than the Kriging model. Another application of the surrogate model is for the uncertainty quantification (UQ). As the computational expensive Monte Carlo method is traditionally used to propagate the uncertain variables, the use of surrogate models can further reduce the number of uncertain samples to obtain the statistic output response. According to the results, the samples are reduced from 100,000 to 100 with an adequate accuracy on the estimation.

Once the statistic of annual energy production (AEP) of the wind farm can be computed efficiently, the optimisation under uncertainty problem is performed with the genetic algorithms. The research considers two different starting layouts, which is a self-generated layout by the algorithm and a well-designed initial layout by the author. The research discovers that the optimisation brings the self-generated case to a performance that is similar to the well-designed layout. In the other case, the optimiser seems to be stuck at the well-designed initial layout and can not easily find a better one. Nonetheless, the computational time of OUU is reduced by using the surrogate UQ model, which reaches a similar time to the deterministic optimisation with the medium-fidelity model. Furthermore, OUU provides a layout that may be more robust to the inter-year variation as indicated by a higher P90 value of AEP than the layout from the deterministic optimisation.

Acknowledgements

First, I would like to express my appreciation to my supervisor Dr. Ir. Michiel Zaaijer. Even though we are not able to meet up for discussion due to the corona measures, I still feel his great responsibility and care over my research project. His guidance continually inspires me, and he has really been a role model for me as a professional.

Also, special thanks to Prof. Dr. Ir. Simon Watson and Dr. Ir. Richard Dwight for providing me insights and feedback on my research.

Many thanks to my friends in the Netherlands. The whole journey of my Master degree would have been different without them. They are supportive and leading me to overcome the difficulties through my life in the Netherlands.

Finally, I would like to thanks my girlfriend and my family for their support, patience and true love. The unlimited support from them has provided strength to every day of my life.

Dank je wel!

*Delft, The Netherlands,
November 24, 2020*

Jia-Shian Chuang

Table of Contents

Abstract	V
Acknowledgements	VII
Glossary	XIX
List of Acronyms	XIX
List of Symbols	XX
1 Introduction	1
1-1 Background information	1
1-2 Problem analysis	1
1-3 Objectives	3
1-4 Methodology	4
1-5 Thesis outline	5
2 Wind Farm Annual Energy Yield	7
2-1 Introduction to the chapter	7
2-2 Wind farm energy estimation	7
2-2-1 Introduction	7
2-2-2 Annual energy production of a single turbine	7
2-2-3 Annual energy production of a wind farm	8
2-3 Wake phenomenon and modelling	9
2-3-1 Overview	9
2-3-2 Jensen model	10
2-4 Wind farm related data	12
2-4-1 Variation of wind speed and wind direction	12
2-4-2 Wind turbine model and wind farm layout	12

3	Surrogate Modelling	15
3-1	Introduction to the chapter	15
3-2	Surrogate modelling	15
3-3	Polynomial Chaos Expansion (PCE)	16
3-3-1	Introduction	16
3-3-2	Calculation of coefficients	18
3-4	Kriging	20
3-4-1	Introduction	20
3-4-2	Kriging trend	23
3-4-3	Correlation function	23
3-4-4	Estimation of hyperparameters	24
3-5	Verification of the surrogate models	25
3-5-1	The original model	25
3-5-2	Verification of the PCE model	28
3-5-3	Verification of the Kriging model	34
3-6	Comparison of the surrogate models	38
3-7	Summary of the chapter	41
4	Effectiveness of Surrogate Models	43
4-1	Introduction to the chapter	43
4-2	Effectiveness on different wind farm layouts	43
4-2-1	Overview of wind farm layouts	43
4-2-2	Results of surrogate models on different wind farm layouts	46
4-3	Randomness from the sampling scheme	48
4-4	Summary of the chapter	49
5	Uncertainty Quantification	51
5-1	Introduction to the chapter	51
5-2	Background overview	51
5-2-1	Introduction to uncertainty quantification	51
5-2-2	Wind resource inter-year uncertainty	52
5-2-3	Statistical inference on wind data	54
5-2-4	Quantity of interest	55
5-3	Uncertainty propagation methods	57
5-3-1	Monte Carlo method	57
5-3-2	Propagation with surrogate models	57
5-4	Implementation of the surrogate models	59
5-4-1	Overview	59
5-4-2	Verification of different configurations	61
5-5	Sensitivity analysis on the uncertain parameters	67
5-5-1	Introduction	67
5-5-2	The effect of Weibull parameters	67
5-5-3	The effect of sector probabilities	69
5-6	Summary of the chapter	70

6	Wind Farm Layout Optimisation	71
6-1	Introduction to the chapter	71
6-2	Optimisation problem definition	71
6-3	Optimisation algorithms	73
6-3-1	Overview	73
6-3-2	Introduction to genetic algorithm	74
6-3-3	Binary genetic algorithm	75
6-3-4	Real-coded genetic algorithm	76
6-3-5	Optimisation toolbox options	76
6-4	Deterministic optimisation	77
6-4-1	Optimisation with the original model	78
6-4-2	Optimisation with the surrogate models	79
6-5	Optimisation under uncertainty	83
6-5-1	Optimisation with the UQ model	84
6-6	Summary of the chapter	86
7	Conclusion & Recommendations	89
7-1	Conclusions	89
7-2	Recommendations for future work	90
A	Surrogate model flowchart	93
B	Kriging modelling parameters	95
C	WAsP wind farm report with wake model output	97

List of Figures

1-1	Flowchart of two different optimisation procedures (adopted from [43]).	2
1-2	Flowchart of the Master thesis research.	4
2-1	The power curve and wind speed probability distribution.	8
2-2	Illustration of the wind rose used in this research.	9
2-3	Illustration of Jensen single wake model and when the wake partially intersects with a downstream turbine.	11
2-4	The probability distribution of variables constructed from the weather data of the IJmuiden weather station in a year.	12
2-5	Overview of the wind farm layout used in this study.	13
3-1	Illustration of two orthogonal polynomials.	18
3-2	An illustration of the prior (left) and posterior Gaussian process (right) [30].	21
3-3	The wind turbine power curve and the power output rose of the wind farm under investigation using the original model.	26
3-4	The two-dimensional power production surface plot of the layout for every wind speed and wind direction using the original model.	26
3-5	Different order of PCE quadrature models on the NREL 5MW turbine PV curve	29
3-6	The difference [%] of power output between the estimation by PCE and the original model for various polynomial basis order with different PCE methods at 10 [m/s] wind speed.	30
3-7	The difference [%] of power output between the estimation by PCE and the original model for varying number of sampling points on wind direction with different orders at 10 [m/s] wind speed.	30
3-8	PCE power output surrogate models on wind direction at 10 [m/s] wind speed.	31
3-9	The two-dimensional power production surface plots of the investigating layout for every wind speed and wind direction using PCE models.	32
3-10	The difference [%] of power output between the estimation by PCE and the original model for varying number of collocation points on wind speed and wind direction.	33

3-11	The difference [%] of the AEP approximation between the surrogate model and the original model using different collocation points in wind speed and wind direction.	33
3-12	The Kriging PV surrogate model of the NREL 5MW wind turbine	34
3-13	The difference [%] of the power output between the estimation by Kriging and the original model for varying number of sampling points on wind direction with different Kriging methods at 10 [m/s] wind speed.	35
3-14	Kriging power output surrogate models on wind direction at 10 [m/s] wind speed.	36
3-15	The two-dimensional power production surface plot of the investigating layout for every wind speed and wind direction using the Kriging model.	37
3-16	The difference [%] of the AEP between the estimation by Kriging and the original model for varying number of collocation points on wind speed and wind direction.	37
3-17	The difference [%] of the AEP approximation in between the surrogate model and the original model using different collocation points in wind speed and wind direction.	38
3-18	Comparison of time required and accuracy for the two surrogate models.	38
3-19	The two-dimensional power production surface plot of the investigating layout for every wind speed and wind direction using the two surrogate models with different sampling schemes.	40
4-1	Overview of the 6 additional wind farm layouts under investigation.	44
4-2	Power roses of each layout when wind speed is at 10 [m/s].	45
4-3	The AEP percentage error of all wind farm layout cases evaluates by the original model and the surrogate models.	47
4-4	The percentage error from the random sampling schemes for the two surrogate models on different wind farm layouts.	48
5-1	Flowchart of a non-intrusive forward uncertainty propagation [56].	52
5-2	A comparison of joint probability distribution of wind speed and wind direction in different years.	53
5-3	The uncertain variables concluded from the 50 years measurement data.	54
5-4	Illustration of P90 values with two normal distributions with the same mean value but different standard deviation [10].	56
5-5	Illustration of cumulative density function with P90, P50, and P10 values.	56
5-6	Variation of mean and standard deviation AEP calculated by different number of samples per Monte Carlo simulation.	59
5-7	The PDF and CDF of the AEP value using OM-MC combination.	60
5-8	The difference [%] in the estimation of P50, P90, and model building time with different number of uncertain samples for various configurations.	62
5-9	PDF and CDF of the AEP value using the PCE-MC combination.	64
5-10	PDF and CDF of the AEP value using the KG-MC combination.	64
5-11	PDF and CDF of the AEP value using the PCE-PCE combination.	64
5-12	PDF and CDF of the AEP value using the PCE-KG combination.	65
5-13	PDF and CDF of the AEP value using the KG-PCE combination.	65
5-14	PDF and CDF of the AEP value using the KG-KG combination.	65

5-15	The percentage error of the estimated statistical quantities from different combinations.	67
5-16	Sensitivity of the original model to Weibull parameters.	68
5-17	Sensitivity of capacity factor to Weibull scale and shape parameters.	69
5-18	Sensitivity of the output response (AEP) to the dominant wind direction sector probability.	70
6-1	An overview of the steps in a genetic algorithm.	75
6-2	Binary representation of the chromosome from a wind farm layout.	76
6-3	Deterministic optimisation of wind farm layout with original model.	79
6-4	Convergence history of the BGA optimisation with the surrogate model. No initial layout is given to the optimiser.	81
6-5	Convergence history of the BGA optimisation with the surrogate model. An initial layout is given to the optimiser.	81
6-6	Deterministic optimisation of wind farm layout with surrogate models.	83
6-7	Optimisation under uncertainty of wind farm layout with UQ model.	85
6-8	AEP probability distribution of the optimised layout compared to the deterministic model.	85
A-1	Flowchart of constructing a surrogate model.	94
B-1	Kriging models based on the linear correlation function.	95
B-2	Kriging models based on the exponential correlation function.	96
B-3	Kriging models based on the Gaussian correlation function.	96

List of Tables

2-1	Properties of the NREL 5MW Wind Turbine.	13
3-1	Examples of different surrogate models (adopted from [52]).	16
3-2	Corresponding orthogonal polynomials to the common probability distributions (adopted from [42]).	17
3-3	Commonly seen correlation functions and their expressions (adopted from [62]).	23
3-4	Comparison between the original model and values from the WAsP software on the wind farm performance.	27
3-5	The parameters and settings used to construct the optimum surrogate models.	39
3-6	Comparison of OM model and the surrogate models on AEP and the time required to construct the models.	40
3-7	The computational time required to build and evaluate each surrogate model.	41
4-1	Summary of the AEP of all cases along with the capacity factor of the layouts.	46
4-2	The AEP ranking of every cases using different wind farm power models.	47
5-1	Uncertain variables and the corresponding statistical properties from the inferred data.	55
5-2	The performance of the OM model compared to OM-MC method for uncertainty quantification.	60
5-3	Summary of the performance of every UQ surrogate model.	66
6-1	Parameters settings for the genetic algorithms.	77
6-2	Summary of the optimisation results regarding different optimisation methods with the original model.	78
6-3	Summary of the optimisation results regarding different optimisation methods with surrogate models and a self-generated layout.	80
6-4	Summary of the optimisation results regarding different optimisation methods with surrogate models and a given initial layout.	81

6-5	Summary of the optimisation results regarding different optimisation methods with surrogate models using fixed linearly spaced samples.	82
6-6	The improvement in AEP of the optimised layouts using RGA with respect to their starting layout.	82
6-7	Summary of the optimisation results regarding different optimisation methods with UQ models.	84
6-8	Comparison of P90 value of AEP on the optimised layout with deterministic optimisation and optimisation under uncertainty using RGA.	86

Glossary

List of Acronyms

PDF	probability density function
CDF	cumulative density function
QoI	quantity of interest
AEP	annual energy production
LCOE	levelized cost of energy
COE	cost of energy
NPV	net-present value
NREL	National Renewable Energy Laboratory
WFLO	wind farm layout optimisation
GA	genetic algorithm
OUU	optimisation under uncertainty
CFD	computational fluid dynamic
PCE	polynomial chaos expansion
SC	stochastic collocation
MC	Monte Carlo simulation
UQ	uncertainty quantification
BGA	binary genetic algorithm
RGA	real-coded genetic algorithm
WAsP	Wind Atlas Analysis and Application Program
KNMI	Koninklijk Nederlands Meteorologisch Instituut
OAT	one-at-a-time
OM	original model
OLS	ordinary least square

LARS	least angle regression
LHS	Latin hypercube sampling
LSS	linearly spaced sampling
FCR	fixed charge rate

List of Symbols

α	Decay constant
β	Vector of Kriging coefficients
δ	Wind direction [°]
ϵ	Percentage error [%]
ϵ_T	Truncation error [%]
$\hat{\beta}$	Estimate of β
λ	Weibull scale parameter
μ	Mean value
Ω	Domain of the uncertain variable
$\omega(x)$	Weight function
$\phi_i(x)$	Univariate polynomials of degree i
$\Psi(X)$	Polynomial basis function
ρ	Probability [%]
σ	Standard deviation
σ^2	Variance
θ	Hyperparameters
k	Weibull shape parameter
ξ	Uncertain variables
$\hat{Y}(x)$	Predicted output at x
\mathbb{E}	Expected value
\mathcal{L}	Likelihood function
\mathcal{M}	True model
\mathcal{N}	Gaussian or Normal distribution
\mathcal{R}	Correlation functions
\mathcal{Y}	Output responses
$\tilde{\mathcal{M}}$	Surrogate model
A	Area of the rotor [m ²]
C	Cost
C_T	Thrust coefficient
d	Distance between wind turbines [m]
D_R	Rotor diameter [m]

N	Number of samples
N_T	Number of turbines
P	Power [kW]
R	Radius of the rotor [m]
S	Distance between the wind turbines [m]
T_{yr}	Total hours in a year [hr]
U	Wind speed [m/s]
u_0	Undisturbed wind speed [m/s]
u_n	Disturbed local wind speed at turbine n [m/s]
U_{Max}	Cut-out wind speed [m/s]
U_{min}	Cut-in wind speed [m/s]
X	Random variables
x	Distance behind the wake [m]
x_i	Coordinates of the i-th wind turbine in x direction [m]
y_α	Coefficients to the polynomial basis
y_i	Coordinates of the i-th wind turbine in y direction [m]
$Z(x, \omega)$	Stationary Gaussian process

Chapter 1

Introduction

1-1 Background information

As climate change and global warming becomes a crucial topic nowadays, renewable energy has been developing rapidly in recent decades [1]. Wind Energy, in particular, has been one of the most popular sources of energy among all. In 2019, the global wind power installations exceeded 60 GW and thus brought the global cumulative installed capacity to more than 650 GW [55]. In contrast to fossil fuels, wind energy is both abundance and independent of the supply concerns.

However, similar to most renewable energy, the intermittent wind source requires a challenging economic evaluation for a potential wind farm project. Therefore, feasibility and risk analysis are often carried out by the investors and the banks before the start of the project to identify the uncertainties [2]. The uncertainties exist in the wind farm output and turbine lifetime, and thus the investment risk increased. Recently, it has become a standard measure in the industry to have a conservative evaluation that includes the effect of this uncertainty behaviour. It is because that by implementing a probabilistic approach that quantifies the effect of uncertainties could possibly reduce the costs of wind energy [56].

1-2 Problem analysis

According to Gonzalez et al. [17], the uncertainty in the returns on investment lies within the future price, costs, and wind. As profitability is the primary concern for the investors, the cost of construction, operation, and even the final decommissioning of the wind farm should be dealt with care. A common approach is to classify the uncertainty into aleatory and epistemic uncertainties [43]. According to the definition, epistemic uncertainties are the scientific uncertainty in the model of a process that results from the limited data, and knowledge [19]. They are reducible once the knowledge of the field is increased.

On the other hand, aleatory uncertainties come from the natural randomness in a process. When an increase in knowledge appears, this kind of uncertainty can be better characterised

but not reduced. In the field of wind energy, the stochastic properties such as wind speed and direction are classified as aleatory uncertainties while the wake model uncertainty is one of the epistemic uncertainties as it can be further understood with research. In this research, the uncertainties in the future price, construction costs, and investors decision are ignored. The main focus is on the future wind speed and direction distribution, which are aleatory uncertainties. In other words, the natural characteristic of wind will be investigated as it could heavily influence the annual energy production (AEP) of a wind farm. Despite the farm lifetime production is essential, Quaeghebeur [45] pointed out that the year-to-year variation in the wind resources is significant and poses financial risks. The uncertainty of inter-year variation in wind resource is therefore investigated.

The design of a wind farm often aims to generate electricity at its optimum output over its facility life span or to find the lowest cost of energy. Wind farm layout optimisation (WFLO) is a method of reaching the goal. The optimisation places the wind turbines at optimum positions within a wind farm area. It avoids excessive wake effect, which has a severe impact on energy production as the wake effect would reduce the wind speed for downstream turbines. The main objective of the layout optimisation is commonly defined by the objective function such as AEP, net-present value (NPV), and levelized cost of energy (LCOE) [20]. Except for LCOE, which aims to minimise, most of the objective functions aim to find a maximisation situation. Nevertheless, a WFLO problem is highly non-linear, where a small uncertainty can be amplified and resulting in non-smooth responses [25]. Therefore, a recent development, namely optimisation under uncertainty (OUU), is used to take the uncertainties into account.

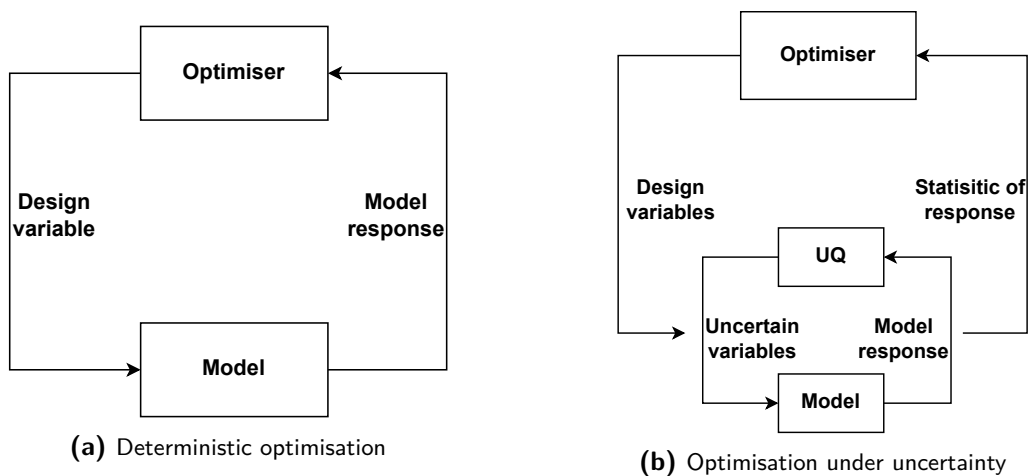


Figure 1-1: Flowchart of two different optimisation procedures (adopted from [43]).

OUU is a robust design approach that takes into account the uncertainties in the input variables by conducting uncertainty quantification (UQ) in the analyser. An example flowchart is presented in Figure 1-1b. OUU differs from the deterministic optimisation in Figure 1-1a. It has a nested loop of uncertainty quantification to compute and form a statistics of response output for the optimisation process [43]. However, many model evaluations are needed to compute the relevant statistics for every optimisation step. This results in the characteristic that OUU is computationally expensive. Furthermore, as the size of a wind farm increases, the number of simulations for the WFLO become too large and infeasible. Several approaches have been proposed to tackle the issues, such as improving the efficiency of the model or set-

ting up a better optimisation problem formulation and algorithm [43]. Nevertheless, the most recent trend for this issue is to develop a better treatment for the stochastic nature of the problem.

As a result, surrogate models, also known as metamodel or emulator, are proposed. The surrogate models emulate and approximate the computational expensive model or simulation. In general, surrogate models aim to reduce the evaluation time and cost to make a large and complex optimisation under uncertainty problem feasible. The principle is to use a limited number of runs at experimental design values. Stochastic collocation (SC), polynomial chaos expansion (PCE), and Kriging (Gaussian Processes) are some popular surrogate model techniques for uncertainty quantification. Different from Monte Carlo simulation (MC) that is based on random samples, SC and PCE utilise a smart sampling technique for the experimental design values, which significantly reduces the number of model evaluations [56]. A comparison of the two techniques was provided in the work of Eldred and Burkardt [9].

This research provides an assessment of the different surrogate models for wind energy production. The method by which the surrogate model can be constructed is also investigated. The surrogate models serve as a means to alleviate two computationally expensive processes. The first one being the wind farm performance assessment, where the evaluation of the wake effects is time-consuming. The experimental design for this stage is the wind speed and wind direction. The second implementation is for the uncertainty quantification, where the inter-year variation of the Weibull parameters and wind rose sector probabilities are uncertain. Once the statistics of the response are obtained, they are treated as the objective in the optimisation algorithm to conduct wind farm layout optimisation. It is expected that the resulting optimised wind farm layout will have a more consistent relative performance to the inter-year variation of the uncertain wind resource.

1-3 Objectives

By conducting optimisation under uncertainty studies, one wishes to understand or aims to reach an optimal design that is robust under uncertainties. With a robust optimisation, a system is more likely to handle the uncertainties in its source. Most importantly, the uncertainty quantification should be feasible as regards the computational cost and within a range of accuracy. Therefore:

The main research objective of this thesis is *to* examine an optimisation under uncertainty problem, WFLO, *by* implementing surrogate models in the wind farm performance calculation and uncertainty quantification on the inter-year variation of the wind resource.

The thesis aims to achieve the main research objective by comparing different surrogates models, namely PCE and Kriging, to the brute force UQ method, Monte Carlo simulation, on the P_{90} value of AEP. The outputs are later used to conduct optimisation on the wind farm layouts to have a robust layout.

1-4 Methodology

The following research questions can summarise the objective of this research:

1. What effects do the surrogate models have on the result of the AEP calculation?
2. What is the consequence of implementing the surrogate method in the uncertainty quantification process?
3. What is the effect of uncertainty quantification with surrogate models on the result of wind farm layout optimisation?

The methodology to answer the research questions is as follows. First research question focuses on the surrogate models for the wake effect during the estimation of the wind farm energy assessment. Surrogate models are developed to alleviate the computational cost caused by extensive model evaluations. The difference in the AEP estimation from different surrogate models for wind farm performance is measured. It is also important to acknowledge the changes in the computational cost from using the surrogate models to the original model.

The second step aims to answer the second research question. It consists of investigating the influence of the surrogate models when using in an UQ process to look for P90 of AEP. The changes in the mean and the standard deviation should be noted as well as the computational cost.

Once the previous two steps are finished, a statistic response of the AEP should be obtained. The optimisation under uncertainty problem, WFLO, is then conducted. The third research question is answered by comparing the optimised layouts from deterministic optimisation with optimised layouts from OUU, that includes surrogate models.

A detailed flowchart that summarised the steps mentioned above is shown in Figure 1-2.

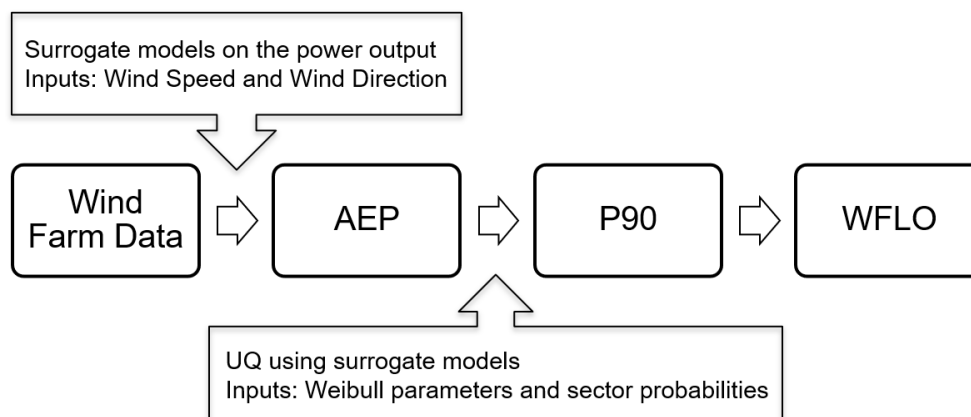


Figure 1-2: Flowchart of the Master thesis research.

1-5 Thesis outline

This chapter has introduced the problem and some background concerning the OUU and WFLO problem. Following the introduction in Chapter 1, the contents of the thesis are divided into the following chapters:

- Chapter 2 (Wind Farm Annual Energy Yield) presents the concept of wind farm performance and its indicators, namely AEP. A brief introduction of the wake effect is also included. The last section of the chapter introduces the weather data and the wind farm model of the research.
- Chapter 3 (Surrogate Modelling) demonstrates the methodologies to construct the surrogate model with PCE and Kriging. Both surrogate models are implemented to conduct the calculation of wind farm performance. The models are verified using the computation expensive medium-fidelity model. A comparison between the two models is made as well.
- Chapter 4 (Effectiveness of Surrogate Models) validates the effectiveness of surrogate models constructed in Chapter 3 on different wind farm layouts. Furthermore, the randomness from the random sampling scheme is also quantified.
- Chapter 5 (Uncertainty Quantification) provides a review of uncertainty propagation methods using surrogate models and the Monte Carlo simulation approach. A sensitivity analysis is also conducted to investigate the effect of uncertain inputs on the function output.
- Chapter 6 (Wind Farm Layout Optimisation) is dedicated to performing the wind farm layout optimisation using the developed UQ method. The research focuses on two types of genetic algorithm, that is binary genetic algorithm and real-coded genetic algorithm. Both introduction and the results of the algorithms are presented in the chapter.
- Chapter 6 (Conclusion and Recommendations) addresses the conclusion of this thesis research, and the research objectives and questions are reviewed along with some future recommendations.

Wind Farm Annual Energy Yield

2-1 Introduction to the chapter

In this chapter, the methods of how to calculate AEP are presented. AEP is commonly used as an objective function for a wind farm layout optimisation problem. The wake effects in the wind farm are later briefly introduced. Lastly, the research data, such as the wind turbine model and the wind farm layout, are presented. The purpose of this chapter is to establish the relationship between the wake model and wind farm layout optimisation. It also points out the reason of why many model evaluations are required to perform wind farm energy assessment.

2-2 Wind farm energy estimation

2-2-1 Introduction

In the planning or evaluation stage of a wind farm project, annual energy production (AEP) is commonly used as a measure of wind farm performance [40]. Different from the calculation of a solitary turbine, the calculation of AEP of a wind farm requires to take the wake effects into account. The following subsections give a detail introduction to the methods of energy estimation.

2-2-2 Annual energy production of a single turbine

It is essential to realise calculation by looking into a solitary turbine and then extending to the entire wind farm. There is a fundamental difference between the calculation of a single turbine and a whole wind farm.

The AEP of a single wind turbine is generally expressed as:

$$AEP_{WT} = \frac{hr}{yr} \int_{U_{min}}^{U_{Max}} P(U) \cdot \rho(U) dU \quad (2-1)$$

where $\frac{hr}{yr}$ is the number of hours in a year (typically 8766) while U_{min} and U_{Max} denote the cut-in and cut-out wind speed of the turbine model. P is the power produced by the turbine and is a function of the hub-height wind speed U . An electrical power curve of the turbine, as shown in Figure 2-1a is needed from the manufacturers of the wind turbine to obtain the corresponding power value. Lastly, for $\rho(U)$, ρ is the probability of the occurrence of the wind speeds.

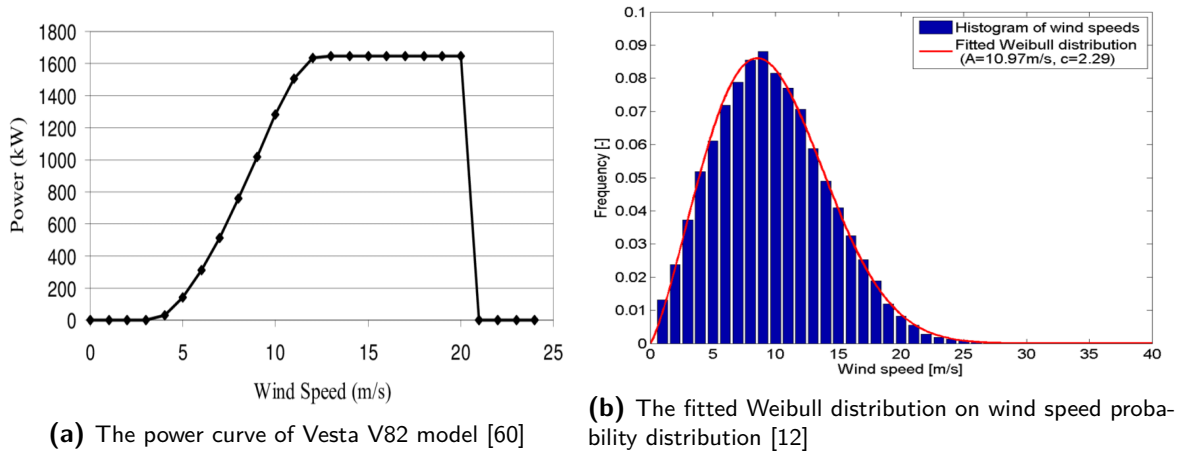


Figure 2-1: The power curve and wind speed probability distribution.

A probability density function (PDF) of the wind condition at the target site like Figure 2-1b can provide the information for $\rho(U)$. Base on empirical experiences, the Weibull distribution is commonly the preferred distribution for the wind speed distribution

$$\rho(U|\lambda, k) = \frac{k}{\lambda} \left(\frac{U}{\lambda}\right)^{k-1} e^{-\left(\frac{U}{\lambda}\right)^k} \quad (2-2)$$

There are two essential parameters for the Weibull distribution, namely, the scale parameter λ and the shape parameter k in Eq. (2-2). These two parameters can be determined by fitting a Weibull distribution to the histogram of wind speed distribution as seen again in Figure 2-1b.

2-2-3 Annual energy production of a wind farm

The calculation for a wind farm requires additional information due to the wake effect caused by different wind direction. Eq. (2-3) is a formula of the annual energy production for the entire wind farm, in which N_T is the total number of the wind farm and n is the n -th turbine under investigation. To determine $P_n(U)$ in Eq. (2-3), wake effect needs to be evaluated to obtain the local wind speed at turbine n . It is due to the wake effect that every time the wind direction changes, the local wind speed at turbine varies. Further, a wind rose as shown in Figure 2-2 which represents the PDF of the wind speed over the wind direction, δ , is

frequently needed to obtain information on $\rho(U, \delta)$. It denotes a bivariate probability density function that consists of free-stream wind speed and wind direction.

$$AEP_{Farm} = \frac{hr}{yr} \sum_{n=1}^{N_T} \int_0^{360} \int_{U_{min}}^{U_{Max}} P_n(U) \cdot \rho(U, \delta) dU d\delta \quad (2-3)$$

Estimating AEP for a wind farm is much more complicated as the wake effects heavily influence the wind speed received by some turbines in certain wind directions. The topic of wind turbine wakes has been a focus of wind energy research for quite a long period [59]. This phenomenon is receiving more attention since the size of the wind turbine and the wind farm are both increasing rapidly. Next, a section regarding the topic of wake phenomenon is presented.

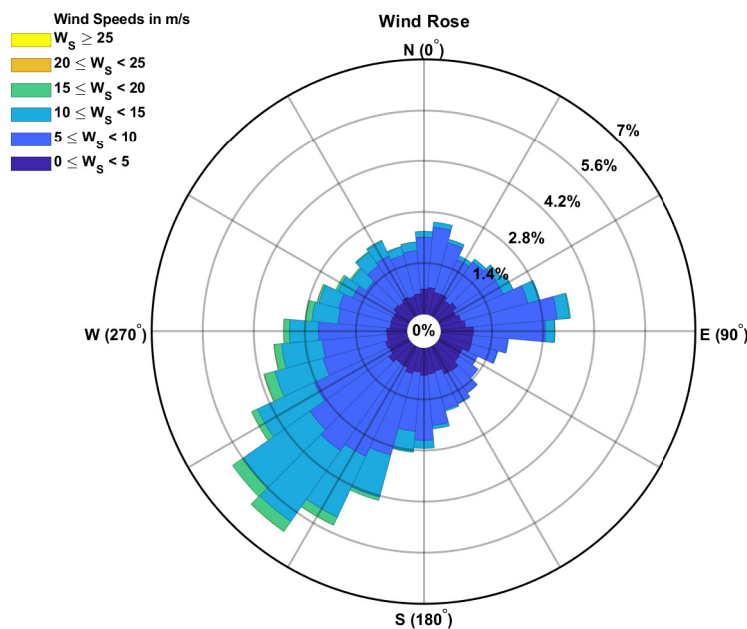


Figure 2-2: Illustration of the wind rose used in this research.

2-3 Wake phenomenon and modelling

2-3-1 Overview

A wind turbine converts the wind energy into usable energy by converting the kinetic energy from the wind into mechanical energy, which then transfers into electrical energy [50]. As the energy from the wind is extracted, the wind speed behind the rotor reduced, and the airflow begins to swirl [37]. Therefore, the downstream wind turbine produces less energy since it receives lower mean wind speed and higher turbulence intensity. In addition, wind turbine

wake can be divided into the near wake and far wake [59]. The near wake region is more dominated by the turbulence and far wake region mainly consists of convection and turbulent diffusion.

Clifton et al. [7] defined losses as the amounts of energy that a wind farm should be able to extract but failed to due to technical or other reasons and wake effect is one of the losses. The estimation of losses should be dealt with care since it will heavily influence the estimation of wind farm energy production. The losses from the wake effect are essential as it is a crucial aspect of the wind farm economic evaluation [14]. In order to take the wake effect into account, various wake models have been developed. In general, the far wake region is more important than the near wake region in a WFLO problem [50]. The first wind turbine wake modelling approach was the kinematic model introduced by Lissaman [32], which was based on self-similar velocity deficit profiles. The kinematic model, however, does not take the initial expansion region of the wake nor the turbulent intensity in the wake behind the turbine into account.

As developed over time, modelling the wake effect can be done with the help of computational fluid dynamic (CFD) software. The Direct Numerical Simulation, Reynolds Averaged Navier-Stokes, and Large-Eddy Simulation fall in this category. However, even though CFD models are able to reduce the epistemic uncertainty from the analytic model, their complexity and high computational requirements are still infeasible for large optimisation. For example, Direct Numerical Simulation could take up to weeks to complete its simulation. Therefore, an analytical model is still usually preferred and can already reach quite satisfying results. Those computationally expensive approaches thus do not fall into the scope of this research. Alternatively, wake models, which are more computationally inexpensive while providing results with sufficient accuracy, are investigated by previous works [13, 14, 29, 49]. Among them, the Jensen model can provide an acceptable trade-off between the computational cost and prediction errors for wind farms, according to VanLuvanee [57]. The research, therefore, chooses the Jensen model to be the wake model in use. The Jensen model is explained in more detail in the next subsection.

2-3-2 Jensen model

The Jensen model is the most widely used wake model, and it is also known as the Park model. An illustration of the model is shown in Figure 2-3a. However, the model proposed by Jensen [21] is a single wake model. It was later modified by Katic et al. [24] to calculate the effect of multiple wakes. The model is constructed upon the mass conservation law in a control volume which is applied to the flow passing the wind turbine. The velocity in the wake region can be described as a function of distance x from the upstream turbine, and the wake is assumed to expand linearly. The wake model heavily depends on a dimensionless scalar, α , which is the decay constant that determines the rate of the wake recovery. α is a parameter that could be influenced by, for example, surface roughness and atmospheric stability. In order to calculate the velocity in a wake, thrust coefficient, C_T , and rotor diameter, D_R , from the wind turbine model are also needed.

In order to calculate the wake effect of multiple wind turbines in a cluster, such as a wind farm, Katic et al. [24] modified the original Jensen model. It is because the wake effect is more severe in a cluster. For such a case, it is essential to take the shadowing condition of the

turbines into account. This is one of the critical reasons that optimisation of the wind farm layout is needed. As the wake expands, if the downstream turbine is experiencing shadowing from the upstream turbine, then it can be partially or entirely in the wake of the upstream turbine. Figure 2-3b presents the illustration of when the wake partially intersects with a downstream rotor.

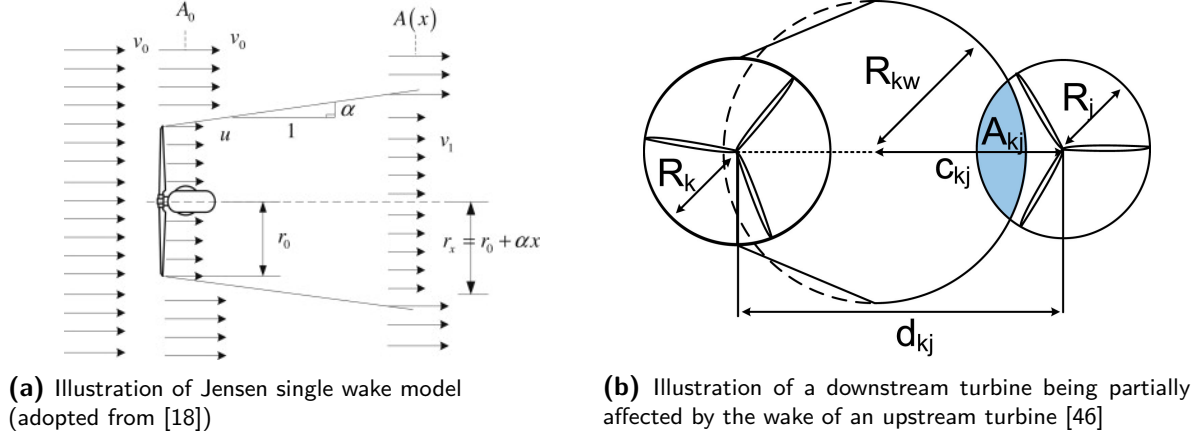


Figure 2-3: Illustration of Jensen single wake model and when the wake partially intersects with a downstream turbine.

Using Eq. (2-4) from Chowdhury et al. [6], the area of intersection, which is the effective area of the wake, can be determined.

$$A_{kj} = R_k^2 \cos^{-1} \left(\frac{d_{kj}^2 + R_k^2 - R_j^2}{2d_{kj}r_k} \right) + R_j^2 \cos^{-1} \left(\frac{d_{kj}^2 + R_j^2 - R_k^2}{2d_{kj}r_k} \right) - \frac{1}{2} \sqrt{(-d_{kj} + R_k + R_j)(d_{kj} - R_k + R_j)(d_{kj} + R_k - R_j)(d_{kj} + R_k + R_j)} \quad (2-4)$$

In Eq. (2-4), A_{kj} is the area of intersection of the wake from the front turbine k to the turbine j . R_k and R_j are the rotor radius of the two wind turbine while R_{kw} is the wake radius. Most importantly, d_{kj} is the distance in between the two turbines as this value is crucial to the wake radius R_{kw} and thus the area of intersection. On the other hands, if the downstream turbine is entirely in the wake of the upstream turbine, the effective area of the wake effect is the rotor diameter itself.

As for the multiple wake effects, they can be treated as a summation of single wake effects. A simple combination of wakes can use Eq. (2-5) while there exist many more accurate models for the summation of the wakes.

$$u_n = u_0 \left[1 - \sqrt{\sum_{i=1}^N \frac{A_{ki}}{A_j} \left(1 - \frac{u_i}{u_0} \right)^2} \right] \quad (2-5)$$

With the information of the area of intersection A_{kj} and the undisturbed wind speed u_0 , the disturbing local wind speed of turbine n , u_n , can be determined. This will be the wind speed that is "felt" by the turbine in a wake.

2-4 Wind farm related data

2-4-1 Variation of wind speed and wind direction

When it comes to the wind resource, wind speed and wind direction are the two common variables. To better understand the distribution of the two variables, a time series measurement data from a real wind site is used. The weather data from the IJmuiden weather station is provided on the Koninklijk Nederlands Meteorologisch Instituut (KNMI) website. IJmuiden locates at the mouth of the North Sea Canal to Amsterdam with the detailed coordinate of $52^{\circ}28' N$ and $04^{\circ}34' E$. The wind mast at the station measures the temperature, air pressure, wind and precipitation from Jan 1, 1971, and continues measuring. The data consists of daily averaged measurement of wind speed and wind direction at a 10-meter height above mean sea level. Thus, it will be needed to extrapolate the wind speed to the turbine hub height. The direction of North is denoted at zero degree, and the degree increases clockwise.

Figure 2-4 presents the probability distributions of the wind direction and wind speed during a year. The wind direction distribution is built upon a linear interpolation of the measurement data while a Weibull distribution is fitted to the wind speed data of all wind directions. The distribution is fitted with a scale parameter $\lambda = 8.25$ and a shape parameter $k = 2.51$, shown in Figure 2-4b.

Nonetheless, there seems to be no suitable statistical distribution to fit the wind direction distribution shown in Figure 2-4a. It is therefore decided to break the wind direction into sectors with a sector width of 30° and characterise the uncertainty in the sector probabilities. This will be further elaborated in the Chapter 5.

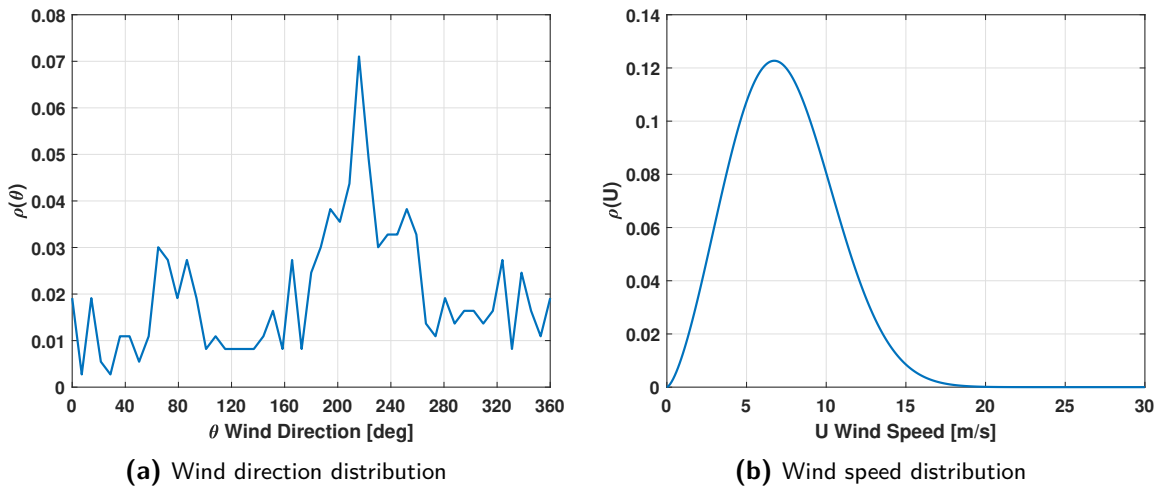


Figure 2-4: The probability distribution of variables constructed from the weather data of the IJmuiden weather station in a year.

2-4-2 Wind turbine model and wind farm layout

For the research purpose, the wind turbine model is the NREL 5MW wind turbine [23]. An overview of the turbine model is shown in Table 2-1. The hub height of the wind turbine is

at 90 meters height as this is important to extract the weather data to make the wind speed representative.

Parameter	Value/description
Rating	5 MW
Hub Height	90 m
Rotor Diameter	126 m
Cut-in, Rated, Cut-out Wind Speed	3 m/s, 11.4 m/s, 25m/s

Table 2-1: Properties of the NREL 5MW Wind Turbine.

A simply two by five wind farm layout that consists of 10 wind turbines is used in most of the research, and it is shown in Figure 2-5. The wind farm layout is designed by the author and is constructed with sufficient spacings between turbines to avoid excessive wake effect. The wake effect utilised the Jensen wake model mentioned in Section 2-3-2. The information provides here will be later used to investigate the use of the surrogate model on the evaluation of the wake model in Chapter 3 and the uncertainty quantification in Chapter 5. Meanwhile, the wind farm layout will be treated as the initial layout in the wind farm layout optimisation problem.

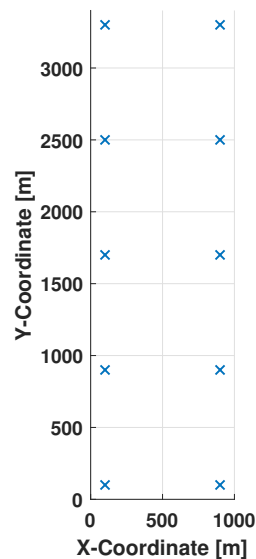


Figure 2-5: Overview of the wind farm layout used in this study.

Surrogate Modelling

3-1 Introduction to the chapter

Surrogate models were proposed due to the high computation expense of extensive simulations as accurate results often require many model evaluations. In this chapter, the general theory and application of polynomial chaos expansion and Kriging are introduced from Section 3-2 to Section 3-4. The readers who are familiar with this could go to Section 3-5 directly. Based on the theory, the surrogate models are used to emulate the model of wind farm power output, and their implementation in wind energy research are presented.

3-2 Surrogate modelling

In many engineering problems, an accurate physics-based simulation can be computationally costly to obtain the details, especially when such computations are repeatedly called. Therefore, a viable option occurs to tackle the issue, which is the use of surrogate models (i.e. meta-models). Surrogate models attempt to emulate a model so that an inexpensive-to-evaluate emulator, $\tilde{\mathcal{M}}$, can replace the expensive-to-evaluate computational model, \mathcal{M} [34]. It is, of course, less detailed than before but a lot faster to evaluate. With this speciality, the application of surrogate modelling in large computation has therefore increased rapidly. In this chapter, the aim is to build a surrogate model for the evaluation of wind farm power output, as mentioned in Chapter 2. In this case, the computational expense comes from the wake model as the wake model requires to evaluate every wind direction for the power output. The anticipated surrogate models are expecting to provide the wind farm power output with less wind speed and wind direction evaluations.

Various surrogate models have been developed and exist in the literature. Some example are given as Polynomial Chaos expansion (PCE) [3], Kriging [16, 38], Polynomial Chaos-based Kriging (PCK) [48], Stochastic Collocation (SC) [9], and Support Vector Machine (SVM) [58]. An overview of each surrogate model can be found in Sudret et al. [52].

In essence, a surrogate model is built by the following steps:

1. Define an initial set of samples in the experimental design space where the true function is evaluated
2. Determine the type of surrogate model (e.g., from Table 3-1) and fit the model to the sampled data points
3. Examine the accuracy of the surrogate model by using a set of validation data

A more detailed flowchart of constructing a surrogate model can be seen in Appendix A.

Name	Response
Polynomial chaos expansions	$\tilde{\mathcal{M}}(x) = \sum_{\alpha \in \mathbb{A}} a_{\alpha} \Psi_{\alpha}(x)$
Kriging	$\tilde{\mathcal{M}}(x) = \beta^T \cdot f(x) + Z(x, \omega)$
Low-rank tensor approximations	$\tilde{\mathcal{M}}(x) = \sum_{l=1}^R b_l (\prod_{i=1}^M v_l^{(i)}(x_i))$
Support vector machines	$\tilde{\mathcal{M}}(x) = \sum_{i=1}^m a_i K(x_i, x) + b$

Table 3-1: Examples of different surrogate models (adopted from [52]).

As the technique develops, more advanced and optimal ways of constructing a surrogate model become available. Adaptive and sequential modelling that involve the iterative procedure of fitting the surrogate model to the actual model are some examples. For simplicity, details are not described here. Nevertheless, since the purpose of a surrogate model is to build an approximation of the expensive model, the accuracy of such approximation needs to be taken into account. Moreover, the effort of constructing the surrogate model is also essential. The advantage of using a surrogate model is lost when the construction itself is complicated and time-consuming [3]. In Murcia et al. [40], the authors study the possibility of using surrogate models in predicting AEP of a wind power plant. Similar to their research, in the following sections, two popular surrogate modelling techniques, namely polynomial chaos expansion (PCE) and Kriging, are introduced with their implementation in this wind energy research.

3-3 Polynomial Chaos Expansion (PCE)

3-3-1 Introduction

Developed by Wiener [61], PCE is a smart sampling method with pseudo-spectral weighting on the samples [56]. The sampling choices are based on the probability distribution of the random variables X as they relate to sets of orthogonal polynomials, $\Psi(X)$. For random variables, there are corresponding classical orthogonal polynomials, as shown in Table 3-2. An example is that a normally distributed variable has the associated orthogonal polynomials, Hermite polynomials. PCE stands out as the polynomial functions can be evaluated faster [35]. Meanwhile, the criteria of using PCE are that it assumes the uncertain variables are independent with each other.

Using the notation from Marelli and Sudret [34], a random vector with independent variables $X = \{X_1, \dots, X_M\}$ is considered with the multivariate orthogonal polynomials that are

orthonormal to the joint probability density function $\rho(X)$. Consider also a computational model that has a response Y defined as

$$Y = \mathcal{M}(X) = \sum_{\alpha \in \mathbb{N}^M} y_\alpha \Psi_\alpha(X) \quad (3-1)$$

where y_α denotes the corresponding coefficients to the polynomial basis. However, in practical applications, Eq. (3-2) is used since the polynomial expansion will need to be truncated to a certain order.

$$\mathcal{M}(X) \approx \mathcal{M}^{PC}(X) = \sum_{\alpha \in A} y_\alpha \Psi_\alpha(X) \quad (3-2)$$

Conventionally, the polynomial basis $\Psi_\alpha(X)$ is built from a set of univariate polynomials $\{\phi_i(x)\}_{i=0}^\infty$, where $\phi_i(x)$ is a polynomial of degree i . The inner product of the polynomials is used to define the orthogonality of the polynomial if

$$\langle \phi_i, \phi_j \rangle \stackrel{def}{=} \int \phi_i(x) \phi_j(x) \omega(x) dx = c_i \delta_{ij} \quad (3-3)$$

with

$$c_i = \langle \phi_i^2 \rangle \quad (3-4)$$

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \quad (3-5)$$

where the $\omega(x)$ is the weight function and can be the PDF of random variable x , and δ_{ij} is the Kronecker delta. When c_i is 1, then the polynomials are so-called orthonormal. For common uncertain variables with given probability density functions (e.g., Normal, Uniform, Exponential), it is possible to use Askey scheme (Table 3-2 shows a subset of classical orthogonal polynomials in the Askey scheme) to find the orthogonal polynomials. The scheme leads to an optimal spectral convergence for the expansion [42]. In Table 3-2, the weight function has a difference with the probability density function due to the normalisation of the probability density function [42]. In Figure 3-1, two classical orthogonal polynomials are illustrated with different polynomial basis degree. On the other hand, if one has arbitrary random variables, it is possible to use *Gram-Schmidt* and *Stieltjes* algorithm to produce custom orthogonal polynomials for the arbitrary random variable.

Distribution	Density function	Polynomial	Weight function	Interval
Normal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	Hermite $He_n(x)$	$e^{-\frac{x^2}{2}}$	$(-\infty, \infty)$
Uniform	$\frac{1}{2}$	Legendre $P_n(x)$	1	$(-1, 1)$
Exponential	e^{-x}	Laguerre $Ln(x)$	e^{-x}	$(0, \infty)$

Table 3-2: Corresponding orthogonal polynomials to the common probability distributions (adopted from [42]).

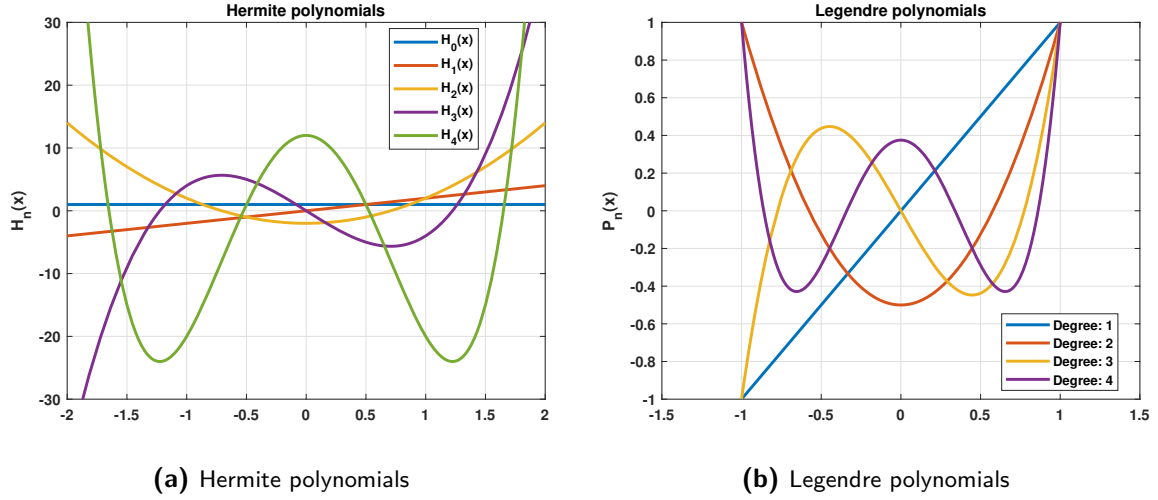


Figure 3-1: Illustration of two orthogonal polynomials.

3-3-2 Calculation of coefficients

From the previous section, the only parameters left are the polynomial chaos coefficients y_α in Eq. (3-2). There exist two approaches to compute the polynomial chaos coefficients. The first one is the *intrusive* method. An example is the *Intrusive Galerkin Projection* method. It is intrusive in the sense that the system of equations is reformulated in order to compute the coefficients. Since intrusive methods generally enlarge and alter the governing equations, *non-intrusive* methods have been more popular in recent days. In contrast, the non-intrusive methods do not need to modify the simulation codes. The non-intrusive methods compute the coefficients by sampling the input variables first and carrying out a set of model evaluation [34] to construct the surrogate model. In the following paragraphs, the non-intrusive *Projection*, *Regression*, and *Least-angle regression* methods, are introduced.

Projection methods

The projection methods is also called as *non-intrusive Galerkin Projection*, and it start from taking the inner product on both sides of the Eq. (3-2) with respect to a polynomial basis $\Psi_b(X)$ which leads to

$$\langle \mathcal{M}, \Psi_b \rangle = \sum_{\alpha \in A} y_\alpha \langle \Psi_\alpha, \Psi_b \rangle \quad (3-6)$$

The right hand side of the equation can make use of the orthogonality of the polynomials where

$$\langle \Psi_\alpha, \Psi_b \rangle = \begin{cases} \langle \Psi_\alpha^2 \rangle, & \text{if } a = b \\ 0, & \text{if } a \neq b \end{cases} \quad (3-7)$$

Eq. (3-6) then turns into solving the coefficient y_α using the following equation:

$$y_\alpha = \frac{\langle \mathcal{M}(X), \Psi_\alpha(X) \rangle}{\langle \Psi_\alpha^2 \rangle} = \frac{1}{\langle \Psi_\alpha^2 \rangle} \int \mathcal{M}(X) \Psi_\alpha(X) \rho(X) dX \quad (3-8)$$

in which the denominator $\langle \Psi_\alpha^2 \rangle$ is known analytically or can be computed inexpensively such as the Gram-Schmidt method. The coefficients can be obtained by the calculation of the expectation value \mathbb{E} , in Eq. (3-9) [34].

$$y_\alpha = \mathbb{E}[\Psi_\alpha(X) \cdot \mathcal{M}(X)] \quad (3-9)$$

One of the projection methods is the quadrature approach. The quadrature method or more specifically, Gaussian quadrature, can be expressed as a weighted-sum approach:

$$y_\alpha = \int \mathcal{M}(x) \Psi_\alpha(x) \rho(x) dx \approx \sum_{i=1}^N \mathcal{M}(x^{(i)}) \Psi_\alpha(x^{(i)}) w^{(i)} \quad (3-10)$$

The weights $w^{(i)}$ and the quadrature points $x^{(i)}$ are computed by using the theory of orthogonal polynomials and making use of the root of the polynomials shown in Table 3-2. When the problem is multi-dimensional, the tensor product quadrature can be used [34]. The number of model evaluations required for PCE by quadrature method is $(p+1)^N$, with p the degree and N the number of input variables. It is, therefore, computationally heavy when the number of input variables is large. This phenomenon is called the *curse of dimensionality*.

Regression methods

Another method to compute the coefficients in the PCE is by tackling a least-squares minimisation problem [34]. A brief introduction of the linear regression method is presented in this section. Eq. (3-1) and its truncated form Eq. (3-2) can be written as

$$Y = \mathcal{M}(X) = \sum_{j=0}^{C-1} y_j \Psi_j(X) + \epsilon_T \equiv y^\top \Psi(X) + \epsilon_T \quad (3-11)$$

where C denotes the number of coefficients and ϵ_T is the truncation error that is supposed to be a zero-mean variable [51]. The equation can be simplified into a linear system, as shown in Eq. (3-12) with m samples and n truncated terms in the polynomial chaos expansion. This results in a matrix-like equation as Eq. (3-13).

$$\Psi y^\top = \mathcal{M}(X) \quad (3-12)$$

$$\begin{bmatrix} \Psi_0(x_1) & \dots & \Psi_{n-1}(x_1) \\ \vdots & \ddots & \vdots \\ \Psi_0(x_m) & \dots & \Psi_{n-1}(x_m) \end{bmatrix} \begin{bmatrix} y_0 \\ \vdots \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} \mathcal{M}_1 \\ \vdots \\ \mathcal{M}_m \end{bmatrix} \quad (3-13)$$

The least-squares minimisation problem can be formulated as

$$\hat{y} = \arg \min \mathbb{E}[(y^\top \Psi(X) - \mathcal{M}(X))^2] \quad (3-14)$$

Commonly, the ordinary least-squares (OLS) method can be utilised to estimate the coefficients in Eq. (3-14). It requires a set of input random vectors X which forms the experimental design and the corresponding model response \mathcal{M} . The solution from the ordinary least square is therefore:

$$\hat{y} = (\Psi^\top \Psi)^{-1} \Psi^\top \mathcal{M} \quad (3-15)$$

Above all, the ordinary least squares method can have an arbitrary number of sampling points to calculate the coefficients.

Sparse PCE: Least-angle regression

When the number of input variables increases, the number of coefficients to be computed also grows. Adaptive sparse PCE is therefore developed using adaptive training algorithms to tackle this issue. Sparse PCE keeps only the most relevant polynomials and sets the coefficient of all the others to 0. In other words, sparse means that only a small number of terms are retained compared to a normal PCE [5]. Blatman and Sudret [5] applied the least-angle regression (LARS) algorithm to obtain the sparse PCE models. LARS is utilised as an effective method for variable selection and can select the polynomial basis that has the most influence on the model response. In general, the algorithm avoids the issue of overfitting when the truncating order is too high and detecting the significant coefficients of the polynomial expansion to lower the computational effort. A detailed mathematical theory and derivation can be found in Marelli and Sudret [34] and Blatman and Sudret [5].

3-4 Kriging

3-4-1 Introduction

Kriging is another popular surrogate modelling technique and is also known as Gaussian process modelling. It was first conceived by Krige [26] in 1951 for the research in geostatistics. Kriging has been a popular choice of surrogate models for the deterministic computer models as it can represent the mapping of the input and output [62]. Besides being able to approximate the output of a computational model, Kriging can provide information about the accuracy on local estimates [30]. It can quantify the interpolation errors in terms of Kriging variance. It also performs better than PCE as a surrogate technique with nonlinear functions, when estimating the stochastic behaviour of non-smooth output responses [25].

The following equation from Santner et al. [47] describes the realisation of a Gaussian process model:

$$\mathcal{M}^K(x) = \beta^\top f(x) + \sigma^2 Z(x, \omega) \quad (3-16)$$

where the first term captures the trend or the mean value of the Gaussian process, and β is the corresponding coefficient of the P arbitrary basis functions $f(x)$. The second term consists of the variance of the Gaussian process σ^2 , and a stationary Gaussian process $Z(x, \omega)$ that has zero mean and unit variance. ω denotes the probability space and is defined by $\mathcal{R} = \mathcal{R}(x_i, x_j; \theta)$ that is a correlation function between two sampled points and is parameterised by a hyperparameter vector θ . Several kinds of Kriging can be classified by, for example, the trend. When the first term in Eq. (3-16) is a constant number, Kriging can be referred to as simple Kriging or ordinary Kriging depending on whether the constant is known or not. Further, universal Kriging is also a type of Kriging as the first term is a combination of prescribed arbitrary functions. Meanwhile, the choice of the correlation functions \mathcal{R} is also crucial to a Kriging model as more detail will follow in the later sections.

The following equations adopt the notation in Lataniotis et al. [31]. Given an experimental design of size N , $\mathcal{X} = \{x^{(1)}, \dots, x^{(N)}\}$ and the corresponding responses $\mathcal{Y} = \{y^{(1)} = \mathcal{M}(x^{(1)}), \dots, y^{(N)} = \mathcal{M}(x^{(N)})\}$, the surrogate model should be able to calculate the prediction

$\mathcal{M}^K(x)$ at a new point x . Once the trend and the correlation function are determined, realisations of the prior Gaussian process can be obtained, as shown in the left of Figure 3-2. The Kriging surrogate model provides the predictions of the model response upon the posterior Gaussian process from the samples as seen in the right of Figure 3-2. The prediction will correspond to a normally distributed variate $\hat{Y}(x)$ with mean $\mu_{\hat{Y}}$ and standard deviation $\sigma_{\hat{Y}}$ as

$$\hat{Y} \sim \mathcal{N}(\mu_{\hat{Y}}(x), \sigma_{\hat{Y}}(x)) \quad (3-17)$$

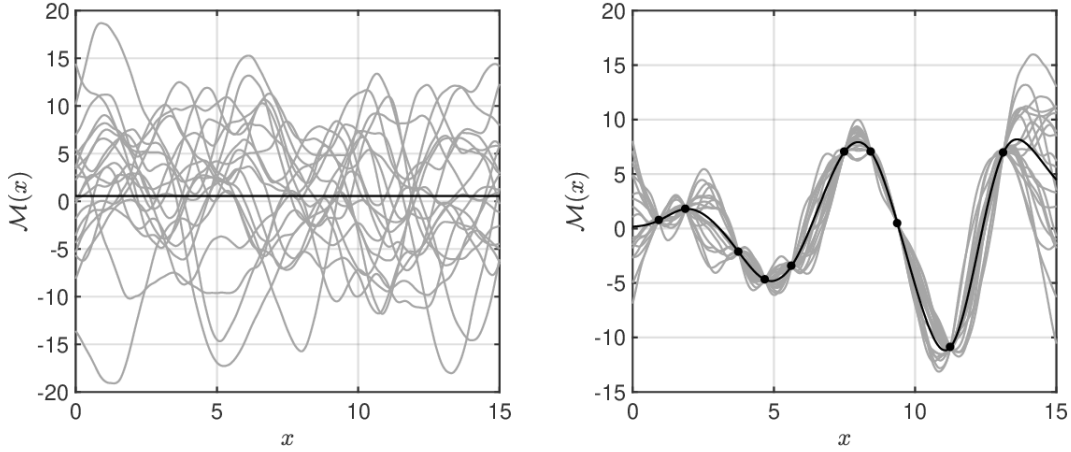


Figure 3-2: An illustration of the prior (left) and posterior Gaussian process (right) [30]. The black lines represent the Gaussian process mean.

The approximation of the computational model is thus the infinite family of the samples of the posterior Gaussian process. For the surrogate purpose, the mean response will be used, and the variance can be a measure of the local prediction error [30]. Based on the Gaussian assumptions, a joint Gaussian distribution, \mathcal{N} , of the observed model response \mathcal{Y} and the prediction at x , $\hat{Y}(x)$, which is determined by

$$\begin{Bmatrix} \hat{Y}(x) \\ \mathcal{Y} \end{Bmatrix} \sim \mathcal{N}_{N+1} \left(\begin{Bmatrix} f^\top(x)\beta \\ F\beta \end{Bmatrix}, \sigma^2 \begin{Bmatrix} 1 & r^\top(x) \\ r(x) & R \end{Bmatrix} \right) \quad (3-18)$$

where the elements are described as follows. First, F is the matrix that consists of observation points for the Kriging trend and it can be expressed as

$$F_{ij} = f_j(x^{(i)}), \quad i = 1, \dots, N, \quad j = 1, \dots, P \quad (3-19)$$

and the expansion is

$$F = [f(x^{(1)}), f(x^{(2)}), \dots, f(x^{(N)})]^\top \quad (3-20)$$

$$= \begin{bmatrix} f_1(x^{(1)}) & f_2(x^{(1)}) & \dots & f_P(x^{(1)}) \\ f_1(x^{(2)}) & f_2(x^{(2)}) & \dots & f_P(x^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x^{(N)}) & f_2(x^{(N)}) & \dots & f_P(x^{(N)}) \end{bmatrix} \quad (3-21)$$

Second, $r_i(x)$ represents the correlation of predicted point x with each design point in \mathcal{X} which is

$$r_i(x) = R(x, x^{(i)}; \theta), \quad i = 1, \dots, N \quad (3-22)$$

To make it easier to explain, the hyper-parameter is rest aside and the expansion reads

$$r(x) = [\mathcal{R}(x, x^{(1)}), \mathcal{R}(x, x^{(2)}), \dots, \mathcal{R}(x, x^{(N)})]^\top \quad (3-23)$$

Third, \mathbf{R} is the correlation matrix of the design points that is given by

$$\mathbf{R} = \begin{bmatrix} \mathcal{R}(x^{(1)}, x^{(1)}) & \mathcal{R}(x^{(1)}, x^{(2)}) & \dots & \mathcal{R}(x^{(1)}, x^{(N)}) \\ \mathcal{R}(x^{(2)}, x^{(1)}) & \mathcal{R}(x^{(2)}, x^{(2)}) & \dots & \mathcal{R}(x^{(2)}, x^{(N)}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{R}(x^{(N)}, x^{(1)}) & \mathcal{R}(x^{(N)}, x^{(2)}) & \dots & \mathcal{R}(x^{(N)}, x^{(N)}) \end{bmatrix} \quad (3-24)$$

Up till now, it is possible to calculate the mean and variance of the Gaussian random variate $\hat{Y}(x)$ using the best linear unbiased predictor (BLUP) [47] properties, which yields

$$\mu_{\hat{Y}} = f(x)^\top \hat{\beta} + r(x)^\top R^{-1}(\mathcal{Y} - F\hat{\beta}) \quad (3-25)$$

$$\sigma_{\hat{Y}}^2(x) = \sigma^2(1 - r^\top(x)R^{-1}r(x) + u^\top(x)(F^\top R^{-1}F)^{-1}u(x)) \quad (3-26)$$

where $\hat{\beta}$ is the generalised least-square estimate of the coefficient β as

$$\hat{\beta} = (F^\top R^{-1}F)^{-1}F^\top R^{-1}\mathcal{Y} \quad (3-27)$$

and

$$u(x) = F^\top R^{-1}r(x) - f(x) \quad (3-28)$$

The above derivation formulates the universal Kriging. Meanwhile, ordinary Kriging mentioned above is a special case of universal Kriging when the trend has a constant but unknown value while simple Kriging is when the trend has a known constant value [62]. In essence, the Kriging model is interpolating all the design points when dealing with noise-free data.

With the theory and derivation of the Kriging model, the following steps are needed to construct a Kriging surrogate model [31]:

1. Select a Kriging trend for the functional basis. This step refers to the mean of the Kriging surrogate model. The step is explained in Section 3-4-2.
2. Select a correlation function $\mathcal{R}(x_i, x_j; \theta)$. With a proper correlation function, the model can thus properly approximate the new points from the design points. Introduction of the step is given in Section 3-4-3.
3. An optimisation needs to be set up to estimate the hyperparameter θ . It can be estimated from the observation data. Details are stated in Section 3-4-4.

Therefore, the following sections give introductions to the first three steps. All others unknown Kriging parameters can be calculated once the optimal value of θ is known. The procedure of calculating these parameters are included in the Section 3-4-4.

3-4-2 Kriging trend

The mean of the Gaussian process can be referred to as the trend. It has been mentioned and discussed in the previous section. Depending on the first term $\beta^\top f(x)$ in Eq. (3-16), different types of Kriging are named which are shown as the following [30]:

- Universal Kriging

Universal Kriging has the most flexibility in terms of formulation. It assumes that the trend is made up of a sum of P arbitrary functions $f_j(x)$ as

$$\beta^\top f(x) = \sum_{j=1}^P \beta_j f_j(x) \quad (3-29)$$

- Simple Kriging

When the trend has a known constant value for a Kriging problem, it can be referred to as simple Kriging. In this case, all the coefficients β are equal to 1.

$$\beta^\top f(x) = \sum_{j=1}^P f_j(x) \quad (3-30)$$

- Ordinary Kriging

In ordinary Kriging, the trend is an unknown but constant value where $P = 1$ and $f_1(x) = 1$.

$$\beta^\top f(x) = \beta_1 f_1(x) = \beta_1 \quad (3-31)$$

3-4-3 Correlation function

Also known as covariance function in the literature, the correlation function \mathcal{R} build up the assumptions for the approximation. It describes the relationship between the observation (i.e., the sample points or experimental design) and the newly predicted points. In other words, it determines the smoothness of the Kriging model[62]. The correlation function \mathcal{R} is usually in the form of $\mathcal{R}(x_i, x_j; \theta)$, where θ is a vector referred to as the characteristic length scale and determines the extent of influence of sampling points.

Correlation family	Expression
Linear	$\mathcal{R}(x_i, x_j; \theta) = \max\left(0, 1 - \frac{ x_i - x_j }{\theta}\right)$
Exponential	$\mathcal{R}(x_i, x_j; \theta) = \exp\left[-\frac{ x_i - x_j }{\theta}\right]$
Gaussian	$\mathcal{R}(x_i, x_j; \theta) = \exp\left[-\frac{1}{2}\left(\frac{ x_i - x_j }{\theta}\right)^2\right]$

Table 3-3: Commonly seen correlation functions and their expressions (adopted from [62]).

As any arbitrary function can not be a valid correlation function, Table 3-3 presents some commonly seen correlation functions. The correlation function needs to satisfy the conditions, such as that the elements need to be positive definite for any number of sampling points N ,

and the function itself needs to be symmetric [31]. All of the shown correlation families are so-called stationary correlation functions as they only depend on the relative position of its two inputs. As an example, a set of one-dimensional inputs, x_i and x_j is used for demonstrating the correlation families as these families are parameterised by θ in Table 3-3. When using different correlation families, the smoothness of the resulting Kriging model changes along with the influence of the neighbouring sampled points.

When the dimension of the input is more than 1, Lataniotis et al. [30] stated it is possible to use the *separable* correlation type such as

$$\mathcal{R}(x_i, x_j; \theta) = \prod_{k=1}^M \mathcal{R}(x_{i,k}, x_{j,k}, \theta_k) \quad (3-32)$$

and the *ellipsoidal* type

$$\mathcal{R}(x_i, x_j; \theta) = \mathcal{R}(h) \quad (3-33)$$

$$h = \sqrt{\sum_{k=1}^M \frac{(x_{i,k} - x_{j,k})^2}{\theta_k}} \quad (3-34)$$

to obtain \mathcal{R} from the one-dimensional correlation families.

3-4-4 Estimation of hyperparameters

In order to build a Kriging surrogate model, the hyperparameters θ are needed, but they are usually unknown and need to be estimated. Setting up an optimisation problem is then required, and β and σ^2 in Eq. (3-16) can be calculated once the hyperparameters are known.

The first estimation method is the maximum-likelihood estimation. To find the Kriging parameters, the maximum-likelihood method can be utilised to find when the likelihood of the observations $\mathcal{Y} = \{\mathcal{M}(x^{(1)}), \dots, \mathcal{M}(x^{(N)})\}$ is maximal. The likelihood function \mathcal{L} reads

$$\mathcal{L}(\mathcal{Y}|\beta, \sigma^2, \theta) = \frac{\det(R)^{-\frac{1}{2}}}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\mathcal{Y} - F\beta)^\top R^{-1}(\mathcal{Y} - F\beta)\right] \quad (3-35)$$

Without diving into the mathematical derivation, the hyperparameters θ can be obtained by solving the following optimisation problem [30] (readers are referred to Santner et al. [47] for proof and more detailed),

$$\theta = \arg \min(-\log \mathcal{L}(\mathcal{Y}|\beta, \sigma^2, \theta)) \quad (3-36)$$

The set of parameters β and σ^2 can thus be calculated where β is referred to Eq. (3-27) and σ^2 is determined by the following equation suggested by Santner et al. [47],

$$\sigma^2 = \sigma^2(\theta) = \frac{1}{N}(\mathcal{Y} - F\beta)^\top R^{-1}(\mathcal{Y} - F\beta) \quad (3-37)$$

The second method is cross-validation, known as the K -fold cross-validation. In general, the method splits all observations, including \mathcal{X} and \mathcal{Y} , into K equal sized subsets [31]. In K -fold cross-validation, the K -th subset is preserved when other $K - 1$ subsets are used to calculate

the Kriging prediction and that K -th set is utilised to evaluate the prediction. The process is then iterated for K times, leading to that every subset is used exactly once to evaluate the prediction. The K results are then averaged for the estimation. Again, without diving into the mathematical derivation, the optimisation problem aims to solve the following objective function [30],

$$\theta = \arg \min \sum_{i=1}^K (\mathcal{M}(x^{(i)}) - \mu_{\hat{Y},(-i)}(x^{(i)}))^2 \quad (3-38)$$

where $\mu_{\hat{Y},(-i)}$ is the mean Kriging predictor. The β can again be determined by Eq. (3-27) and σ^2 is calculated by

$$\sigma^2 = \sigma^2(\theta) = \frac{1}{K} \sum_{i=1}^K \frac{(\mathcal{M}(x^{(i)}) - \mu_{\hat{Y},(-i)}(x^{(i)}))^2}{\sigma_{\hat{Y},(-i)}^2(x^{(i)})} \quad (3-39)$$

A comparison of the surrogate results using the two estimations is included in Appendix B. It is suggested to read through this Chapter first as the comparison is explained in the following sections. For further detailed explanations of maximum-likelihood and cross-validation, readers are referred to Lataniotis et al. [30, 31].

3-5 Verification of the surrogate models

3-5-1 The original model

With the previous theoretical background on the two surrogate models, it is possible to construct surrogate models that are needed for the research. However, it is important to acknowledge the medium-fidelity model that will be emulated by the surrogate models. In this research, the model is named as the *original model* and abbreviates as the **OM** model. The original model consists of the NREL 5MW wind turbine power curve, as shown in Figure 3-3a and the Jensen wake model mentioned in Section 2-3-2.

The original model evaluates the wind farm performance using the recommended interval sizes from Feng and Shen [12], where $\Delta u = 1$ [m/s] and $\Delta \theta = 1^\circ$ [deg]. Even though this interval sizes in wind speed and wind direction could lead to a fine evaluation of the wind farm performance, some research has been looking into the use of surrogate models for this evaluation. The reason that the surrogate models are needed is that the wake model is evaluated in every interval of wind direction resulting in a great computation effort. The total number of the wake model evaluations using the above interval sizes leads to $26 \cdot 360 = 9360$ times, where 26 is the wind speeds interval from 0 to 25 [m/s], and 360 is the wind direction interval from 0 [deg] to 359 [deg]. The evaluation can be reduced to $23 \cdot 360 = 8280$, where the number 23 comes from the interval from the cut-in wind speed to the cut-out wind speed of the wind turbine, which is 3 [m/s] and 25 [m/s], respectively. It is because that the model yields zero power output when the wind speed is lower than the cut-in wind speed or above the cut-out wind speed as the power curve figure shows. It is thus no need to take those condition into the evaluation.

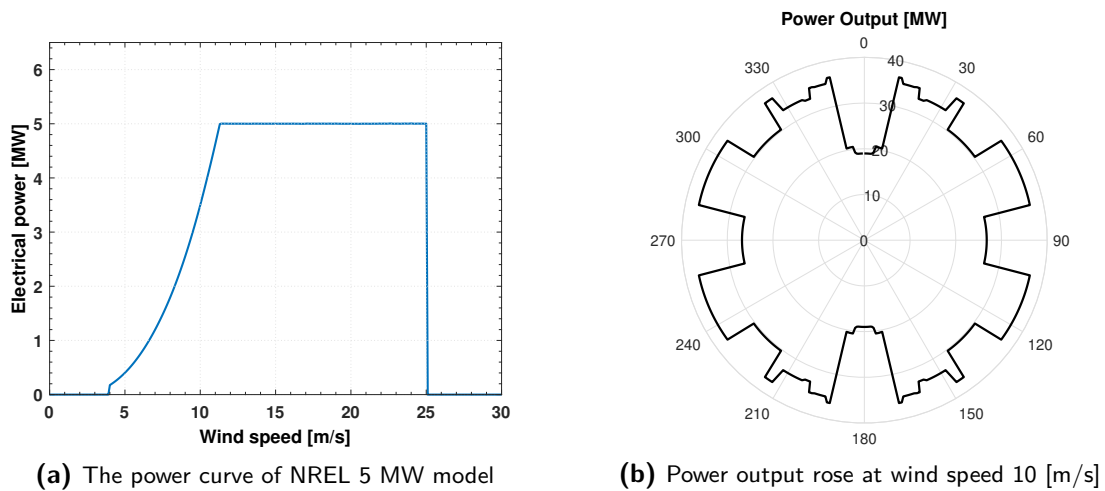


Figure 3-3: The wind turbine power curve and the power output rose of the wind farm under investigation using the original model.

The power output rose when the wind speed is at 10 [m/s] of the investigated wind farm is shown in Figure 3-3b. The wake effect is captured as expected. When the wind direction is 0 [deg] and 180 [deg], the two columns of wind turbines align with the wind direction, causing the wake effect to be most severe. The wake effect also causes the deficits of power output in other wind directions since no other influence factors, such as the availability of the turbines, are taken into account in this research. Lastly, a two-dimensional surface plot of the power production of the layout for every wind speed and wind direction is shown in Figure 3-4. The wake effect within the investigated wind farm can be clearly seen in Figure 3-4 as the power deficits are represented as "dips" in both partial-load and full-load regions. Here, the full-load region is where the wind speed is close to the rated wind speed.

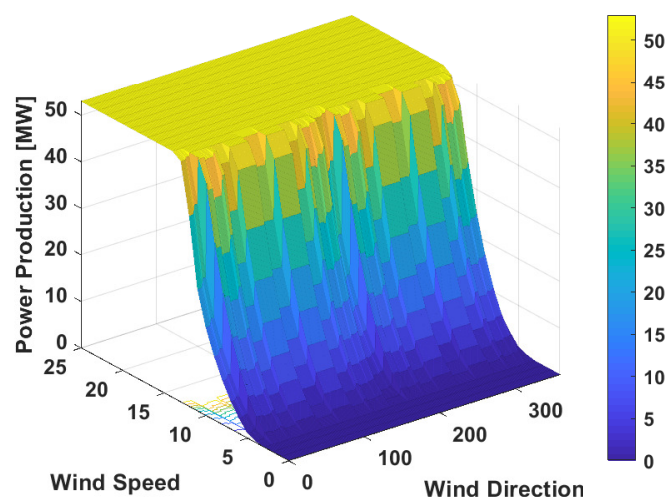


Figure 3-4: The two-dimensional power production surface plot of the layout for every wind speed and wind direction using the original model.

It is also possible to calculate the AEP with the original model using the power output. The AEP is calculated with the wind rose in Figure 2-2 while dividing the wind direction into 12 sectors (i.e. bins) and taking the mean sector probabilities from the 50 years measurement. The Weibull scale and shape parameters are obtained in a similar fashion of using the mean value of the 50 years measurement. Finally, Eq. (2-3) can be used to obtain the AEP.

To verify the original model used in this research is correct, Wind Atlas Analysis and Application Program (WAsP) by Risø, is utilised. WAsP is commonly utilised to perform wind resource assessment, wind farm micro-siting, and for the calculation of wind turbine and wind farm energy yield. The wake effect model implemented in WAsP is approximately the same model used in the current research, the Jensen model (aka Park model), from Katic et al. [24], but with a slight difference in the parameters. For a WAsP analysis to work, users are required to provide wind turbine parameters, wind turbine locations, wind climate data, and terrain analysis. Ten NREL 5 MW wind turbines are again used in this case. The layout of the wind turbine has been shown in Figure 2-5 with 800 meters distance to the closest turbine. The wind climate data is derived from the Global Wind Atlas ¹ website. However, it should be noted that the weather data from KNMI are measured from a meteorological mast of 10 meters high while the Global Wind Atlas translates the data from 100 meters to the hub-height of 90 meters. Due to lack of data, the terrain analysis data used in the research comes from the example data in WAsP, which present the analysis of a Danish offshore wind farm, Tunø offshore wind farm. Since the investigated wind farm is also an offshore wind farm, the study continues with the existing data.

The comparison of the results from the OM model and the WAsP software is presented in Table 3-4. The gross AEP is the value without the wake effect. The OM model directly gives the net AEP where the deficit from the wake effect has been considered. The differences in the net AEP could be explained with the wind climate data source as the original model uses the measurement data from the KNMI mast, and WAsP uses the re-analysis data from Global Wind Atlas. Nevertheless, the results are approximately close to each other, and therefore, the OM model is verified. An extensive WAsP wind farm report is placed in Appendix C for the reader who is interested in the results.

	Original Model	WAsP
Total Gross AEP [MWh]	-	$2.2700 \cdot 10^5$
Total Net AEP [MWh]	$2.2348 \cdot 10^5$	$2.1484 \cdot 10^5$
Capacity Factor [%]	48.17	46.3

Table 3-4: Comparison between the original model and values from the WAsP software on the wind farm performance.

For the purpose of this work, UQLab [33], a MATLAB Toolbox is used to perform surrogate modelling. The toolbox is developed at the Chair of Risk, Safety and Uncertainty Quantification of Swiss Federal Institute of Technology in Zurich (ETH). It is free to academic researchers with non-commercial usage. UQLab provides a framework for uncertainty quantification. It features the ability to carry out uncertainty propagation through Monte Carlo

¹Global Wind Atlas 3.0, a free, web-based application developed, owned and operated by the Technical University of Denmark (DTU). The Global Wind Atlas 3.0 is released in partnership with the World Bank Group, utilising data provided by Vortex, using funding provided by the Energy Sector Management Assistance Program (ESMAP). For additional information: <https://globalwindatlas.info>

simulation (MC), building surrogate models (e.g., polynomial chaos expansions, Kriging), sensitivity analysis, and more.

The following sections present the verification steps of the surrogate model using PCE and Kriging model. For the purpose of evaluating the quality of the surrogate model, it is important to define a measuring method. The percentage error, ϵ , is used in this research and is shown in Eq. (3-40) to measure how accurate the surrogate model is.

$$\epsilon [\%] = \frac{|AEP_{\hat{M}} - AEP_{OM}|}{AEP_{OM}} \cdot 100 \quad (3-40)$$

3-5-2 Verification of the PCE model

After the OM model is tested and verified, the research moves on to the verification of the PCE model. Here, PCE serves as a surrogate technique for power output. First, the random variables of interest need to be defined. For a wind farm model, the wind speed and wind direction will be the choices. The first verification is conducted to the wind speed surrogate model for the PV curve of the turbine. The second one is about the wind direction surrogate model for the wind farm power output. A two-dimensional surrogate model that considers both wind speed and wind direction follows after the two previous models.

After the surrogate model is built, a validation set for the surrogate models is implemented, which varies per cases. Wind speed surrogate model uses an interval of 1 [m/s] from 0 to 25 [m/s] as the validation set. In contrast, the wind direction surrogate model utilises wind direction with an interval of 1° from 0° to 359°. Similarly, the two-dimensional surrogate model also has a validation set that takes the full interval size of wind speed and wind direction.

Surrogate on wind speed

The first case under investigation is the surrogate model with wind speed, which affects the wind turbine power production. The power curve of the wind turbine is thus the original model upon which the surrogate model is constructed. Considering a case with a single wind turbine and using the quadrature method, Figure 3-5a demonstrates the results of the surrogate models on the NREL 5MW wind turbine power curve.

It can be seen with the zeroth-order expansion that the surrogate does not represent the real model well. As the order increases, the surrogate starts to become a better fit. Figure 3-5b further supports this finding, where the performance of the surrogate model becomes closer to the true response. Nevertheless, since the power curve is not a polynomial function, it is hardly possible to have an exact surrogate model. This is also in agreement with the work of Bailleul [3]. The difference [%] of the AEP between the estimation by PCE and the PV model is defined by using the surrogate model in the AEP calculation as shown in Eq. (2-1). Similar to what has been seen in fitting the PV model, the differences in AEP decreases as the order of quadrature increases. It eventually converges to a percentage error around 1% when the order is more than 6, which is considered a sufficient approximation.

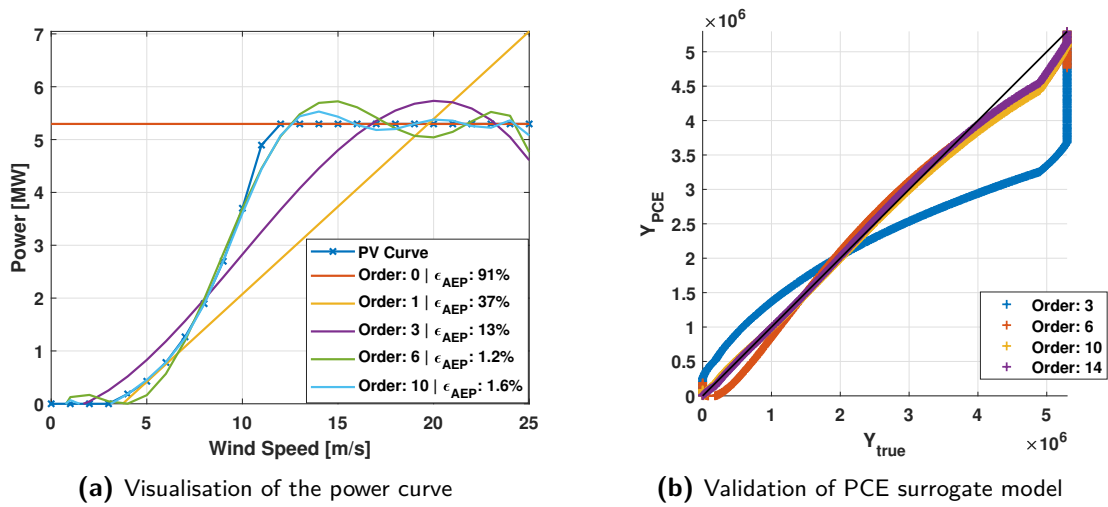


Figure 3-5: Different order of PCE quadrature models on the NREL 5MW turbine PV curve

Surrogate on wind direction

The second case is the investigation of the surrogate model with the wind direction. This case is important when there are multiple wind turbines in a wind farm, causing the wake effect to take place. The wind speed is fixed at 10 [m/s], which is the wind speed in between the cut-in and rated wind speed of the wind turbine, and thus the wake effect can be clearly presented.

As mentioned in Section 3-3-2, there are several ways of building up a PCE model. Figure 3-6 and Figure 3-7 show the investigation of two different sampling schemes with a different order of polynomial basis and number of sampling points using different methods. The sampling points are also called the collocation points, where the sampled wind directions are used as the experimental design that evaluates the true power output. Before introducing the two sampling scheme, it is worthy to note that the quadrature method depends on the order of the polynomial basis to determine the number of sampling points, which follows the Gaussian quadrature rule. The sample points for the quadrature method is, therefore, cannot be chosen by the researcher. For OLS and LARS, the first scheme is random sampling, which randomly sampled the wind direction from $[0^\circ \ 360^\circ]$. The random sampling scheme can be based on Monte Carlo sampling or Latin hypercube sampling (LHS), and the research continues with the Latin hypercube sampling scheme. The second scheme is linearly spaced sampling (LSS) where the samples are drawn from $[0^\circ \ 360^\circ]$ with linearly spaced interval.

The difference [%] of the 10 [m/s] power output between the estimation by PCE and the OM model is defined by making the wind direction into 12 sectors (e.g. bins) and using the mean probability of each sector from the 50 years measurement. It also considers the number of sampling points in wind direction. The sector probability is divided by the number of wind directions that are sampled within that sector. With these probabilities, the power yield is calculated, assuming that the wind speed is 10 [m/s] throughout the entire year. In addition, the power yield of the original model is calculated with the interval of one degree for the wind direction from $[0^\circ \ 360^\circ]$, and the PCE model is validated with the same intervals.

Figure 3-6 compares the random sampling scheme, Latin hypercube sampling, to the linearly

spaced sampling with a different order of polynomial basis using different PCE methods. The linearly spaced sampling seems to have a more stable convergence rate on accuracy. Also, based on the convergence plots in Figure 3-7, it is concluded that using the linearly spaced sampling could possibly lead to an earlier convergence for OLS and LARS methods. Under such sampling theme, the order of polynomial basis seems to have little effect on the difference of power output between the surrogate model and the original model. On the other hands, the number of sampling points indicates the accuracy of less than 1% can be achieved when more than 50 points are used for the polynomial order higher than one. The linearly spaced sampling scheme is, therefore, selected to continue the construction of the surrogate model.

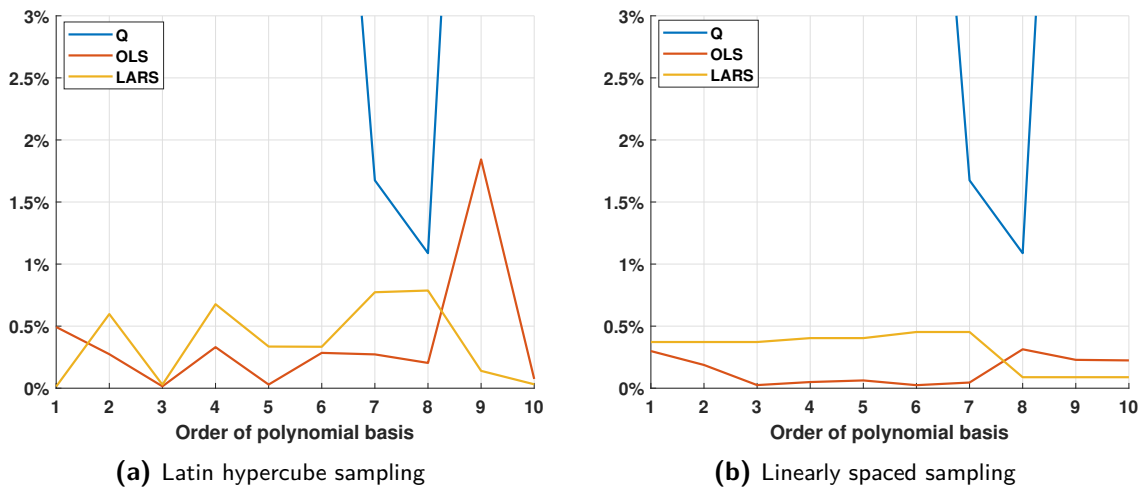


Figure 3-6: The difference [%] of power output between the estimation by PCE and the original model for various polynomial basis order with different PCE methods at 10 [m/s] wind speed.

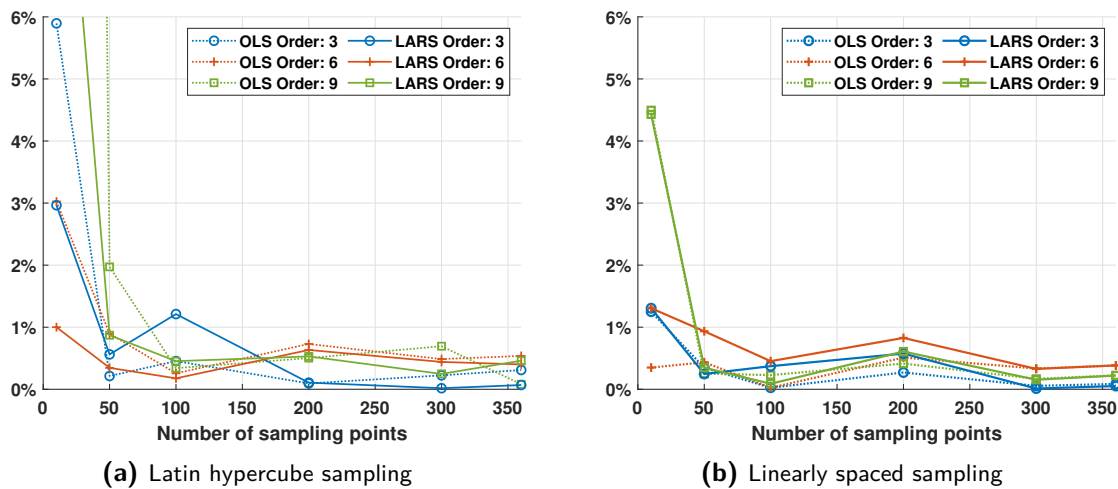


Figure 3-7: The difference [%] of power output between the estimation by PCE and the original model for varying number of sampling points on wind direction with different orders at 10 [m/s] wind speed.

Finally, Figure 3-8 presents the results of the PCE surrogate model in terms of power roses

using different polynomial basis orders. The surrogate was constructed by 100 wind directions with linearly spaced samples and using the OLS method to calculate the polynomial chaos coefficients. The samples are also called collocation points, where the true function is evaluated.

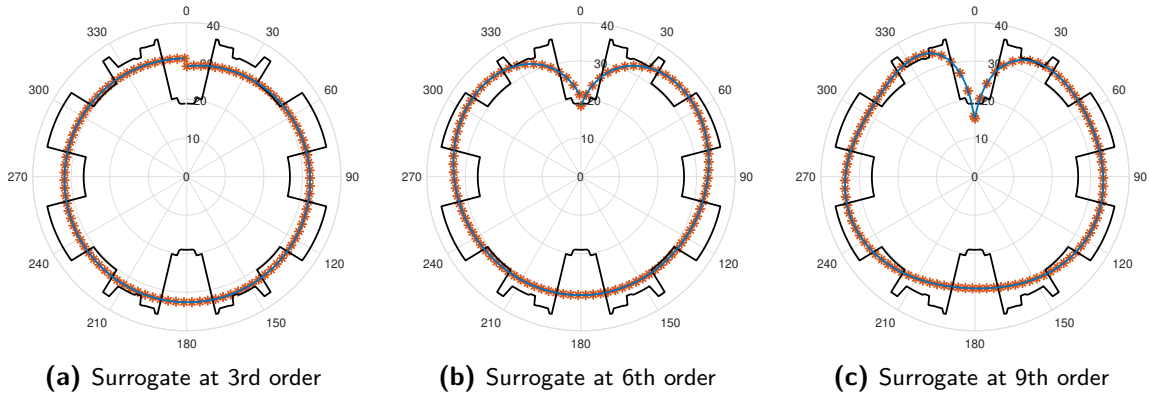


Figure 3-8: PCE power output surrogate models on wind direction at 10 [m/s] wind speed. The black line is the power output from the original model, the blue line comes from the constructed surrogate model, and the markers * denote the collocation points.

As clearly seen, none of the orders is a representative fit. Even with a higher polynomial expansion degree, the surrogate models can only catch the large wake effect in the 0 degree but failed in the 180 degrees. Moreover, with a higher polynomial expansion degree, the surrogate models start to yield discontinuity at 0° or 360° . It has been identified in some literature [3, 42] that the surrogate models work better on wind speed than on wind direction. This might be due to the polynomial being better at representing the PV curve than the power rose of a wind farm as the power rose is more erratic. Nevertheless, the results in Figure 3-8 yield considerable accuracy on the power output with respect to the original model where the difference with the 3rd order is 0.0022%, the 6th order is 0.0739%, and the 9th order is 0.4033%.

Surrogate on wind speed and wind direction

Now, both surrogate models on wind speed and wind direction have been built. The construction of a two-dimensional surrogate model hence follows. It is called a two-dimensional model because the two inputs, wind speed and wind direction are both sampled to calculate the wind farm power output. To showcase the three different PCE methods, Figure 3-9 is created.

The quadrature method depends on the order of the polynomial basis to determine the number of sampling points, which follows the Gaussian quadrature rule. Therefore, the order is fixed at 15th to have a similar amount of collocation points to compared with the regression methods. The result using the ordinary least square method shown in Figure 3-9b and the one utilising the least angle regression method is shown in Figure 3-9c. The methods both use 9th order polynomial basis with 200 samples in wind speed and wind direction, shown as the red markers in the figures. It is worth to note that the power surface plots are plotted using the validation set that is every wind speed and wind direction at integer intervals. From the power surface

plot of the three methods, it is suspected that PCE is not able to catch the power deficit from the wake effect that well as one could hardly see dips in the partial-load and full-load region as Figure 3-4 has shown. The reason behind this could possibly be understood from the one-dimensional results as the PCE model can not well capture the wake effect.

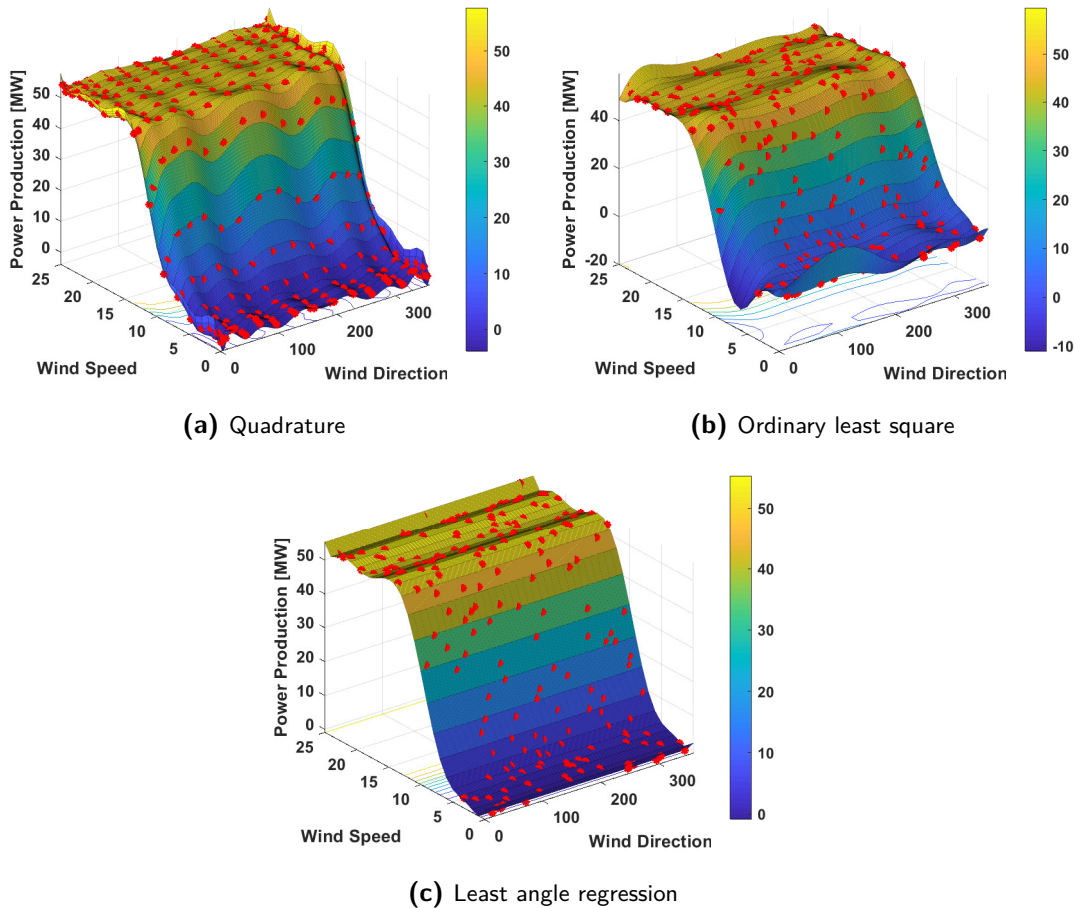


Figure 3-9: The two-dimensional power production surface plots of the investigating layout for every wind speed and wind direction using PCE models. The red markers denote the collocation points.

An investigation of the accuracy using different methods and sampling points is also conducted. The result is presented in Figure 3-10. In the figure, PCE using quadrature method is indicated as PCE-Q, PCE using least square regression is denoted as PCE-OLS, and PCE using least-angle regression is PCE-LARS. The difference [%] of the AEP between the estimate by PCE and the OM model is again defined by making the wind direction into 12 sectors and using the mean probability of each sector from the 50 years measurement. The same goes to the Weibull parameters. Based on the result, both regression methods converge to a certain level when the number of samples becomes more than 200. The difference is that PCE-LARS performs better even with smaller sample size, and this might be because least-angle regression is a sparse PCE, which is a more advanced and efficient method that is able to select the polynomial basis that has the most significant impact on the model response [5].

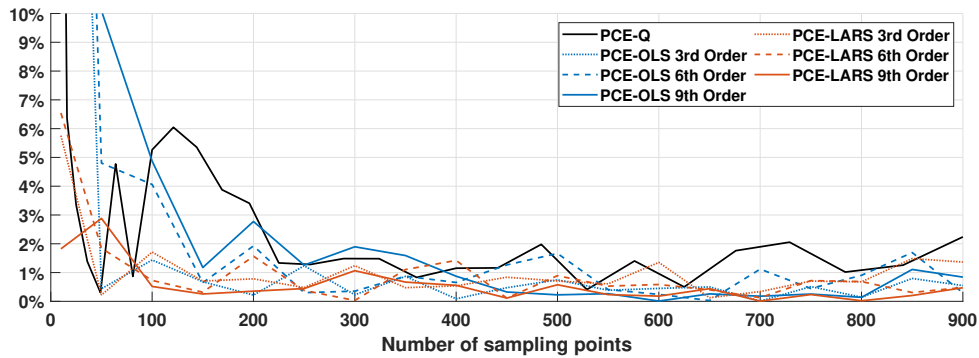


Figure 3-10: The difference [%] of power output between the estimation by PCE and the original model for varying number of collocation points on wind speed and wind direction.

A further detailed investigation is conducted to see the effect of taking a different number of sampling points in wind speed and wind direction. To assign different sampling points in the two dimensions, the linearly spaced sampling scheme needs to be used. The results of difference in AEP approximation is shown in the contour plots of Figure 3-11a and its detailed view of Figure 3-11b. Based on the results, it can be concluded that the number of samples in wind speed is more important than the number of samples in wind direction. For example, an accuracy of 2% can be reached when the number of collocation points for wind speed is more than 6, whereas that is 10 for the wind direction. The surrogate model was constructed using a polynomial basis of 9th order with the LARS method.

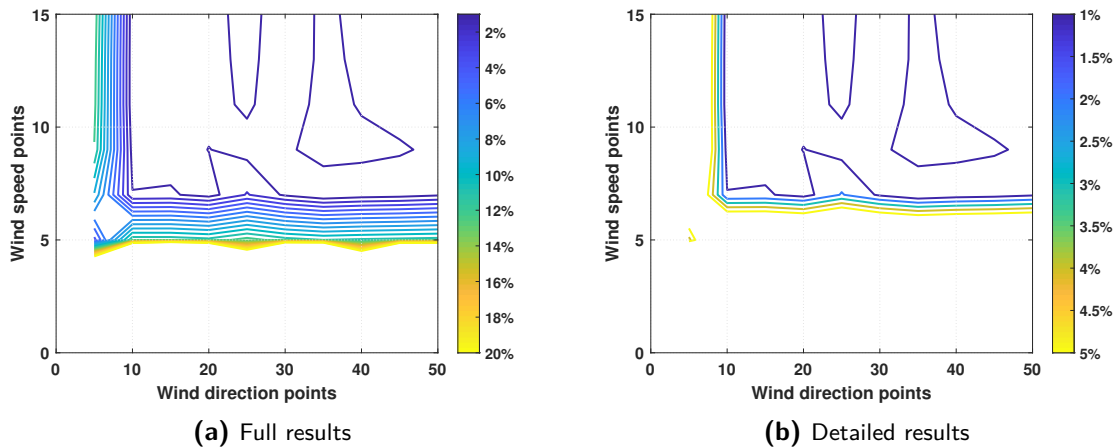


Figure 3-11: The difference [%] of the AEP approximation between the surrogate model and the original model using different collocation points in wind speed and wind direction.

3-5-3 Verification of the Kriging model

Next to the PCE surrogate model, the verification of the Kriging surrogate model follows. In a similar manner, the wind speed surrogate model is investigated first and then the wind direction one. Once the two models are constructed, a two-dimensional Kriging surrogate model is made. The verification also consists of studying different Kriging setting and number of samples used to build the surrogate model. The same validation set with the one used for the PCE model is also implemented once the surrogate model is built. In general, the validation set takes every wind speed and wind direction into account.

Surrogate on wind speed

The Kriging model is also utilised to create the PV curve surrogate model. Visualisation of the surrogate model is made and can be seen in Figure 3-12. Based on the number of sampling points (i.e., experimental design) used to construct the model, the accuracy of the approximation varies. Moreover, the Kriging model can also provide information regarding the variance of the approximation. In Figure 3-12, the blue lines denote the mean value of the Kriging model, and the grey areas show the 95% confidence interval based on the variance. As a demonstration, both models are built with universal Kriging of 3-degree polynomial, exponential correlation function, and use the cross-validation method to estimate the hyperparameters. The difference in AEP is determined in the same way as mentioned in the PCE verification. According to the results, the Kriging model is in good agreement with the original PV model when using 15 samples as using this surrogate PV model can reach a difference with the original model within 1%.

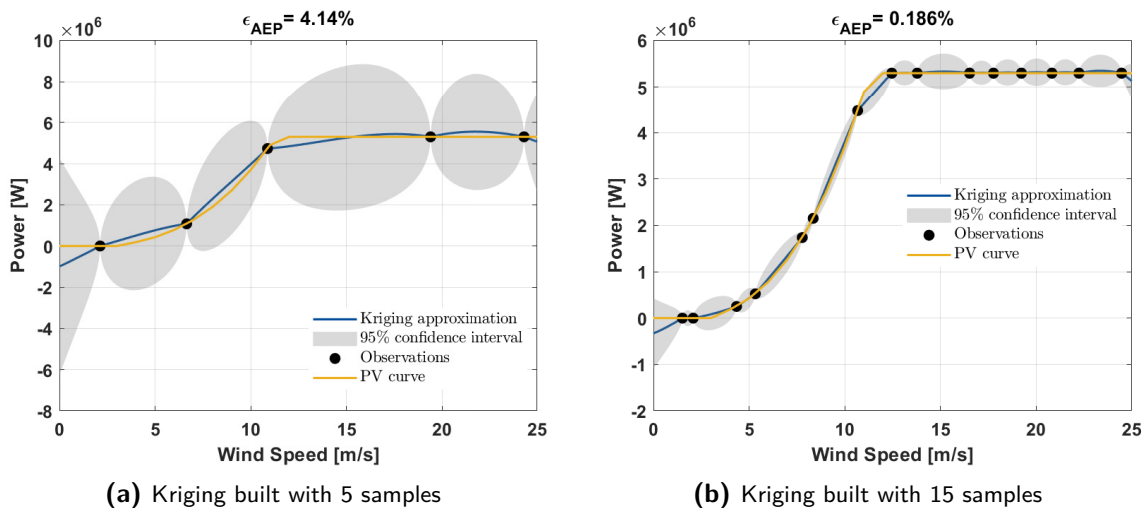


Figure 3-12: The Kriging PV surrogate model of the NREL 5MW wind turbine

Surrogate on wind direction

The second investigating case for Kriging is also the surrogate models based on wind direction. The wind speed is again fixed at 10 [m/s] to ensure that the wake effect is clearly presented. Similar to the PCE modelling, the two sampling schemes are tested to see the effect on the power output differences. The results are shown in Figure 3-13. The results are similar to the ones shown before, as the linearly spaced sampling performs slightly better in terms of the rate of convergence. It is not expected that this sampling scheme turns out to have such similar results with a different order of universal Kriging trends, especially when the number of sampling points on wind direction is more than 50. In general, the differences in AEP from the OM model and the Kriging model reach less than 2.5% when the number of sampling points is more than 50. Regarding the method used to construct the Kriging surrogate model on the wind direction, the results can hardly tell which method performs the best as the differences are relatively small as clearly seen in Figure 3-13b.

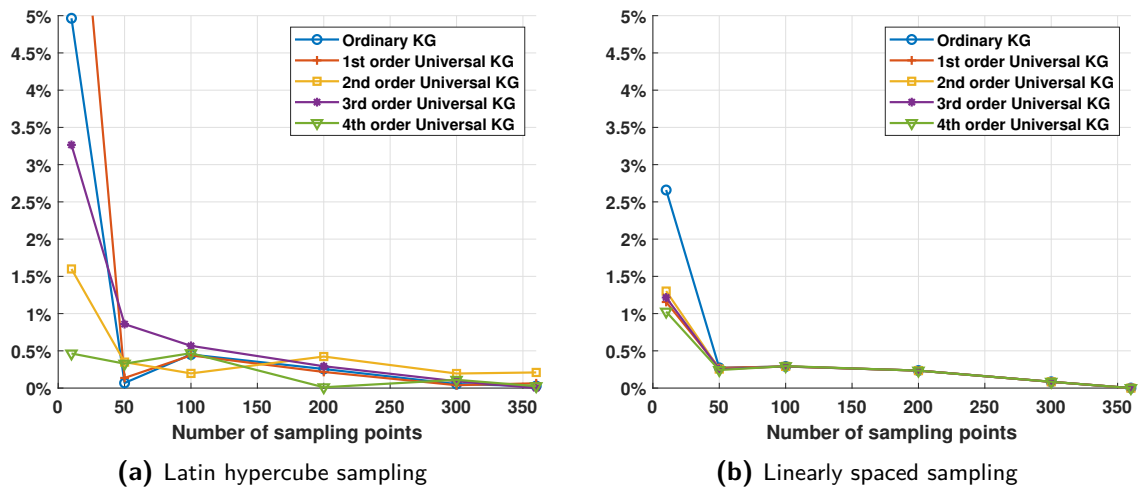


Figure 3-13: The difference [%] of the power output between the estimation by Kriging and the original model for varying number of sampling points on wind direction with different Kriging methods at 10 [m/s] wind speed.

Based on the convergence plots, it is concluded that using the linearly spaced sampling with 4th order universal Kriging results in a reasonable accuracy of power output approximation. Figure 3-14 shows the power roses with different sample numbers for the Kriging surrogate and plotted with equally spaced points from $[0^\circ, 360^\circ]$ degree. According to the power roses, it can be said that the Kriging model performs better than PCE models when it comes to fitting the wind direction as it catches the wake effect better even with small amount of sampling points. The differences between the power yield from the Kriging model and the original model for the three power output roses is 0.2139%, 0.3604%, and 0.0217%, respectively. Nevertheless, it should also be noted that the construction of Kriging is much more complicated and computationally expensive than the construction of the PCE model. Many parameters need to be defined, and the hyperparameter itself is an optimisation problem that requires several iterations. A comparison of the time requirement of the two surrogate methods will be carried out in Section 3-6.

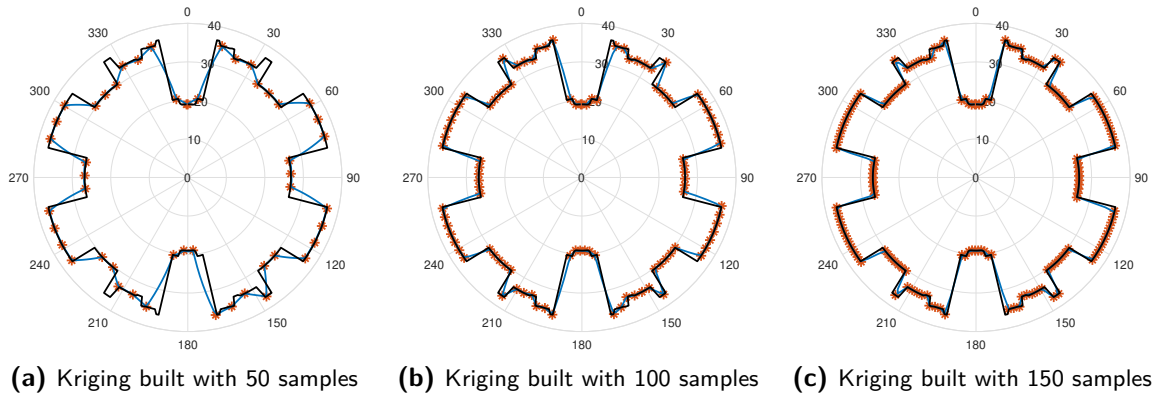


Figure 3-14: Kriging power output surrogate models on wind direction at 10 [m/s] wind speed. The black line is the power output from the original model, the blue line comes from the constructed surrogate model, and the markers * denote the collocation points.

Surrogate on wind speed and wind direction

Now, surrogate models are created for the power curve and the wake evaluation using the Kriging model. Wind speed and wind direction are now both considered to build a two-dimensional surrogate model. To showcase the Kriging surrogate models, Figure 3-15 is created. All construction methods of the Kriging surrogate model use 200 random samples (shown as the red markers in the figures) to create the power output surface plots. After the construction of the models, the validation set is used to evaluate the power output using every wind speed and wind direction at integer intervals. The ordinary Kriging presented in Figure 3-15a reaches a difference between the surrogate model and the original model of 0.3081%. The 2nd order, 3rd order, and 4th order universal Kriging displayed in Figure 3-15b to Figure 3-15d reach 0.1945%, 0.5179%, and 0.5666%, respectively. Based on the surface plots of the power output, it is possible to conclude that Kriging surrogate model captures the power deficit from the wake effect much better than the PCE model as the "dips" can be clearly seen in regions like wind direction at 180 [deg].

The construction of the surrogate model requires an investigation of the methods to build a Kriging model, and the number of samples needed to achieve a certain accuracy. The selection of Kriging settings is based on the investigation using different construction methods shown in Figure 3-16. It is seen from the result that all settings converge to roughly the same difference as the sampling numbers reach 300. However, the 2nd order universal Kriging outperforms other methods in lower sampling points in wind speed and wind direction.

A further detailed investigation is conducted to see the effect of taking a different number of sampling points in wind speed and wind direction. The surrogate model was constructed using 2nd order universal Kriging with linearly spaced sampling scheme. This parameter configuration is obtained from the convergence plot in Figure 3-16. The result of difference in AEP approximation is shown in the contour plots of Figure 3-17a and its detailed view of Figure 3-17b. Similar to the PCE case, the number of sampling points in the wind speed is more important than the one in the wind direction. According to the results, it can be seen that accuracy of 5% can be reached when the number of collocation points for wind speed is more than 5, whereas this is 10 for the wind direction.

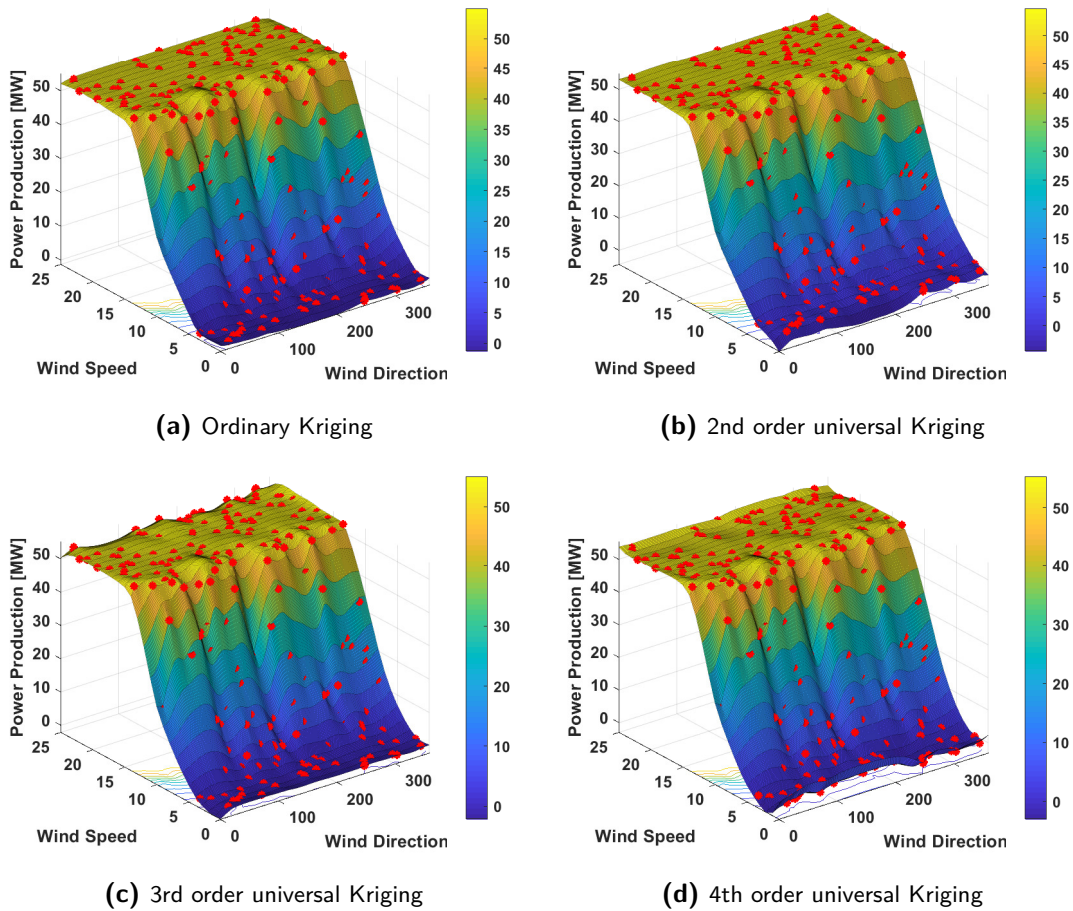


Figure 3-15: The two-dimensional power production surface plot of the investigating layout for every wind speed and wind direction using the Kriging model. The red markers denote the collocation points.

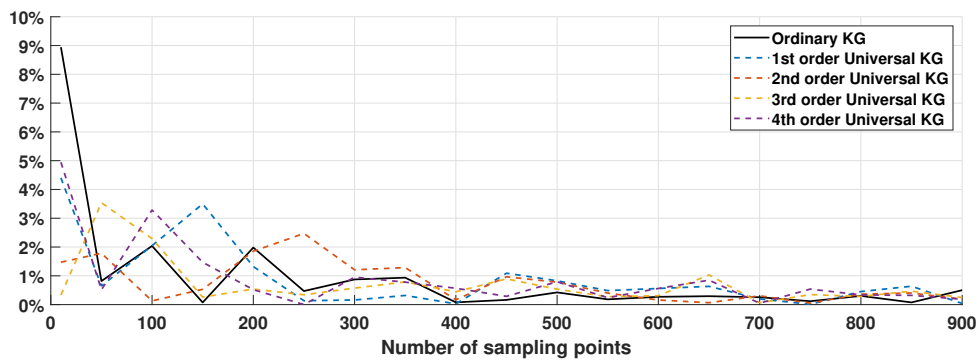


Figure 3-16: The difference [%] of the AEP between the estimation by Kriging and the original model for varying number of collocation points on wind speed and wind direction.

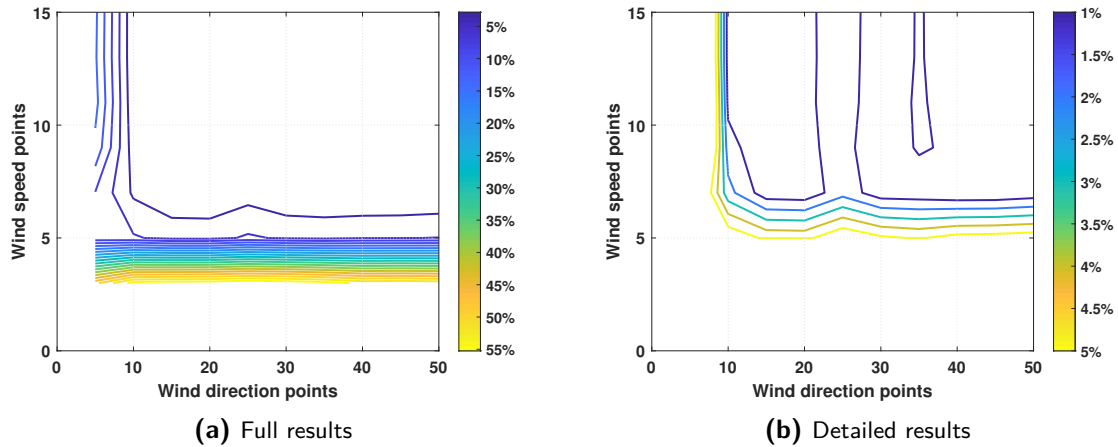


Figure 3-17: The difference [%] of the AEP approximation in between the surrogate model and the original model using different collocation points in wind speed and wind direction.

3-6 Comparison of the surrogate models

Two different surrogate models with different techniques have been constructed to provide a faster evaluation for the wind farm power output. This section aims to provide a comparison of the two surrogate models and establish the optimal parameters for the models. Based on the previous verification, the polynomial basis order for the PCE model and the universal Kriging order has been determined as 28th and second order, respectively. The configurations are results from several attempts that look for the best estimation.

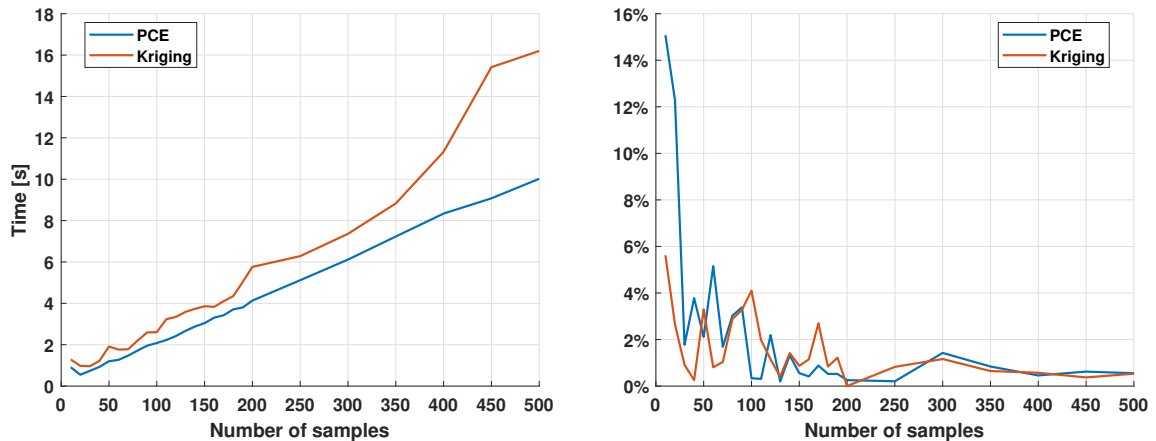


Figure 3-18: Comparison of time required and accuracy for the two surrogate models.

Therefore, with the fixed orders, the first comparison is on time required to build the surrogate model. Since the computational time depends on the hardware used, the relative differences are more crucial than the absolute time needed. For an OM model, it requires 215.2038 seconds for MATLAB to complete the 9360 evaluations of the wind farm power out-

put. Figure 3-18a shows the time required by the surrogate models with a different number of samples. It is noteworthy that the random sampling scheme is used here to have the freedom of varying sample size. Based on the result, Kriging seems to require a little bit longer time than the PCE model. Second, the effect of the number of samples on the accuracy of the models compared to the OM model is investigated. The result is shown in Figure 3-18b and is consistent with what has been shown in the previous sections. The Kriging model is slightly more accurate than the PCE model despite that they both are able to have accuracy within 2% when more than 200 samples are used.

The optimal setting for the two models after the verification is presented in Table 3-5. Both models have an experimental design of 200 samples. As mentioned before, the PCE model uses the LARS method with 28th order of polynomial basis while the Kriging model uses the 2nd order universal Kriging.

	Number of samples	Method	Settings
PCE	200	LARS	28th Order
Kriging	200	2nd Degree Universal Kriging	Exponential Correlation Family, Cross-Validation Estimation

Table 3-5: The parameters and settings used to construct the optimum surrogate models.

The results are visualised in Figure 3-19. The 200 sampling points are further categorised by how they are generated, either by Latin Hypercube sampling or using linearly spaced samples. From Figure 3-19a and Figure 3-19b, the difference between the estimation by PCE model and the original model is considered small but the power output surface plots vary a lot. The error of the models is shown in Table 3-6. When using a random sampling scheme for the collocation points, there are discontinuities at the boundary of wind speed and wind direction sampling ranges. On the other hands, PCE model using linear sampling scheme performs better but fails to capture the power deficit caused by the wake effect. Nevertheless, the differences between both PCE models to the original model on the AEP value are within 1%, which is not expected by looking at the figures at first sight, especially for the one with the random sampling scheme. The reason maybe because of the PCE model tends to smooth out the power deficits caused by the wake effect as already seen in Figure 3-8. When using linearly spaced samples as in Figure 3-19b, the sampling scheme only sampled certain wind speed and wind direction, which may cause the model to be biased toward those sampled points and averaging out the power yield. In contrast, using Latin hypercube samples, the model is less biased by the sampled points as every wind speed, and wind direction are all likely to be sampled to evaluate the true power output.

Meanwhile, both Kriging models are able to capture the wake effects as the dips in the surface plots are easy to spot. Even though the Kriging model with linearly spaced samples seems to be a better representation of the original model when looking at the figure, the difference of it to the original model is surprisingly higher compared to the case where random samples are used. The reason behind it could be similar to what has been discussed for the PCE model. Such that as long as there are an adequate amount of sampling points, using the random sampling scheme (i.e., Latin hypercube sampling) could lead to a better result. It is because that all wind speed and wind direction are possible to be sampled for the true evaluation.

Lastly. a comparison of the time required to build and evaluate the surrogate models men-

tioned in the figure is presented in Table 3-7. As can be concluded from the table, the time difference of the two sampling schemes for the surrogate models does not show a large difference.

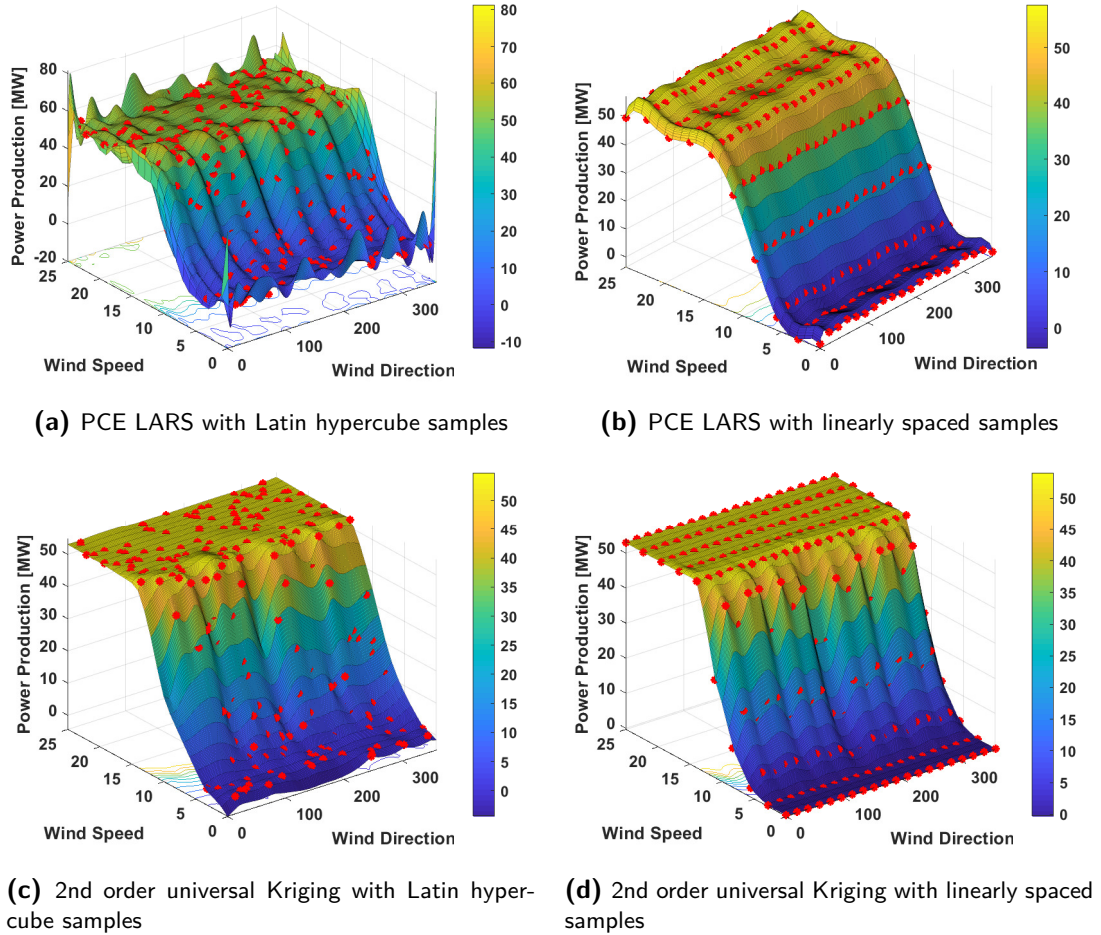


Figure 3-19: The two-dimensional power production surface plot of the investigating layout for every wind speed and wind direction using the two surrogate models with different sampling schemes.

	AEP [MWh]	CF [%]	Time [s]	Error [%]
Original Model	$2.2348 \cdot 10^5$	48.17	215.2038	-
PCE LHS Model	$2.2379 \cdot 10^5$	48.24	4.8327	0.1365
PCE LSS Model	$2.2407 \cdot 10^5$	48.30	4.7239	0.2636
Kriging LHS Model	$2.2350 \cdot 10^5$	48.18	6.4206	0.0089
Kriging LSS Model	$2.2275 \cdot 10^5$	48.01	6.4719	0.3295

Table 3-6: Comparison of OM model and the surrogate models on AEP and the time required to construct the models.

To sum up, it is suggested to use random sampling schemes for the surrogate models when one is mainly concerned about the AEP value but not about the wake effect being fully

represented. Nevertheless, both PCE and Kriging model are considered accurate surrogate models to the OM model. The Kriging model outperforms the PCE model regarding the ability to catch the wake effect but fails on the time required to build the model. In the current research on the surrogate model for power output, it is seen that PCE is a better method to construct surrogate models than Kriging as it can give a better accuracy compared to the original model. Furthermore, regarding the procedure of setting the surrogate models, PCE is far more straightforward. The overall complexity of setting up a Kriging model and the computational time is more intense than for other methods.

	PCE LHS	PCE LSS	Kriging LHS	Kriging LSS
Model building (s)	4.7976	4.6820	6.2808	6.3267
Evaluation (s)	0.0351	0.0419	0.1398	0.1452
Total time (s)	4.8327	4.7239	6.4206	6.4719

Table 3-7: The computational time required to build and evaluate each surrogate model.

3-7 Summary of the chapter

This chapter explains and illustrates the surrogate modelling techniques, namely polynomial chaos expansion and Kriging. The surrogate models aim to emulate the expensive computational procedure of estimating the wind farm power output. Both surrogate models reach satisfactory accuracy, where the percentage error of AEP lies within 2% when using an adequate amount of samples. Kriging has shown a slight advantage in capturing the wake effect over PCE. Nevertheless, it requires additional computational effort in terms of time compared to the other methods, and the accuracy is not higher than PCE under the same amount of samples. Overall, PCE can be considered a better method when considering both evaluation effort and accuracy.

Effectiveness of Surrogate Models

4-1 Introduction to the chapter

From Chapter 3, it has been proven that surrogate modelling, both PCE and Kriging, can be a good replacement of a time-consuming wind farm power model. However, the investigation was built on a specific layout that is a gridded and aligned one. The influence of different layouts is still not known, and this chapter, therefore, focuses on the effectiveness of the surrogate model on different wind farm layouts. Furthermore, it has also been found that when using a random sampling scheme (e.g., Latin hypercube sampling) to sample the collocation points, the surrogate models produce different results due to the stochastic characteristic of the scheme. Hence, Section 4-3 is dedicated to study and quantify the randomness from the sampling scheme.

4-2 Effectiveness on different wind farm layouts

4-2-1 Overview of wind farm layouts

Besides the wind farm layout mentioned in Figure 2-5, the research further looks into six additional wind farm layouts. Here, the six additional wind farm layouts are named as case layout from 2 to 6 while case 1 is the layout that has been investigated in Chapter 3, namely the one shown in Figure 2-5. The number of wind turbines is still fixed to 10. From Chapter 3, it has been acknowledged that using the original model to calculate the energy yield is time-consuming, so it is possible to use a surrogate model to improve the calculation speed of a wind farm layout optimisation problem. The main goal is to study if the surrogate model that has been constructed only works properly on a specific type of wind farm layout. This study is useful if the surrogate model is implemented in wind farm layout optimisation to reduce the computational effort. As many layouts are evaluated while the optimiser looks for a layout with a maximum wind farm energy output, it is essential to know the limit of the surrogate models.

The wind farm layout for case 2 to 7 is shown in Figure 4-1 while case 1 is the layout mentioned in Chapter 2 and Chapter 3. These cases are determined by the author based on three different aspects. The first one is changing aligned layouts into staggered ones. The two layouts, in case 2 and case 3, are determined based on this aspect. The reason for the staggered layout is to reduce the wake effect. Instead of an aligned layout, Figure 4-1a and Figure 4-1b shift the rear wind turbines away from the front rows to avoid wake effect. The second aspect is creating as many wake effect as possible while maintaining the area constraints and safety distance. This aspect can be seen in Figure 4-1c and Figure 4-1d, where the turbines are placed close to each other. The third aspect is changing the regular grid layout to irregular or random ones. Figure 4-1e shows an L-shape layout, while Figure 4-1f is randomly generated to mimic the progress within a wind farm layout optimisation.

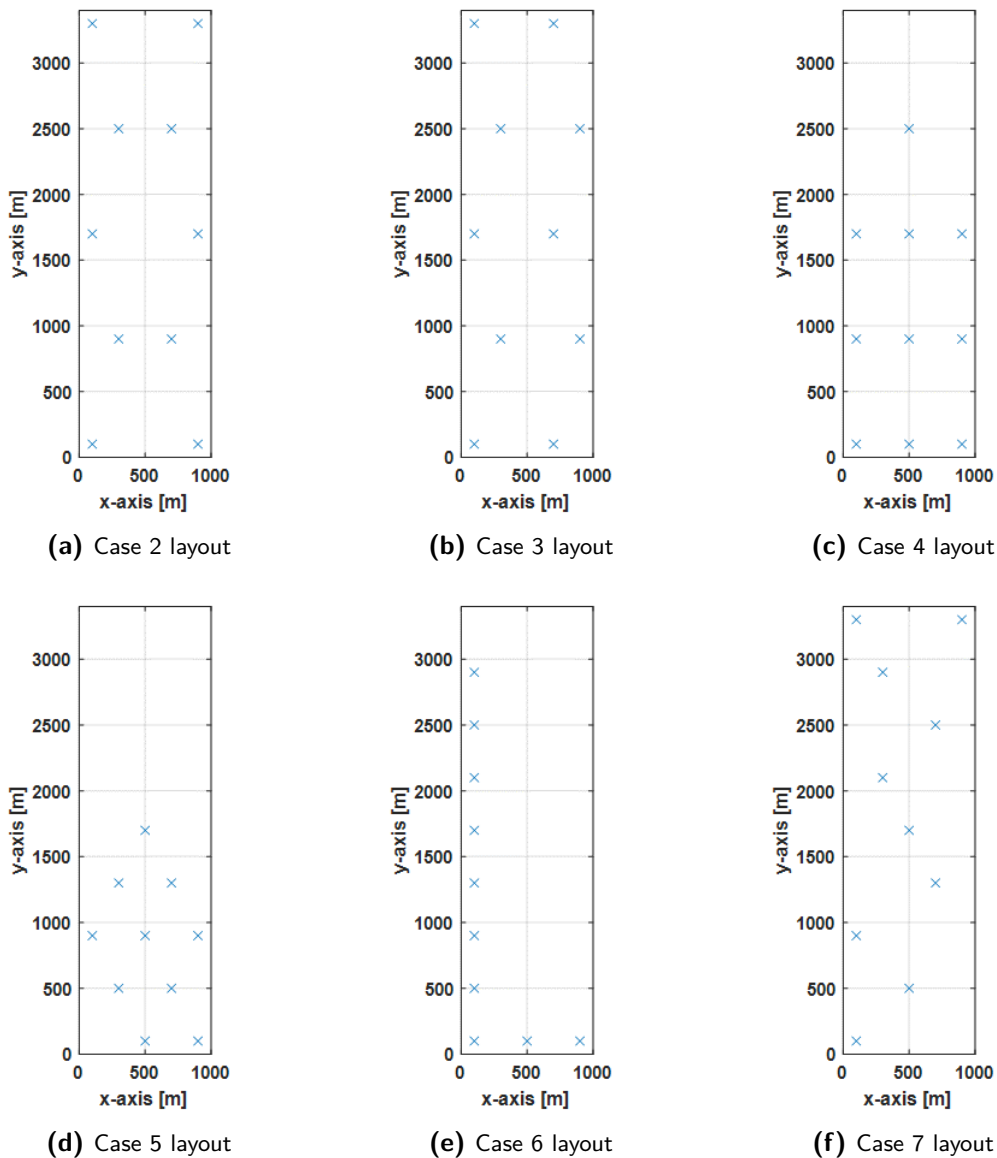


Figure 4-1: Overview of the 6 additional wind farm layouts under investigation.

In order to have a clear visualisation of the wake effect, the power rose for wind speed at 10 [m/s] is a preferred choice as this wind speed is below rated and above cut-in wind speed. The power rose of case 1 has been presented in Figure 3-3b, and the power rose of case 2 to 7 at 10 [m/s] using the original model are shown in Figure 4-2.

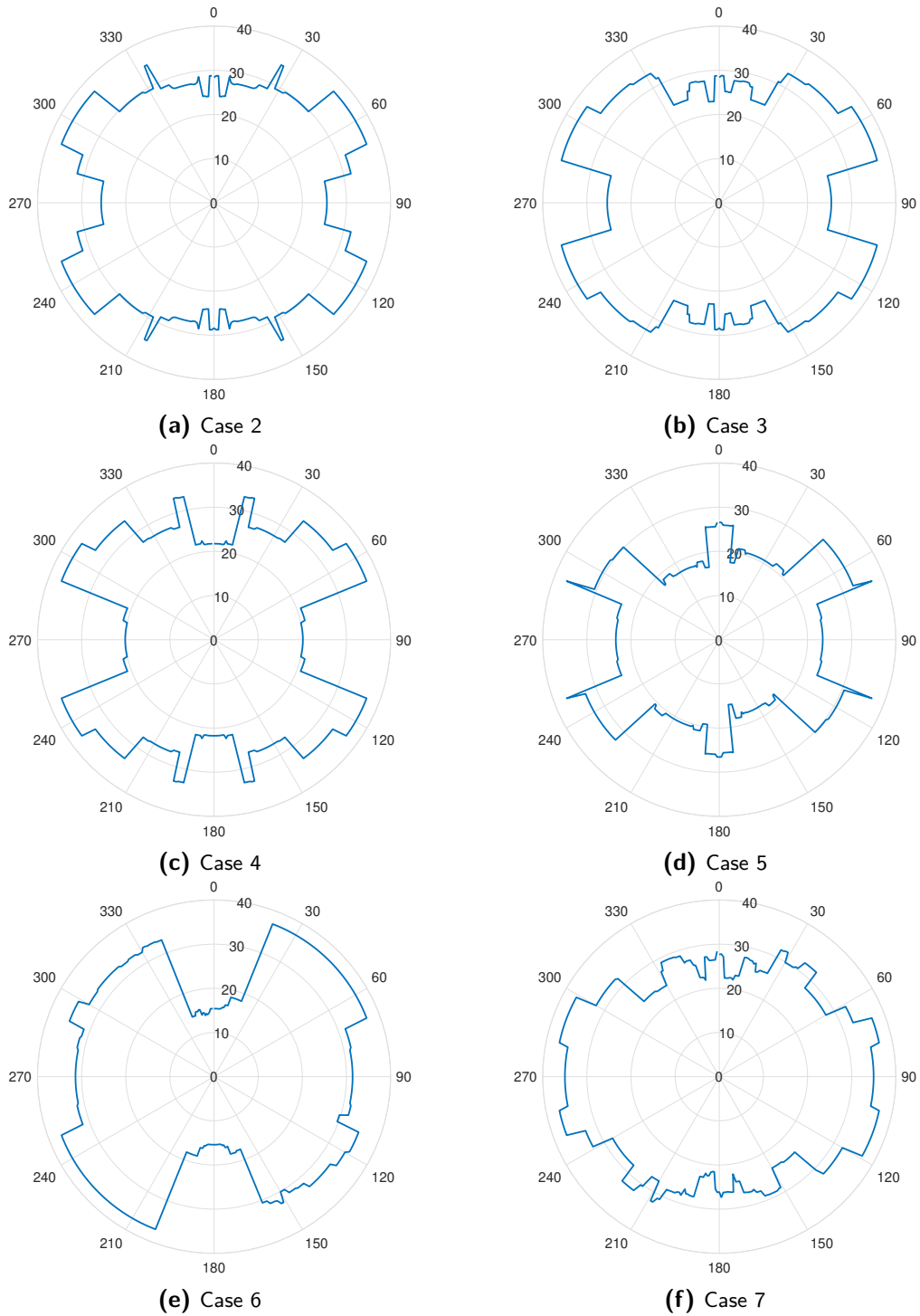


Figure 4-2: Power roses of each layout when wind speed is at 10 [m/s].

Several findings can be identified from Figure 4-2. It should be noted that the following comparisons are based on the AEP in Table 4-1. The AEPs are calculated with the full range of wind speeds and wind directions using the original wind farm power model. Firstly, the staggered layouts in case 2 and case 3 do not surpass the aligned layout in case 1 as expected. This could be because the dominant wind direction of the site is from the southwest direction, as shown in Figure 2-2. The dominant wind direction is not aligned with the closely spaced lines of wind turbines in case 1. As for case 2 and case 3, the staggered formation of the wind turbines coincide to this dominant wind direction and results in an inferior performance by the wake effect.

	AEP [MWh]	Capacity factor [%]
Case 1	$2.2348 \cdot 10^5$	48.17
Case 2	$2.1967 \cdot 10^5$	47.34
Case 3	$2.2075 \cdot 10^5$	47.58
Case 4	$2.0991 \cdot 10^5$	45.24
Case 5	$1.9387 \cdot 10^5$	41.78
Case 6	$2.1818 \cdot 10^5$	47.02
Case 7	$2.1787 \cdot 10^5$	46.96

Table 4-1: Summary of the AEP of all cases along with the capacity factor of the layouts.

Second, for the wake-intense cases, such as case 4 and case 5, a significant reduction in AEP can be seen. The power deficit from the wake effect is worst when the wind turbines are placed too close as case 5 implies.

Last, irregular layouts are hard to predict their performance. The L-shaped layout seems to experience a considerable wake effect when the wind direction is 180 degree, but it is expected that this layout can overcome the dominant southwest wind direction. Nevertheless, the AEP from this layout does not surpass the grid layout in case 1. A ranking on the AEP of all 7 cases using the original model is shown in the first row of Table 4-2. The best layout is case 1 with a standard gridded layout, follows by the two staggered layouts. The two irregular layouts are ranked 4th and 5th while the two wake-intense layouts are the worst ones.

4-2-2 Results of surrogate models on different wind farm layouts

Both PCE and Kriging surrogate models mentioned in Chapter 3 are implemented for the 6 additional wind farm layouts. The configuration of the wind farm power surrogate is kept the same for all layouts to have a valid comparison.

The study first looks into the percentage error of all cases on AEP for both surrogate models with different sampling schemes, namely the Latin hypercube sampling samples (LHS) and the linearly spaced samples (LSS). The comparison is shown in Figure 4-3. The differences show whether the surrogate models work better or worse on some wind farm layout. Based on the results, the differences between the estimation by the two surrogate models to the original model mostly lie within 2%. It is hard to tell from the comparison whether there is a kind of wind farm layout that cannot be handled by the surrogate models. The only observation is that the wake-intense layouts, such as case 4, show higher differences using both surrogate models.

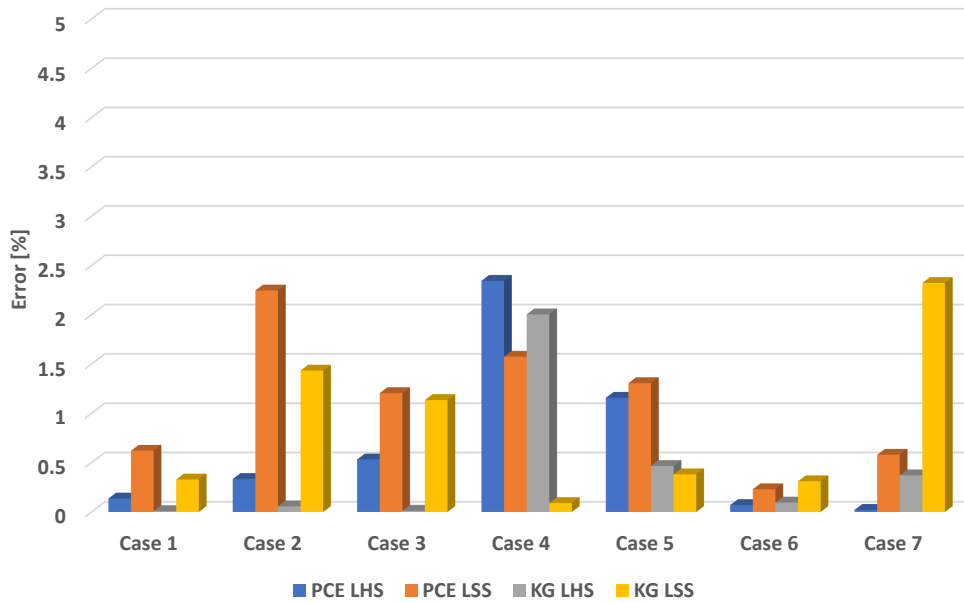


Figure 4-3: The AEP percentage error of all wind farm layout cases evaluates by the original model and the surrogate models.

Furthermore, the effectiveness of the surrogate models is compared. It has decided to present the effectiveness of the surrogate model as ranking for the easiness of interpretation. The advantage of showing ranking is that it could simulate the procedure within the wind farm layout optimisation. The wind farm layout optimisation compares the performance of every possible layout and chooses the best one out of it. As a consequence, the importance of checking the effectiveness of the surrogate models stands out as a false optimised layout may occur if one includes an ineffective surrogate model in the layout optimisation.

Besides the ranking using the original model, Table 4-2 also presents the ranking using the two surrogate models with two sampling schemes. Based on the ranking results, only PCE with LHS obtained the same ranking as the original model while other cases deviate from what has been known from the original model. Even though the ranking order from the 7 cases is different, all models reach the same conclusion that the two wake-intense cases perform the worst. Based on the ranking order, both surrogate models with LHS scheme seem to have a more accurate approximation than using the linearly spaced sample scheme. Overall, there is a possibility that the surrogate models may be effective for most kind of realistic wind farm layouts as the types of wind farm layout does not seem to have a significant influence.

		Ranking of cases												
	OM	Case 1	>	Case 3	>	Case 2	>	Case 6	>	Case 7	>	Case 4	>	Case 5
PCE	LHS	Case 1	>	Case 3	>	Case 2	>	Case 6	>	Case 7	>	Case 4	>	Case 5
	LSS	Case 2	>	Case 1	>	Case 3	>	Case 7	>	Case 6	>	Case 4	>	Case 5
KG	LHS	Case 1	>	Case 3	>	Case 2	>	Case 7	>	Case 6	>	Case 4	>	Case 5
	LSS	Case 3	>	Case 7	>	Case 2	>	Case 1	>	Case 6	>	Case 4	>	Case 5

Table 4-2: The AEP ranking of every cases using different wind farm power models.

4-3 Randomness from the sampling scheme

As the name of the sampling scheme implies, the random sampling scheme randomly samples from the wind speed and wind direction to construct the surrogate model for wind farm power output. As mentioned before, Latin hypercube sampling is used for this scheme. From Chapter 3 and the previous investigation on the effectiveness of the surrogate models, it was concluded that this sampling scheme had provided a better approximation to the results of the original model. However, it is found that this sampling scheme further causes uncertainty on the results as the samples could be different in each construction. This section looks at the results of the surrogate models with the Latin hypercube sampling scheme. It should be noted that as long as no predefined samples are specified, this phenomenon will exist for any random sampling methods. This randomness is actually epistemic uncertainty, and more simulations can better understand it. On that account, the models are iterated for 50 times, and the results are compared to see the uncertainty and the differences in different wind farm layouts.

Figure 4-4 presents the percentage error of surrogate models for different case layouts. As the uncertainty exists in the constructed surrogate models due to the random samples, box plots are used. According to the results, 50% of the error samples (denotes as the blue box), lie within 1.5%. The blue box represents the interquartile range of the data set. If outliers are taken into account, the uncertainty from both surrogate models in each case lies within 4%. The median of error (denotes as the red line in the box plot), is below 1% and more specifically, around 0.7% for every case. It has been learned in the previous section that PCE and Kriging surrogate models do not seem to be sensitive to different types of wind farm layout. The results from the figure confirm the previous findings.

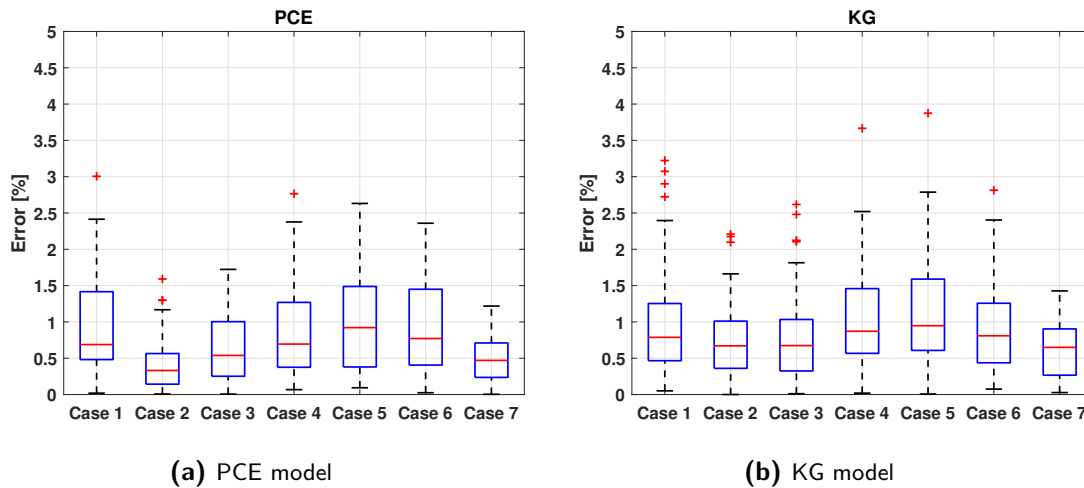


Figure 4-4: The percentage error from the random sampling schemes for the two surrogate models on different wind farm layouts.

The LHS sampling scheme has been proven to be a better one than the linearly spaced sampling scheme from both Chapter 3 and this chapter. However, as the upper bound of the medium shows an approximately 1% error in all cases, it results in additional uncertainty within the approximation of the surrogate model to the original model. Therefore, a solution

for this is proposed, which is to have a pre-defined sample set that is generated via Latin hypercube sampling. The use of the pre-defined samples preserves the advantage of sampling throughout the entire wind speed, and wind direction domain as linearly spaced samples are not capable of doing so. Moreover, it also keeps the variable sampling density of LHS while eliminating the randomness in each construction of surrogate models. This pre-defined sample set is then used for the construction of the surrogate models. It is possible to use this method to eliminate the uncertainty in random sampling and leave only the approximation error of the model itself.

4-4 Summary of the chapter

This chapter looks into the effectiveness of the surrogate models on different types of wind farm layout besides the one discussed in Chapter 3. The first type of layout is the staggered layout, the second type is the wake-intense layout, and the last one is the irregular layout. According to the results, the effectiveness of the surrogate models is considered accurate enough except some slight deviation in cases with a wake-intense layout. Nevertheless, the percentage error for all layouts lies within 2.5%. Next, the randomness from the random sampling scheme is discussed. It introduces additional uncertainty in the approximation of surrogate models. Based on the findings, the upper bound of the medium in error in all cases is approximated 1%. This error is then considered as the additional uncertainty caused by the Latin hypercube sampling scheme. A solution has also been proposed to avoid this uncertainty. It is by using a pre-defined sample set that is generated via Latin hypercube sampling. In this way, it is possible to preserve the advantage of sampling throughout the whole domain and the variable sampling density of LHS while eliminating the randomness in each construction of surrogate models.

Uncertainty Quantification

5-1 Introduction to the chapter

Fundamental knowledge of the uncertainty quantification is introduced in this chapter. In various research in the uncertainty quantification field, Monte Carlo simulation is most commonly seen. However, it has been identified that the high computational cost in this traditional uncertainty quantification method yields problems for high-fidelity and large simulation. Chapter 3 has verified that the use of surrogate models can reduce the computational effort of determining wind farm power and AEP. In Chapter 5, the surrogate models will be used for uncertainty quantification of AEP as a function of stochastic wind climate, which focuses on the year-by-year variation of Weibull distribution and wind rose. Lastly, a sensitivity analysis of the uncertain inputs is carried out to study how the inputs affect the output value.

5-2 Background overview

5-2-1 Introduction to uncertainty quantification

The research on uncertainty quantification (UQ) has seen rapid growth in almost all domains of sciences and engineering in the last decades. Deterministic-based modelling is gradually replaced by stochastic modelling to have a better understanding of the uncertainty in physical phenomena, and measurements [34]. In this case, it is crucial to analyse the statistical moments, such as mean and standard deviation of the probability density function (PDF) and cumulative density function (CDF) of performance functions, in other words, quantity of interest (QoI). The field of UQ covers a lot of topics, such as parameter estimation and calibration and sensitivity analysis [56]. In this research, UQ is used as a method to include the statistics and randomness from the natural wind phenomenon into the evaluation of annual wind farm yield.

Most commonly, an UQ process starts with a characterisation step that determines the nature of the uncertainties. Besides the distinction between aleatory and epistemic uncertainties,

another critical part of characterisation is to use probability theory, for example, finding the PDF for the uncertain variable. Once the characteristics of the uncertainties are known, a propagation step is conducted. The uncertainties are propagated through a computational model to obtain the relation between inputs and QoI. UQ can be further distinguished into intrusive and non-intrusive based on the propagation method. The difference is that the intrusive methods require one to change the mathematical model and demand a great coding effort while the non-intrusive methods treat the models as a *black box*. An example of a non-intrusive UQ flowchart is shown in Figure 5-1. The non-intrusive method is the choice of this research and is most widely applied in the wind energy community [56].

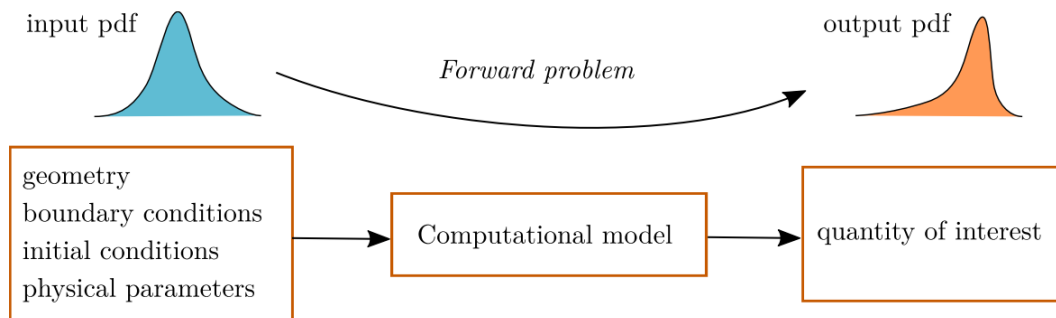


Figure 5-1: Flowchart of a non-intrusive forward uncertainty propagation [56].

However, the black box models used in UQ could be prohibitively computationally expensive. Therefore, various methods are proposed to obtain accurate outputs with as few samples as possible [56]. In particular, the high computational cost in traditional UQ via sampling-based approaches (e.g., Monte Carlo method) yield problems for high-fidelity (e.g., CFD codes) and large simulation. Thus, surrogate models are again proposed to alleviate the cost and making the process more efficient by smartly sampling the input parameters, which will be addressed in the later sections. The UQLab toolbox is also utilised to conduct the uncertainty quantification.

5-2-2 Wind resource inter-year uncertainty

Mentioned in Chapter 1, a wind farm project typically contains significant financial risks while developing and operating. Consequently, to lower such risk, several researchers have investigated the topic of uncertainty characteristics within wind energy. Discussed in Quaeghebeur [45], the lifetime production of a wind farm is important, but so is the yearly production. A large inter-year variation could lead to financial risks, and the author concluded that the inter-year wind resource varies massively in real measurement. In order to investigate the inter-year uncertainty of the wind resource, the weather data is divided into yearly data. The results of post-processing the weather data, the joint probability distributions in Figure 5-2, clearly show the inter-year difference in wind resource. The shown inter-year difference supports the statement made in Quaeghebeur [45]. The research continues to find a method to characterise the uncertainty for the year-to-year variation in the wind resource.

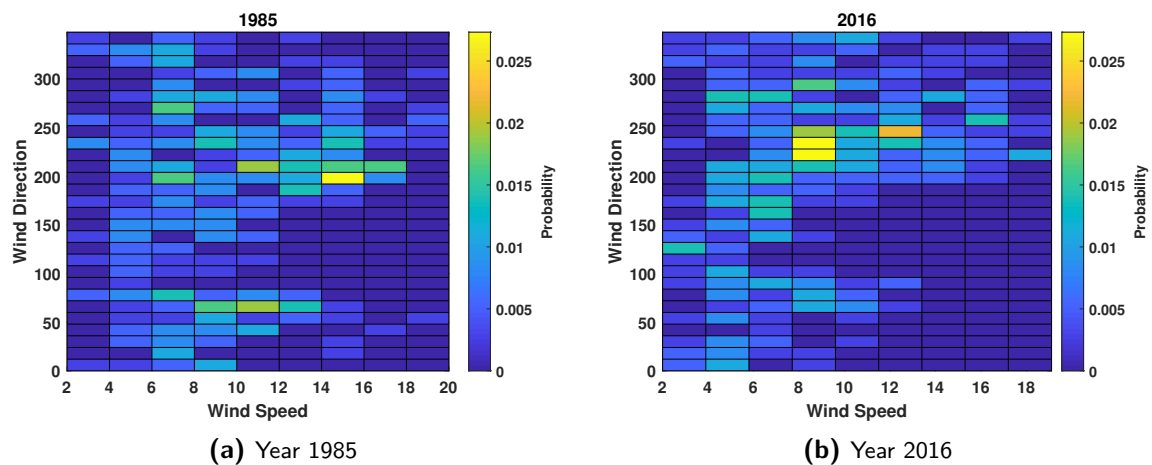


Figure 5-2: A comparison of joint probability distribution of wind speed and wind direction in different years.

The first attempt is to characterise the Weibull distribution within each wind direction sector. In this sense, each wind direction sector will have independent Weibull parameters to other sectors. These Weibull parameters vary from year-to-year and can therefore be considered to be stochastic variables. The statistics of the yearly wind speed distribution, as expressed in the probability distribution of the Weibull parameters, is then dependent on wind direction. Nevertheless, this approach turns out to introduce a significant amount of uncertain variables if one wants to characterise the uncertainty in the wind speed distribution within every wind sector. For example, if one divides the wind direction into 12 sectors, there will be $12 \cdot 2 = 24$ uncertain variables for wind speed if a two-parameter Weibull distribution is used. This number of uncertain variables has only taken the wind speed into account, and it will be even more when uncertainty in the probability of occurrence per wind sector is considered as well. Therefore, a second attempt that utilises an aggregated Weibull distribution for all wind direction sectors is implemented. The precision in estimating AEP will reduce, but the yearly variation will still be captured. This approach is suitable for this research as the research aims to deal with the inter-year variability in wind resource. The uncertain variables for the wind speed, thus reduce to two, which are the Weibull scale and shape parameters.

The 50 years measurement from the website of KNMI is processed to have information about the yearly wind speed and wind direction distributions. After sorting out the occurrence of wind speed in individual years, Figure 5-3a can be obtained. In Figure 5-3a, each colour represents a year within the 50 years, and from each Weibull distribution curve, the Weibull scale and shape parameters can be extracted. In contrast, there is no suitable mathematical probability distribution for the wind direction, as shown in Figure 2-4a, and is therefore normally presented by a wind rose. It is decided to break down the wind direction into sectors (e.g. bins). Based on the number of occurrences of wind direction that falls inside a sector, a sector probability is assigned. The uncertainties related to the wind direction are thus the sector probabilities in the wind frequency rose. The different colours in Figure 5-3b denote different wind frequency roses of each year with different colours. The figure evidently presents the inter-year variation in wind direction. Further, the changes in probability inside each sector can also be clearly seen.

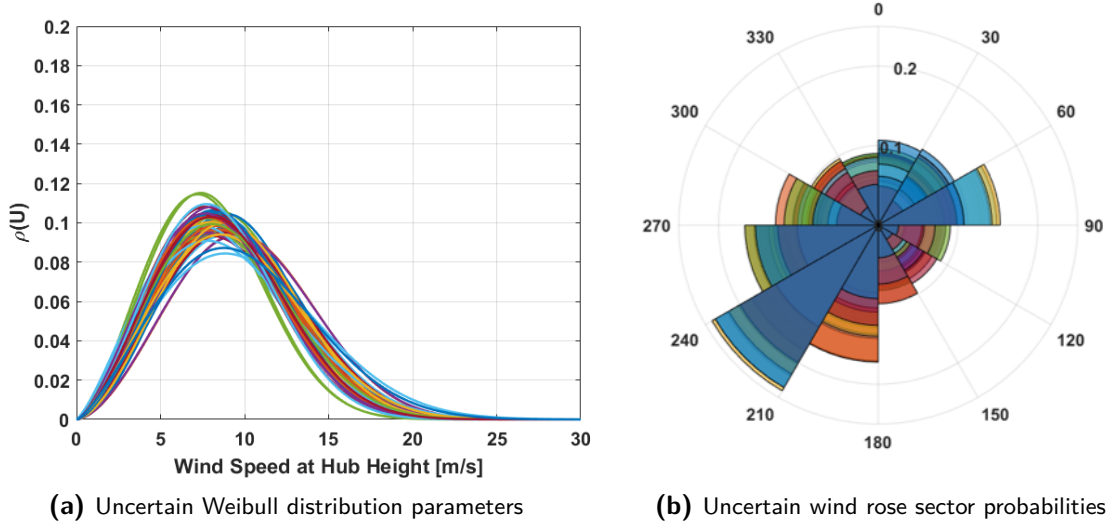


Figure 5-3: The uncertain variables concluded from the 50 years measurement data.

5-2-3 Statistical inference on wind data

Using the 50 years of weather data from the IJmuiden station, the uncertain variables (e.g., Weibull parameters and wind rose sector probabilities) are obtained by post-processing the daily data to investigate the yearly differences. These uncertain variables are subjected to distribution fitting using the UQLab toolbox. There are two main steps for conducting statistical inference. The first one being fitting different probabilistic models to the data, and the second is to select the one with the best fitting performance[54]. In this research, the data is fitted with the parametric inference that gives probability distributions with fixed analytical expressions. The fitted distribution generally utilised the maximum likelihood solution, in which the likelihood function \mathcal{L} is the maximum:

$$\arg \max_{\theta} \prod_{h=1}^n \rho_X(\hat{x}^{(h)}; \theta) \stackrel{def}{=} \arg \max_{\theta} \mathcal{L}(\theta) \quad (5-1)$$

Eq. (5-1) [54] considers a case with a random vector X where ρ_X is the probability distribution function. The fitted distribution parameters θ are determined by n independent observations \hat{x} of X . The final distribution is determined by the one that has the minimum Akaike Information Criterion (AIC), where k is the number of model parameters and \mathcal{L} is the likelihood function.

$$AIC = 2k - 2\log(\mathcal{L}) \quad (5-2)$$

Table 5-1 summarised the probability distribution and the statistical moments of the inferred uncertain variables. It has to be noted that it is not possible to simply infer all 12 sectors of wind direction. As the probability of the sectors is the uncertain variable, it still needs to satisfy the rule of probability that the probability from every sector has to sum up to one. As the sector probabilities become uncertain variables, they are now represented by PDF instead of a fixed value. When sampling from the PDF of each wind sector, the summation could likely be more or less than one. Therefore, to avoid such matters, the wind sector of

[210°, 240°] is intentionally left out as its the most dominant sector among all sectors. This makes the fact that only 11 out of 12 sectors are statistically inferred. The omitted sector is determined by the rest of the sectors. It comes from the total possibility one subtracts the summation of all others 11 sector probabilities. This guarantees that the total probability of all wind sector can be "one", as stated in the mathematical and probability theory.

Uncertain variables	Distribution	Mean	Standard deviation
Weibull scale parameter λ	Gaussian	10.1746	0.5728
Weibull shape parameter k	Gaussian	2.5063	0.1191
Wind sector probability [0°, 30°]	Beta	0.0645	0.0176
Wind sector probability [30°, 60°]	Laplace	0.0575	0.0208
Wind sector probability [60°, 90°]	Beta	0.0947	0.0250
Wind sector probability [90°, 120°]	Weibull	0.0529	0.0171
Wind sector probability [120°, 150°]	Beta	0.0439	0.0138
Wind sector probability [150°, 180°]	Beta	0.0617	0.0150
Wind sector probability [180°, 210°]	Lognormal	0.1113	0.0213
Wind sector probability [240°, 270°]	Logistic	0.8888	0.0400
Wind sector probability [270°, 300°]	Gaussian	0.0801	0.0177
Wind sector probability [300°, 330°]	Weibull	0.0649	0.0172
Wind sector probability [330°, 360°]	Weibull	0.0646	0.0143

Table 5-1: Uncertain variables and the corresponding statistical properties from the inferred data.

5-2-4 Quantity of interest

In order to lower the risk for an investment of a wind farm even more, a statistic confidence level $P90$ could be utilised. In the name of Pxx , xx can range from 1 to 99, and denotes the probability percentage that the value will be exceeded in a statistic response. For example, due to the uncertainty in the wind condition, AEP of a wind farm is actually not a deterministic value but a statistical response. To be more specific, the inter-year uncertainty in wind resource is the reason that makes the AEP into a statistic response. The AEP that has been mentioned in the previous chapters is a mean, which is obtained by:

$$\mu_{AEP} = 8766 \cdot \mu_P \quad (5-3)$$

$$\mu_P = \mathbb{E}[P(X)] = \int_{\Omega} P(X)\rho(X)dX \quad (5-4)$$

where $X = (X_1, X_2, \dots, X_n)$ is a vector of variables, which contains wind speed U and wind direction θ . $\rho(X) = \prod_{k=1}^n \rho_k(X_k)$ is the joint probability density function of the random variables, Ω is the domain of the variable, and P is the power production by the wind farm. When the statistic response results in a symmetrical probability distribution, such as the normal distribution, the equations denote the $P50$ of the statistic responses. The value 50 after P indicates that there is a probability of 50% that the actual AEP of the wind farm will exceed this $P50$ value. For a normal distribution, this value is also the mean value μ or the expected value \mathbb{E} as shown in Eq. (5-4).

For a statistic response like the normal distribution, the $P90$ value is smaller than the μ_{AEP} . In this sense, $P90$ means that the actual AEP is expected to have the 90% chance that the value will be larger than the $P90$ value, and there is only 10% chance that it will be lower. It should be noted that the standard deviation σ is also crucial when calculating the $P90$ value. The $P90$ value is different for two normal distribution with the same mean μ but different standard deviation σ , as seen in Figure 5-4. The need for accurately performing uncertainty quantification is thus required as an inaccurate approach could lead to a consequence of the wind farm financing [10].

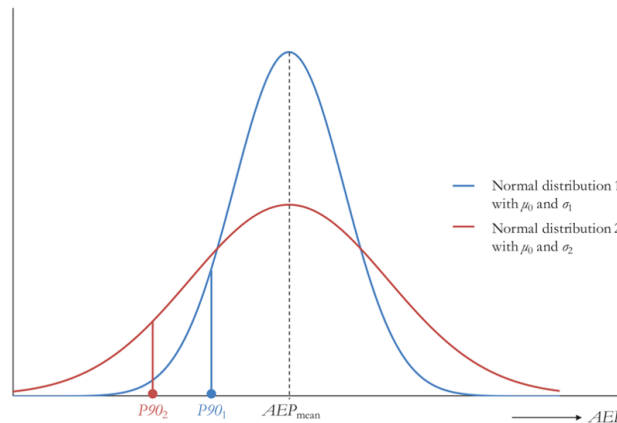


Figure 5-4: Illustration of $P90$ values with two normal distributions with the same mean value but different standard deviation [10].

Nevertheless, it should be noted that when the probability distribution is not symmetrical, the mean will not be the $P50$ value as what has been for a normal distribution. In this case, the QoI will then obtain from the cumulative density function of the statistical response, as shown in Figure 5-5. The $P90$ value will correspond to the 10% cumulative probability, while the $P50$ value denotes the 50% cumulative probability.

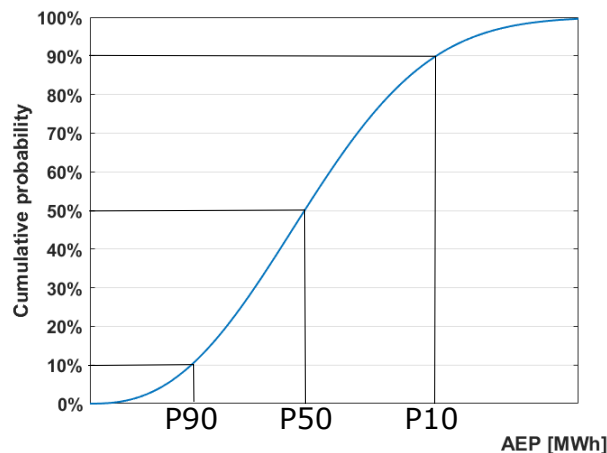


Figure 5-5: Illustration of cumulative density function with $P90$, $P50$, and $P10$ values.

5-3 Uncertainty propagation methods

5-3-1 Monte Carlo method

One of the non-intrusive methods of uncertainty propagation via sampling is the Monte Carlo (MC) method. The method is based on drawing random samples from the probability distribution of the random variables. The mean and the variance of the QoI can be approximated by

$$\mu_Y = \mathbb{E}[Y] \approx \frac{1}{N} \sum_{i=1}^N Y \quad (5-5)$$

$$\sigma_Y^2 = \text{Var}[Y] \approx \frac{1}{N} \sum_{i=1}^N (Y - \mathbb{E}[Y])^2 \quad (5-6)$$

where Y is the model response and N is the number of samples.

Advantages of the Monte Carlo simulation (MC) are the ease of implementation, and it is free of the curse of dimensionality. However, the method can be expensive when applied to a high-fidelity simulation code that requires hours for a single simulation [15]. This is because the accuracy for MC increase with $\mathcal{O}(1/\sqrt{N})$, where N is the number of samples. Therefore, for an accurate uncertainty analysis, MC may require more than thousands of samples or model evaluations. It is, therefore, exploring the possible use of surrogate models for uncertainty propagation.

5-3-2 Propagation with surrogate models

As mentioned in the previous paragraph, even though the Monte Carlo method is robust and versatile for uncertainty quantification, it can yield an excessive large computational expense. Due to this situation, the application of surrogate modelling in UQ has been investigated. For example, it is possible to acquired statistical information of the output (i.e., QoI) analytically from the PCE coefficients.

The uncertain inputs for the UQ model have been mentioned in the previous section. Based on the experience from the previous chapter, it has been acknowledged that PCE is faster in term of construction and providing statistical output. However, when the research focuses on the wake effect phenomenon, the Kriging surrogate model is able to provide a better solution. The conclusion is based on the results of building a surrogate model for the wind farm power output. It will be tested whether the conclusion also holds when using the surrogate model as a mean in the uncertainty quantification process.

Applying the PCE model in UQ

The PCE has been a popular surrogate technique in the uncertainty quantification field, especially as an alternative method to Monte Carlo simulation method [39]. The use of calculating the polynomial chaos coefficients for uncertainty quantification is not yet clarified. Before

knowing the application, some statistical knowledge needs to be explained. Statistically, the mean (i.e., expected value) of a function $\mathcal{M}(X)$ can be defined as

$$\mu_{\mathcal{M}} = \mathbb{E}[\mathcal{M}] = \int_{\Omega} \mathcal{M}(X)\rho(X)dX \quad (5-7)$$

With the polynomial expansion in Eq. (3-2), the above expression of mean is then transformed into

$$\mu_{\mathcal{M}} = \int_{\Omega} \sum_{i=0}^p y_i \Psi_i(X)\rho(X)dX \quad (5-8)$$

The expression can be further simplified using the orthogonality of the polynomials. The mean can therefore be obtained and is thus the zeroth coefficient as

$$\mu_{\mathcal{M}} = \mathbb{E}[\mathcal{M}^{PC}(X)] = y_0 \quad (5-9)$$

The variance has a more complex expression but can also be simplified using the polynomial approximation and the mean as

$$\sigma_{\mathcal{M}}^2 = \mathbb{E}[(\mathcal{M}^{PC}(X) - \mathbb{E}[\mathcal{M}^{PC}(X)])^2] \quad (5-10)$$

$$= \mathbb{E}[\mathcal{M}^{PC}(X)^2] - (\mathbb{E}[\mathcal{M}^{PC}(X)])^2 \quad (5-11)$$

$$= \int_{\Omega} \mathcal{M}(X)^2 \rho(X)dX - \mu_{\mathcal{M}}^2 \quad (5-12)$$

$$= \sum_{i=1}^p y_{\alpha}^2 \langle \Psi_i^2(X) \rangle \quad (5-13)$$

Therefore, once the coefficients are computed, the statistics of the model response can be inexpensively computed as the mean and variance are functions of the coefficients.

Applying Kriging model in UQ

The statistical moments of the output response is a must-know property for most of the UQ problem. Similar to PCE, it is possible to use the theory of Kriging model that mentioned in Chapter 3 to the uncertainty quantification process. Once the Kriging model has been built, the statistical moment can be efficiently computed via Monte Carlo analysis. The mean of the response can obtain again by Eq. (5-7) and the variance can be calculated by:

$$\sigma_{\mathcal{M}}^2 = \int_{\Omega} (\mathcal{M}(X) - \mu_{\mathcal{M}})^2 \rho(X)dX \quad (5-14)$$

Even though Kriging model does not have the advantage in analytically obtain the mean and variance of the output response from the polynomial coefficient, it is still more efficient than the traditional Monte Carlo method once the surrogate model is built.

5-4 Implementation of the surrogate models

5-4-1 Overview

Once the surrogate models on the wind farm power production were built as shown in Chapter 3, uncertainties in the inputs (i.e. Weibull parameters and wind rose sector probabilities) are propagated using MC, and surrogate models using PCE and Kriging. Different combinations of surrogate techniques for the AEP model and uncertainty quantification are investigated with a reference case. The reference case is where an original model with uncertainty propagation through a MC method. Here, MC is regarded as the ground truth for the UQ results. There is no use of surrogate modelling techniques, and thus, it is regarded as a reference case for the comparisons. The reference case is illustrated first and is abbreviated as the OM-MC combination.

However, an analysis of the number of uncertain sample or number of trials needed for MC is required. It is because one of the disadvantages of MC is that a single run of MC simulation may not be reliable and the result may vary a lot from a different simulation. Nevertheless, the greater the number of samples in a MC simulation, the more stable will be the output response. It is, therefore, looking into the number of samples that is adequate enough for the MC simulation. Figure 5-6 demonstrates the investigation and presents the finding that with 100,000 number of samples, the mean AEP and the standard deviation AEP converges to a very small variation. It is especially important for the standard deviation to present this convergence as it represents the uncertainty of the measurand [11].

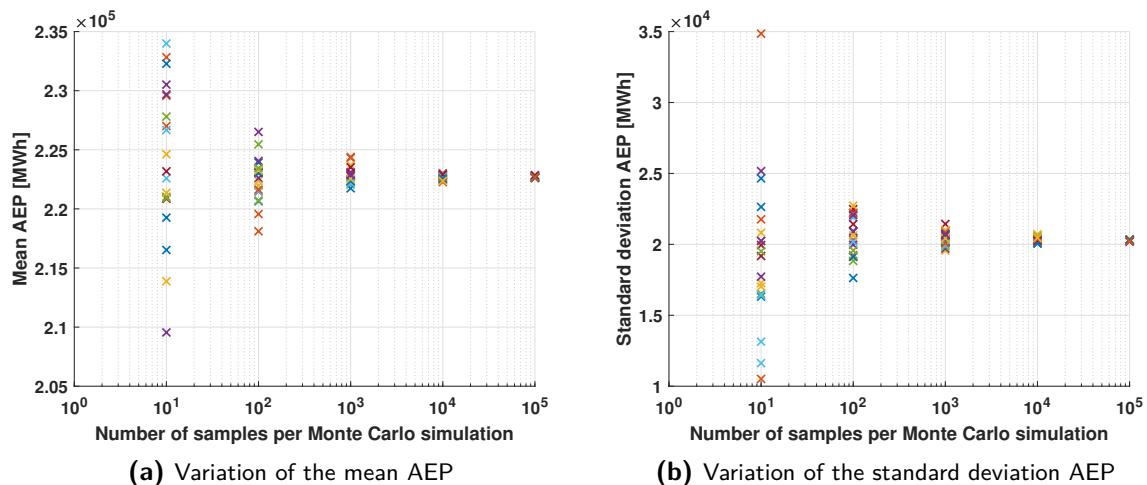


Figure 5-6: Variation of mean and standard deviation AEP calculated by different number of samples per Monte Carlo simulation. 20 different Monte Carlo simulation are trialed per number of samples.

Therefore, the MC simulation propagates the uncertain inputs with 100,000 samples. The P50 AEP value in this case is $2.2270 \cdot 10^5$ [MWh] with a capacity factor of 48%. The value differs a bit from what has been obtained from the original model of the wind farm power production. However, this is expected as the original model uses the average of the Weibull scale and shape parameters and the average probability per wind rose. Since the P50 value is not always the same as the mean value, the values are reasonable. The P90 value of AEP

is $1.9231 \cdot 10^5$ [MWh]. In addition to that, the histogram of the number of occurrence of AEP value, PDF, and CDF obtained by OM-MC are shown in Figure 5-7. The shape of the PDF curve does not appear to be a Gaussian distribution as it slightly displayed a negative skewness. This phenomenon could be a result of what has been shown in Table 5-1, where all different input distributions are mixed up with each other while some inputs have larger effects over others. A sensitivity analysis of the input variables to the output response could help to understand this effect more and Section 5-5 is dedicated to this.

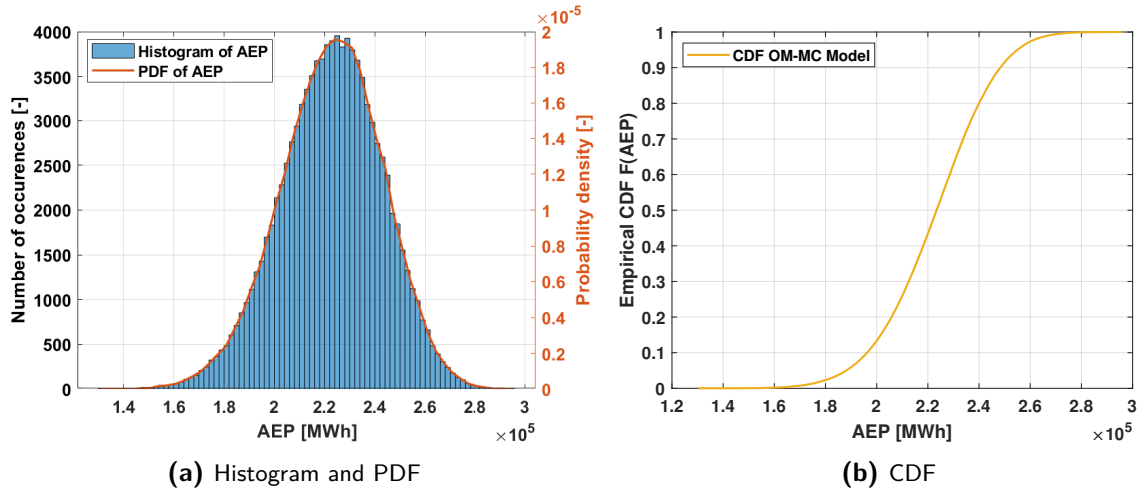


Figure 5-7: The PDF and CDF of the AEP value using OM-MC combination.

In order to conduct a validation of the stochastic approach, the original model mentioned in Chapter 3 is referred to as the deterministic approach. The deterministic approach uses the mean value of the uncertain variables (i.e. Weibull parameters and sector probabilities) from the 50 years measurement. The deterministic AEP for the investigated wind farm is computed as $2.2348 \cdot 10^5$ [MWh] while the stochastic approach of this chapter considers a value with 90% probability of exceedance. As a result, the cumulative chance of 10% (i.e., P90) for the AEP is expected to return a value that is less than $2.2348 \cdot 10^5$ [MWh]. Table 5-2 presents the comparison between the two approaches, where the AEP value from the deterministic OM model and the P50 AEP value of OM-MC are similar, and the P90 value is indeed smaller than the value from the OM model. The difference in evaluation time indicates that an additional step is conducted to propagate the yearly uncertain inputs by MC simulation.

	Deterministic OM	OM-MC
Number of uncertain samples	-	100,000
AEP/Mean AEP [MWh]	$2.2348 \cdot 10^5$	$2,2275 \cdot 10^5$
P50 AEP [MWh]	-	$2,2270 \cdot 10^5$
P90 AEP [MWh]	-	$1.9623 \cdot 10^5$
Std AEP [MWh]	-	$2.0287 \cdot 10^4$
Time [s]	215.2038	217.0501

Table 5-2: The performance of the OM model compared to OM-MC method for uncertainty quantification.

Next, a comparison of the multiple combinations of surrogate models is performed. The goal

is to find the combination that can provide the best estimation on the P50 and P90 value of AEP with a small number of samples and time effort. For simplicity, the name of the combination is shortened with the abbreviations where the first acronym denotes the model used to calculate the wind farm power output and the second one indicates the model used to propagate the yearly uncertain inputs.

5-4-2 Verification of different configurations

Figure 5-8 presents the difference in the estimation of P50 and P90 value of AEP and model building time with a different number of samples from uncertain inputs for various combination configurations. There are in total of 6 combinations that are under investigation, namely, PCE-MC, PCE-PCE, PCE-KG, KG-MC, KG-PCE, and KG-KG.

First of all, the most apparent differences between the 6 cases are the time required. Based on the figure, the combinations that use the PCE surrogate model for the wind farm power output lead to a faster computational time except for the PCE-KG case. The reason leads back to Chapter 3 where it indicates that Kriging requires more time to construct the surrogate model. The same behaviour is observed in the cases where Kriging takes part in, either as the surrogate model of power output or as a means to propagate the uncertain inputs. It is worthy to note that model building time mentioned here is regarding the UQ surrogate model. Both PCE-MC and KG-MC cases do not consist of the UQ surrogate model, so only the total computational time is shown for the two cases.

Regarding the accuracy of the estimations on P50 and P90 value of AEP, the approximation of the P50 value of AEP is better than the P90 value. Moreover, when PCE surrogate model is used as an UQ model, the required number of samples to reach an adequate accuracy is less than using Kriging as an UQ model. In general, PCE UQ model requires at least 100 samples for P50 and P90 value to converge while the Kriging UQ model commonly requires 200 samples from the uncertain inputs. Lastly, the two configurations that use the Monte Carlo method to propagate the uncertainty need more samples to reach the accuracy of using the surrogate model configuration, which is often more than 500 samples.

The following paragraphs present the PDFs and CDFs of the AEP, determined with the implementation of the six mentioned combinations.

1. PCE-MC

After introducing the reference OM-MC case, the investigation first focuses on the uncertainty quantification with the PCE power production surrogate model. The configuration for the first surrogate model follows the suggestion in Chapter 3. The first combination under this scenario is the combination of PCE-MC, where a PCE surrogate model on wind farm power performance is used, and the uncertainty inputs are propagated with MC simulation with 1,000 samples. The number of samples for the uncertain inputs is based on the results in Figure 5-8a, where a sufficient accuracy is reached with an acceptable computational time. As shown in Figure 5-9, PCE-MC model has an acceptable agreement with the OM-MC model, yielding a P50 value of $2.2519 \cdot 10^5$ [MWh] and a P90 value of $1.9837 \cdot 10^5$ [MWh].

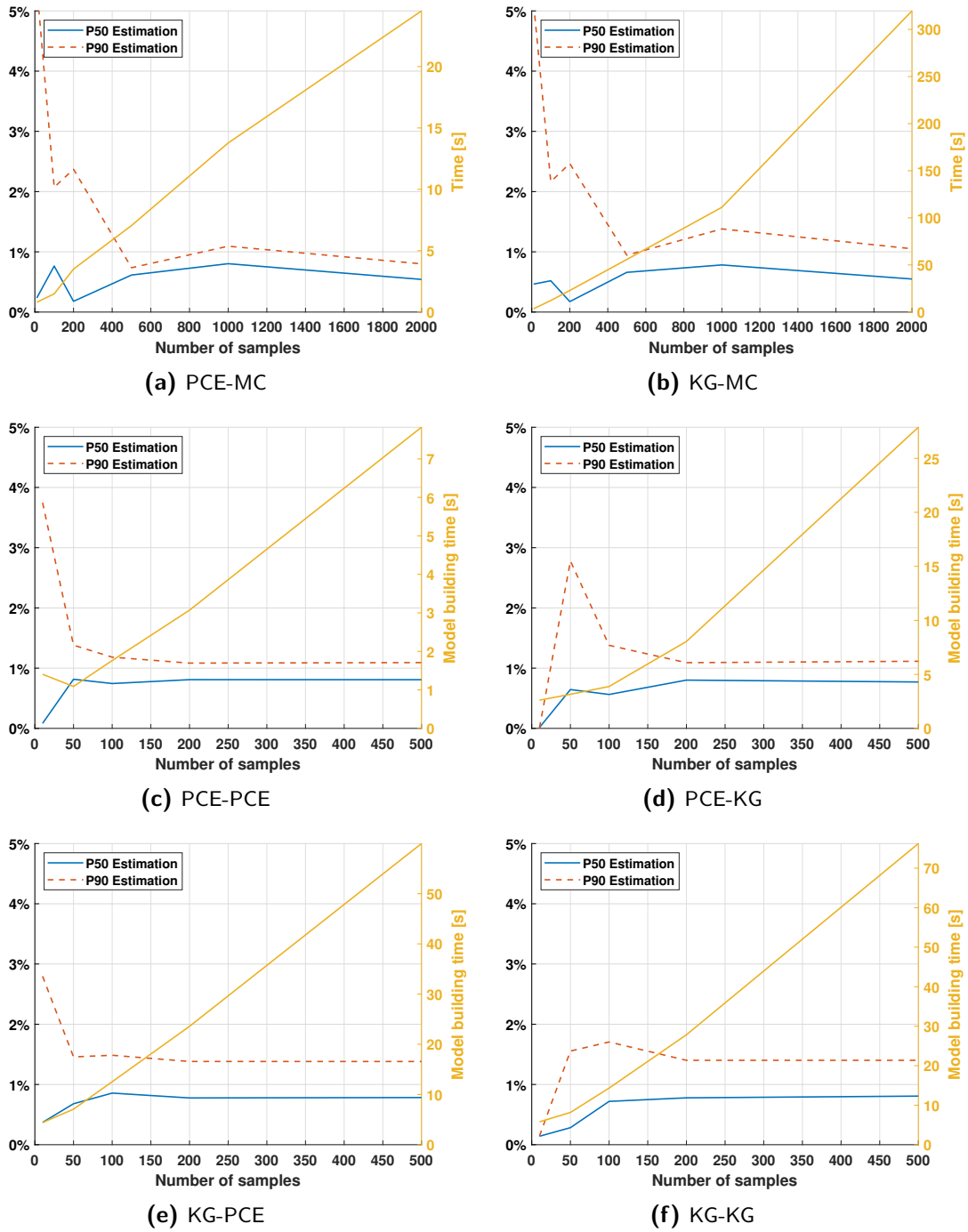


Figure 5-8: The difference [%] in the estimation of P50, P90, and model building time with different number of uncertain samples for various configurations.

2. KG-MC

Similar to the PCE-MC model, a Kriging surrogate model on wind farm power performance is used, and the uncertain inputs are propagated with MC simulation with 1,000 samples. Based on the results in Figure 5-8b, it seems that 1,000 samples can already

provide an accurate estimation as the PDF and CDF plot in Figure 5-10 verifies it. Both models, PCE-MC and KG-MC, utilise MC simulation for the uncertainty propagation demonstrate a slightly over-estimate but still a good fit to the OM-MC model. According to the results, a P50 value of AEP using this combination is $2.2514 \cdot 10^5$ [MWh] and $1.9893 \cdot 10^5$ [MWh] for the P90 value. It is observed that the estimations of both P50 and P90 values of AEP are inferior to PCE-MC as the values are further away from the OM-MC model.

3. PCE-PCE

In PCE-PCE combination, a PCE surrogate model on wind farm power performance is built and followed by UQ that propagates uncertain variables using the PCE techniques. The configuration for the second surrogate model that propagates the uncertain inputs is determined by trial and error method, after which a 4th order LARS is chosen as it can provide a better estimation. With 100 samples from the uncertain inputs, the PDF and CDF are displayed in Figure 5-11.

4. PCE-KG

In PCE-KG case, a PCE surrogate model on wind farm power performance is built and followed by UQ that propagates uncertain variables with the Kriging techniques. The PCE model for the wind farm power evaluation follows the one suggested in Chapter 3 and the second one is determined by searching for the best estimation from several attempts. The configuration is set to second-order Universal Kriging. From the attempts, the author finds that the Kriging model can not handle large sample size while using a high order of Universal Kriging as a means to propagate uncertain variables. It results in significantly long computation time and does not guarantee an accurate estimation. With 100 samples from the uncertain inputs, the PDF and CDF are displayed in Figure 5-12.

5. KG-PCE

A Kriging surrogate model on wind farm power performance is built and is followed by UQ that propagates uncertain variables with the PCE techniques with 100 samples from the uncertain inputs. Same with the PCE-PCE case, the configuration for the second surrogate model is a 4th order LARS as it can provide a better estimation. Compared to the PCE-PCE model where the PCE technique is also used to propagate the uncertain inputs, both KG-PCE and PCE-PCE provides a good agreement with the OM-MC model.

6. KG-KG

Lastly, a Kriging surrogate model on wind farm power performance is built and followed by another Kriging model that propagates 100 samples from the uncertain variables. The UQ configuration is also set to second-order Universal Kriging.

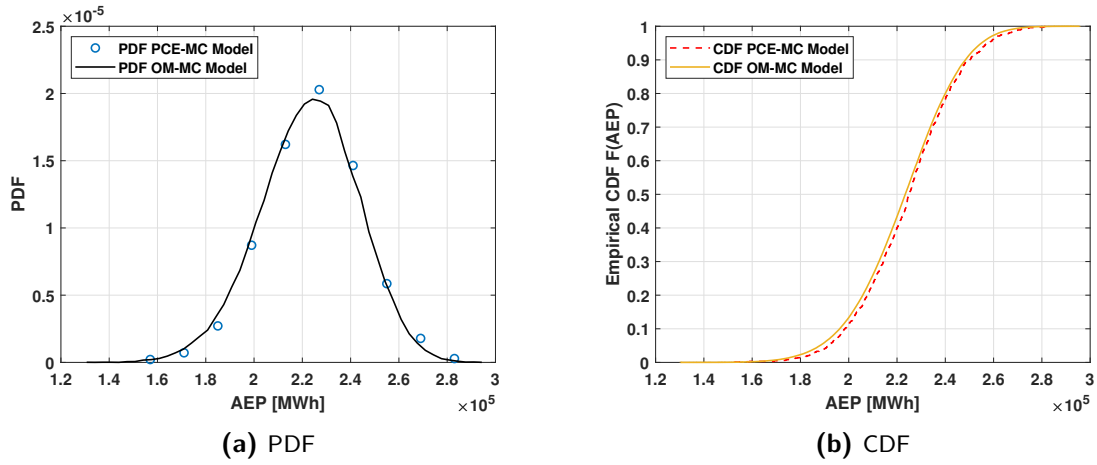


Figure 5-9: PDF and CDF of the AEP value using the PCE-MC combination.

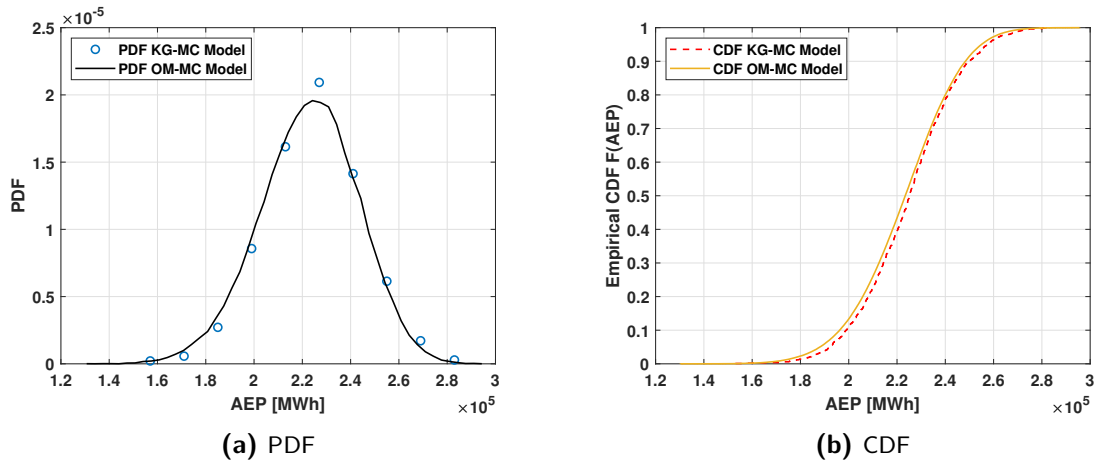


Figure 5-10: PDF and CDF of the AEP value using the KG-MC combination.

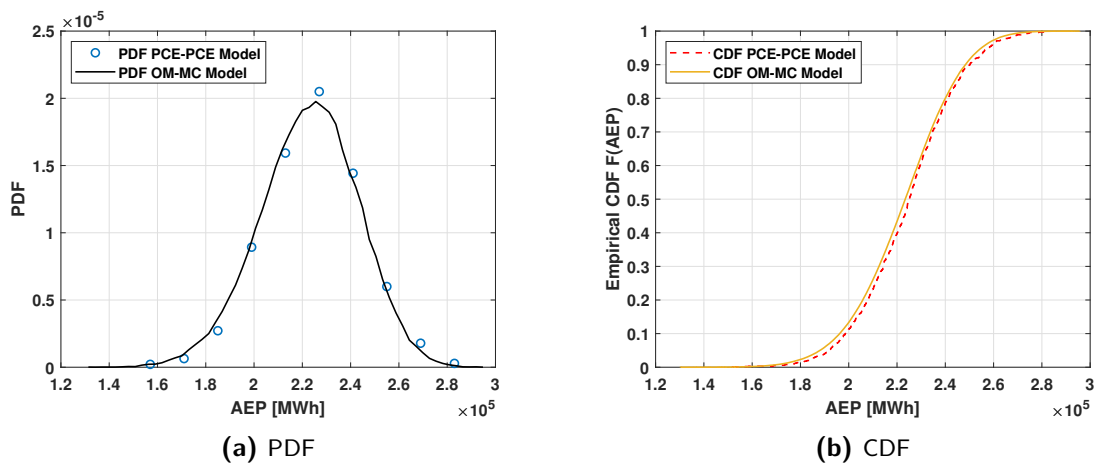


Figure 5-11: PDF and CDF of the AEP value using the PCE-PCE combination.

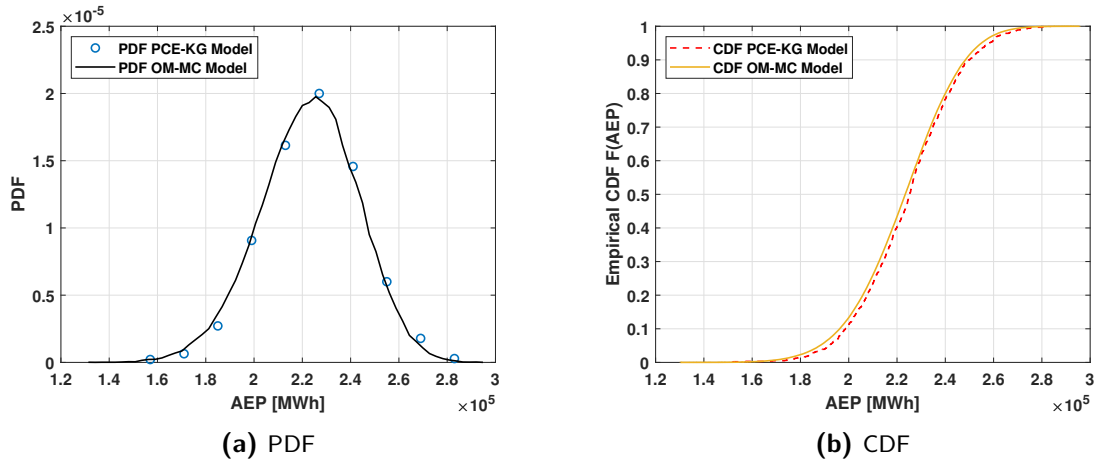


Figure 5-12: PDF and CDF of the AEP value using the PCE-KG combination.

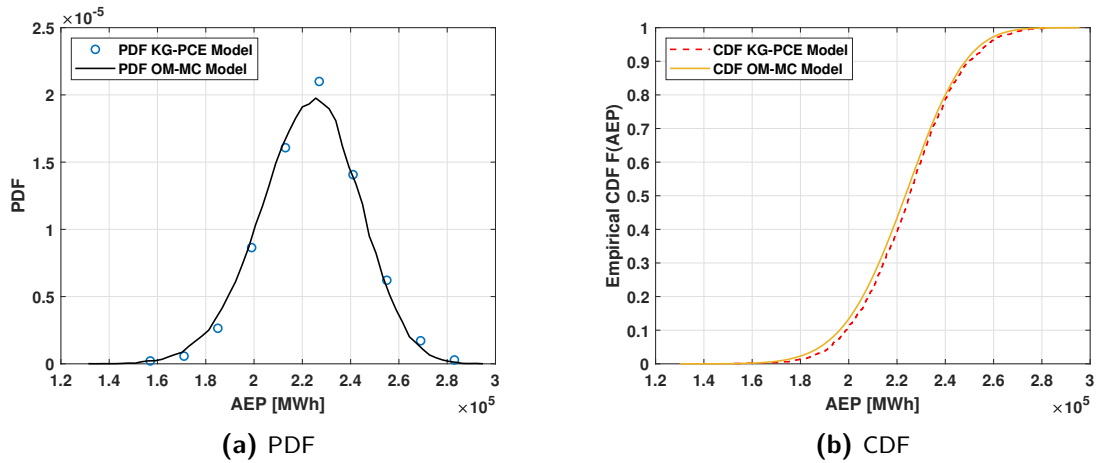


Figure 5-13: PDF and CDF of the AEP value using the KG-PCE combination.

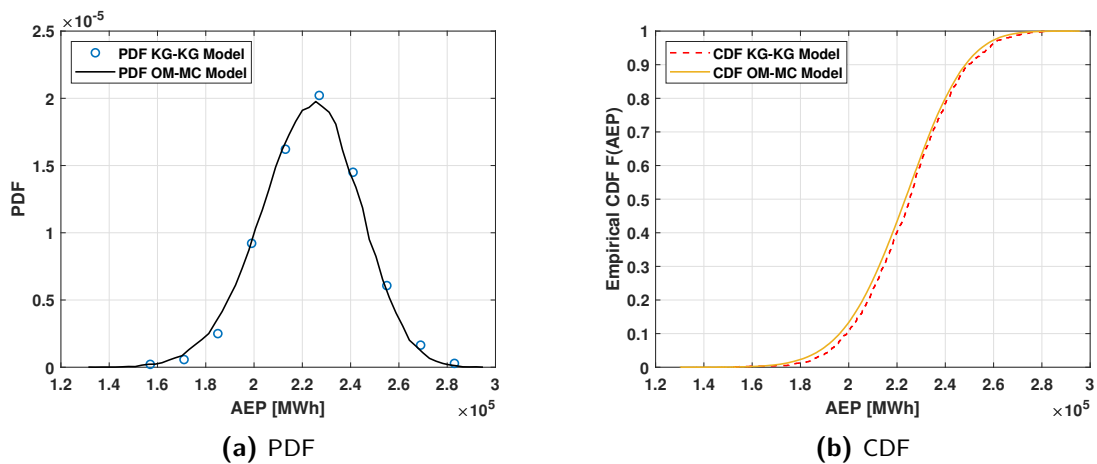


Figure 5-14: PDF and CDF of the AEP value using the KG-KG combination.

Table 5-3 gives a summary of the selected configurations for all combinations. In general, the computational time for every combination depends heavily on the power output surrogate modelling. Further, Figure 5-15 displays the percentage error of the estimated QoI and the standard deviation using different combinations. The estimation on the P50 and P90 value of AEP from all combinations can reach within 2% for P90 and even less than 1% for P50. The estimations of the P50 value of AEP using PCE as the surrogate model for wind farm power yields percentage errors within 1%, which are considered to be very accurate. The same behaviour can be seen when using the Kriging model for wind farm power evaluation.

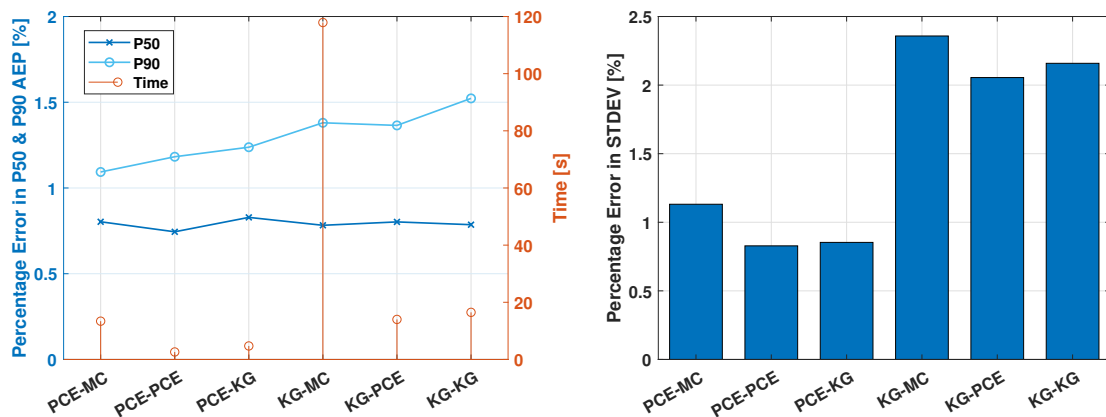
	Number of samples for UQ	P50 AEP [MWh]	P90 AEP [MWh]	Time [s]
OM-MC	100,000	$2.2340 \cdot 10^5$	$1.9623 \cdot 10^5$	217.0501
PCE-MC	1,000	$2.2519 \cdot 10^5$	$1.9837 \cdot 10^5$	13.39
PCE-PCE	100	$2.2506 \cdot 10^5$	$1.9854 \cdot 10^5$	2.61
PCE-KG	100	$2.2525 \cdot 10^5$	$1.9865 \cdot 10^5$	4.41
KG-MC	1,000	$2.2514 \cdot 10^5$	$1.9893 \cdot 10^5$	117.89
KG-PCE	100	$2.2519 \cdot 10^5$	$1.9890 \cdot 10^5$	14.00
KG-KG	100	$2.2515 \cdot 10^5$	$1.9921 \cdot 10^5$	16.48

Table 5-3: Summary of the performance of every UQ surrogate model.

On the other hand, The estimation of the P90 value of AEP varies per cases. While PCE-MC and PCE-PCE yield the most accurate P90 approximations, KG-PCE and KG-KG both give a relatively large percentage error compared to other cases. The time needed to construct and evaluate the combination set-ups is also investigated. It should be noted that PCE-MC and KG-MC use 1,000 samples, which is the reason that KG-MC yields a relatively high computational time. Despite that PCE-MC does not require that much computational time, it still takes longer than the other cases that use PCE as the wind farm evaluation model.

Next to that, the percentage error regarding the standard deviation of AEP via all six combinations is presented. PCE-PCE gives the best standard deviation estimation when compared to other combinations. The optimal configuration depends on the efficiency of the model that takes the time effort and accuracy into account. As shown in the previous PDF and CDF figures, PCE-MC and KG-MC combinations provide satisfying estimations on both P50 and P90 value of AEP. Nevertheless, both PCE-MC and KG-MC show the results from 1,000 uncertain input samples, and when talking about reducing the size of samples, PCE-PCE and KG-PCE are stable in terms of accuracy regarding the different number of uncertain samples. In general, it is possible to agree that PCE-PCE outperforms all other combinations with its fast computational time along with its accuracy.

According to the results, PCE-PCE could be a suitable UQ model to carry out the following study on the WFLO due to its fast calculation with adequate accuracy. The reason is that when the accuracy from all combinations are relatively similar, the time required to construct and evaluate the model becomes essential. The small differences in the required time enlarge when implementing the model in an iterative simulation, such as WFLO.



(a) The percentage error of the estimated P50 and P90 value of AEP from different combinations (b) The percentage error of the estimated AEP standard deviation from different combinations

Figure 5-15: The percentage error of the estimated statistical quantities from different combinations.

5-5 Sensitivity analysis on the uncertain parameters

5-5-1 Introduction

Sensitivity analysis is a study that is closely related to the uncertainty analysis and aims to study the variability of the input variables or their combined effect on the model response. It is also useful as a method to conduct model refinement [4] for a mathematical model or a numerical system that has a large number of input variables. For example, there are in total 13 uncertain inputs for the UQ models as shown in Table 5-1. Assessment of the variables could be done by sensitivity analysis, especially so that the most influential variables can be identified and studied. The remaining parameters can be neglected to make the model into a lower-dimensional one.

Various methods of sensitivity analysis have been developed over time. One-at-a-time (OAT) is the simplest and most common methods [41]. The essence of such a method is to alter one input variable while keeping others the same value. The next step is to return the altered variable to its original value and repeat the process for every other input. OAT is often used to investigate the correlation of the input and output. In the next section, the Weibull scale and shape parameters along with wind rose sector probabilities are the varying input variables to be tested. Even though OAT is easy for implementation, the approach has a drawback of not exploring the entire input space since no combination effect from multiple input variables is taken into account.

5-5-2 The effect of Weibull parameters

As stated in the previous sections, wind speed distribution is commonly approximated by the Weibull distribution. In this research, the 2-parameter Weibull distributions with scale and shape parameters have been implemented to calculate the power output and the AEP of the investigated wind farm. In order to see how exactly the Weibull scale parameter, λ , influences

the simulation result, the other parameters are fixed as the method OAT stated. Figure 5-16a shows the PDF of the scale parameter along with the expected AEP calculated with the original model using different values of the Weibull scale parameter. From the result, the Weibull scale parameter exhibits a positive correlation with the AEP value. In other words, with a larger scale parameter, a wind farm could yield larger AEP when other parameters are fixed. This can be explained by the fact that the Weibull scale parameter is proportional to the mean wind speed at a site [27].

On the other hand, it is also possible to observe how the Weibull shape parameter, k , influence the results of the simulation. The result is presented in Figure 5-16b, and it shows that the Weibull shape factor also has a positive correlation with the AEP value. According to both cases, it is possible to suspect that the Weibull scale parameter poses a much larger influence on the AEP value as the y-axis of the two figures differs in range.

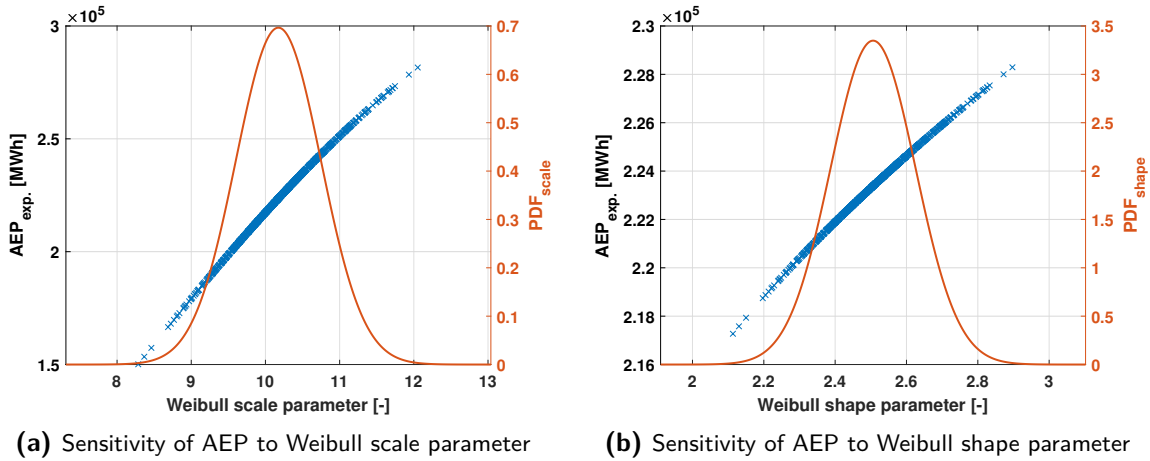


Figure 5-16: Sensitivity of the original model to Weibull parameters.

Until now, it has acknowledged that the AEP value is clearly dependent on the values of both Weibull parameters. Inspired by Lackner et al. [27], Figure 5-17 is created. Instead of using AEP as the output parameter, the capacity factor is used and is defined by:

$$CF[\%] = \frac{AEP}{P_{rated} \cdot N_{WT} \cdot T_{yr}} \quad (5-15)$$

where P_{rated} is the rated power of the wind turbine model, N_{WT} is the number of wind turbines in the wind farm, and T_{yr} is the total hours in a year.

The figure shows the ranges of values of the Weibull scale and shape parameters from the 50 years measurement. As seen from the result, the capacity factor increases as the scale parameter increases, which means AEP also increase. This corresponds to what has been seen in Figure 5-16a. The same behaviour can be seen for the Weibull shape parameter except for the magnitude of increment in capacity factor as the shape parameter increases is not as large as the case for the scale parameter. This can be understood by the effect of shape parameter on the Weibull distribution, where a higher shape parameter makes the distribution more condense than spread out. Therefore, the dependence of the capacity on the shape parameter highly depends on the scale parameter. It can be seen from the figure that when the scale parameter is high, the increase in the shape parameter can result in a

higher increment in the capacity factor. In contrast, the case for lower scale parameter does not have the same magnitude of increment.

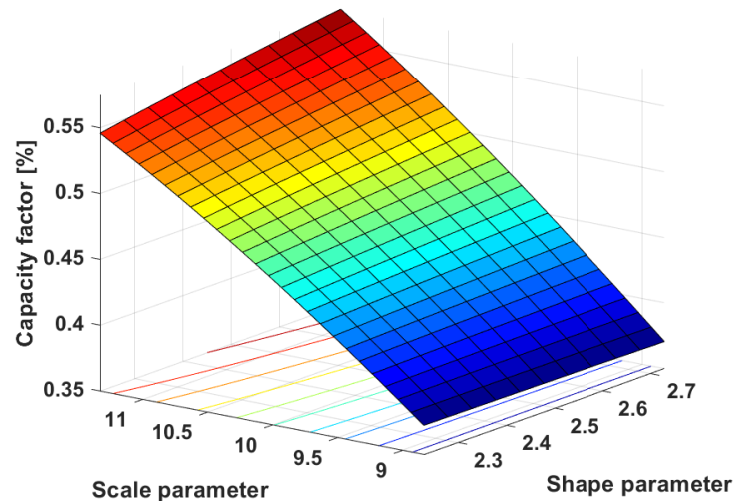


Figure 5-17: Sensitivity of capacity factor to Weibull scale and shape parameters.

5-5-3 The effect of sector probabilities

The analysis on the sector probabilities follows after the analysis of the Weibull parameters. The sector probabilities influence the wind farm yield when there exist multiple wind turbines, and wake effects play a role. However, while the OAT technique works well for the Weibull parameters, it is not the same for the sector probabilities. The reason is that the total probability from all sectors needs to equal one. Due to this reason, sector probabilities are dependent to some degree.

In this case, an alternative method is proposed. As displayed in Figure 2-2 and Figure 5-3b, there is a dominant wind sector, where the wind direction is within $[210^\circ, 240^\circ]$. The sensitivity analysis thus focuses on this specific sector to study how this sector affects the wind farm yield. The approach has a rationale that as the research aims to study the inter-year uncertainty, the most dominant sector prevails other sectors on the probability of occurrence. When this sector has a larger probability, the other sectors will have smaller probabilities and vice versa. Figure 5-18 displays three different estimations of AEP when the dominant wind direction sector has the largest, mean, smallest probabilities of the 50 measured years. The results show that the differences in these three scenarios are not as large as what has been shown for the Weibull distribution parameters but rather small. It is suspected that the sector probabilities do not have the same influence on the wind farm yield as the Weibull scale factor has.

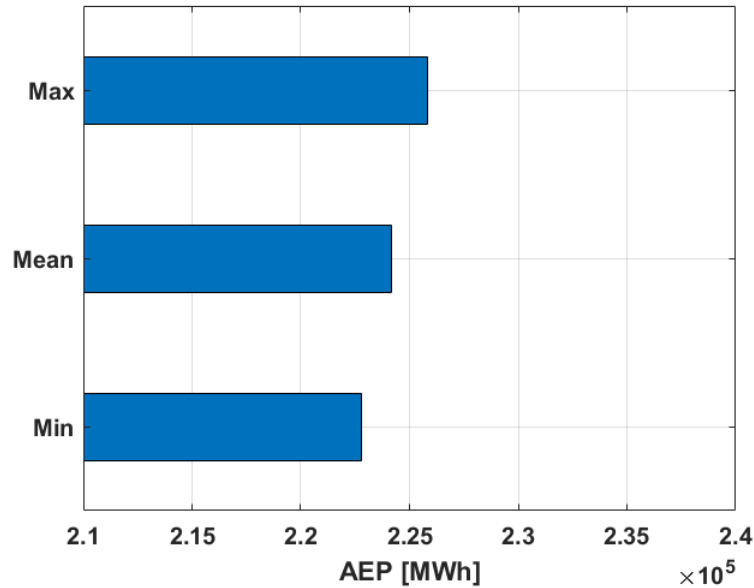


Figure 5-18: Sensitivity of the output response (AEP) to the dominant wind direction sector probability.

5-6 Summary of the chapter

In this chapter, a background overview of the uncertainty quantification for the study has been discussed. The uncertain variables are the year-to-year variation in the wind resource, namely, the Weibull scale and shape parameters, and the wind direction sector probabilities. The research aims to find the P90 value of AEP when taking these uncertainties into account. The propagation of uncertainty is crucial for uncertainty quantification, and the Monte Carlo method and surrogate models have both been implemented to study the accuracy and the computational effort in terms of time required. The Monte Carlo method has shown its strong and weak points that it is simple to implement but computationally heavy compared to the surrogate models introduced here. When using the surrogate models to propagate the uncertainty, the number of samples needed from the uncertain variables drastically decreased while keeping the percentage error with respect to the ground truth model relatively small. A sensitivity analysis has also been conducted to see which uncertain variables affect the wind farm energy yield in the most significant way. It turns out that the Weibull scale parameter has the most extensive influence on the wind farm energy yield.

Wind Farm Layout Optimisation

6-1 Introduction to the chapter

This chapter presents the fundamental knowledge on the optimisation of wind farm layouts. It further applies the uncertainty quantification and surrogate modelling techniques from previous chapters within the wind farm layout optimisation (WFLO). As a consequence, the optimisation problem develops into an optimisation under uncertainty (OUU) problem. The main optimisation algorithm used in the chapter is the genetic algorithm (GA), and two different kinds of it are implemented, namely, real-coded genetic algorithm (RGA) and binary genetic algorithm (BGA). Lastly, the results from the deterministic approach and OUU are discussed in the later sections.

6-2 Optimisation problem definition

As the demand for renewable energy rises, the scale of the wind farm also increases. Due to the substantial investment capital of a wind farm, the optimisation of the wind farm layout is starting to show its important aspect. First, some background knowledge of optimisation theory needs to be reviewed. As stated in Pérez et al. [44], there are four basic elements to start any optimisation problem, which are as follows:

1. *Objective function*

The objective function presents how each variable contributes to the optimisation problem. It also shows whether a decision is good or bad. With different considerations on economic aspects of wind farm development, different objective functions can be chosen. The most widely used metrics in wind energy-related fields are the AEP, LCOE, and NPV [20]. A detailed explanation of AEP has been presented in Section 2-2. Regarding the cost of energy (COE), it requires more information to calculate the value, including AEP and can be expressed by:

$$COE = \frac{C_{Farm}}{AEP_{Farm}} \quad (6-1)$$

where C denotes the total cost of the generation of energy for a year, and it's divided by the total energy generated in a year. It can be defined as the cost per unit of energy and shows either the wind farm has economic profit or loss over its operational lifetime [20]. The definition of the cost has many different approaches. Many research works have reviewed the cost function proposed by Mosetti et al. [36]:

$$C = N_T \left(\frac{2}{3} + \frac{1}{3} e^{-0.00174 N_T^2} \right) \quad (6-2)$$

in which the cost function assumes that the total cost of the wind farm is a function of the number of wind turbines N_T in the farm. It should be noted that the cost function by Mosetti et al. [36] is meant to try-out the optimisation algorithm rather than optimising a real wind farm.

Furthermore, levelised cost of energy (LCOE) shares a similar definition with COE but has a clear difference between the fixed and variable costs, which can be expressed as

$$LCOE = (C_c * FCR + C_{O\&M}) / AEP \quad (6-3)$$

where C_c is the total installed capital cost of the wind farm, and FCR is the fixed charge rate, which includes information about the taxes, insurance, and financing [28]. The parameter $C_{O\&M}$ refers to the annual operation and maintenance cost for the wind farm.

Nevertheless, when more information is taken into account, the uncertainty within the whole calculation increases. To lower the uncertainty and restricted it to the wind resource, the maximum power production is chosen to be the optimisation objective. Therefore, the objective function of this research is still AEP, which should be maximised.

2. Problem variables

In a WFLO problem, the problem or design variables are often the locations of the wind turbines because the locations of the turbines could heavily influence the wake effect and foundation costs. The problem variables are often represented by vectors and can be integrated into a matrix \hat{X} as:

$$\hat{X} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_N & y_N \end{bmatrix} \quad (6-4)$$

where x and y denote the exact location in x -coordinates and y -coordinates of each turbine from the first one to the N th one.

3. Constraints

In an optimisation problem, it is crucial to set some constraints to the problem variables. Otherwise, the problem variables become unbounded, and the optimiser will be searching in an unlimited domain. Constraints could be set due to the availability of area and safety factors. The most common constraints are the minimum distance between

turbines, and the wind farm area limits [44]. Once the constraints are not violated, the problem variables are considered since they are within the feasible region. Based on the wind farm area limits, certain bounds can be applied to the problem variables, which are the wind turbine coordinates. As shown in Figure 2-5, the wind farm has an area of 3400 [m] x 1000 [m]. The x coordinates in Eq. (6-4) are limited to a lower bound of 0 and a higher bound of 1000. In the meantime, the y coordinates are appointed to the range of 0 to 3400.

The minimum distance between the wind turbines is another constraints to the optimisation problem. It is set due to safety reasons as the distance is limited to be larger than two rotor diameters. It can be mathematically expressed as

$$S_{i,j} = |WT_{i(x,y)} - WT_{j(x,y)}| > 2 \cdot WT_D, \quad \text{for } \begin{cases} i, j = 1 \dots N \\ i \neq j \end{cases} \quad (6-5)$$

in which S is the distance between the wind turbines (WT), WT_D is the rotor diameter, x and y are the coordinates of the turbines, and N is the number of turbines.

4. Data

In order to make an optimisation problem, it is essential to have the information needed to calculate the aforementioned objective function and constraints. For example, the wind resource data mentioned in the previous chapters are required. As mentioned before, 50 years of measurement data from KNMI's IJmuiden weather station is utilised. Secondly, the wind turbine data is also needed. The wind turbine model used here is the NREL 5MW model, and the hub height, rotor diameter, thrust coefficient and the power curve can be found in Jonkman et al. [23]. Moreover, the wake effect model follows the one in Jensen [21] and Katic et al. [24] where a suggestion of the decay factor k can be found. Finally, the number of wind turbines in the wind farm and a reference location of them is necessary for the wind farm data.

The current optimisation problem can be expressed in the mathematical notation as:

$$\text{maximise} \quad AEP(\hat{X}) \quad (6-6a)$$

$$\text{subject to} \quad S_{i,j} \geq 2D \quad \forall i, j = 1, \dots, 10 \quad \forall i \neq j \quad (6-6b)$$

$$0 \leq x_i \leq 1000 \quad \forall i, j = 1, \dots, 10 \quad (6-6c)$$

$$0 \leq y_i \leq 3400 \quad \forall i, j = 1, \dots, 10 \quad (6-6d)$$

where $S_{i,j}$ denotes the distance between the two wind turbines as shown in Eq. (6-5), D is the rotor diameters, and x_i and y_i are the coordinates of the wind turbines. Even though using a single straightforward objective function is not the best function for a WFLO problem [53], it is still decided to use AEP as the sole objective function since none of the construction uncertainties was considered in the research.

6-3 Optimisation algorithms

6-3-1 Overview

Several optimisation algorithms have been developed to maximise or minimise the objective function. The work by Mosetti et al. [36] was the first one to use a genetic algorithm (GA) in

the field of WFLO to minimise the COE under three different wind conditions. Following this work, quite an amount of studies looked into the problem using either the same algorithm or others. Typically, the algorithms can be categorised into deterministic and heuristics (i.e., random search-based) [3, 45]. In general, the differences are the ability to find the optimal solutions (i.e., the quality of the solution) and the required computational costs, where the deterministic approach yields a faster but lower quality optimisation result.

The research decides to follow the steps of Mosetti et al. [36] and used GA as the choice of the algorithm after studying different optimisation algorithms and surveying the toolbox in hand. Based on the domain of search space, different kinds of GA have been developed. The following paragraphs give an introduction to two different kinds of GA, namely, binary genetic algorithm (BGA) and real-coded genetic algorithm (RGA). The differences between them will be pointed out in later sections.

6-3-2 Introduction to genetic algorithm

Developed by John [22], genetic algorithms are population-based and are an analogy to the Darwinian-fitness theory of natural selection. The design parameters of the optimisation problem are defined as genes and transformed into arrays of bits or character strings like chromosomes. The selection of the solution is based on the fitness of the objective function. If the criteria are not met, a new population is created by performing selection, crossover, and mutation to replace the old population. As evolution goes on, the algorithms produce more generations. Each generation evaluates its fitness, and once the criteria are met, an optimum solution is obtained from the last generation. Figure 6-1 shows a simplified flow chart of the algorithm, and as presented, GA relies heavily on its operators, which can be summarised as follows:

- **Initialisation**
An initialised population needs to be defined first for GA. The population can be generated randomly or by user-defined values. The size of the population also needs to be taken into account since it can severely increase the computational cost or constrain the solution in a local extremum (i.e., minimum or maximum) if the size is too small.
- **Selection**
In GA, the concept of generation is utilised, and thus the term "parents" and "children" emerged. Within an existing population, a portion of them is chosen to be the parents. The selection is based on the fitness to the objective function. The fitter they are, the more likely they will be chosen to generate the next generation, which is the children. A common option for the selection process is the roulette wheel selection, where the probability of being selected is proportional to the fitness of every chromosome in the population [16].
- **Crossover**
Once the selection is completed, a crossover operation is applied to create the next generation. There are different methods to perform the crossover, for example, single-point crossover and uniform crossover. In a single-point crossover, two chromosomes from the selection process, which act as the parents, will have a randomly chosen crossover

points, where the gene (or string) before and after that crossover points will exchange in between the two chromosomes. As a consequence, it emerges two new chromosomes.

- Mutation

Besides the crossover operation, the mutation is also a critical operator in GA. The operator randomly alters one or more values in a chromosome. It does not occur in every creation of a new generation, but it is an essential operation in GA. It ensures the evolution occurs in the algorithm and does not trap in the local solution or extremum.

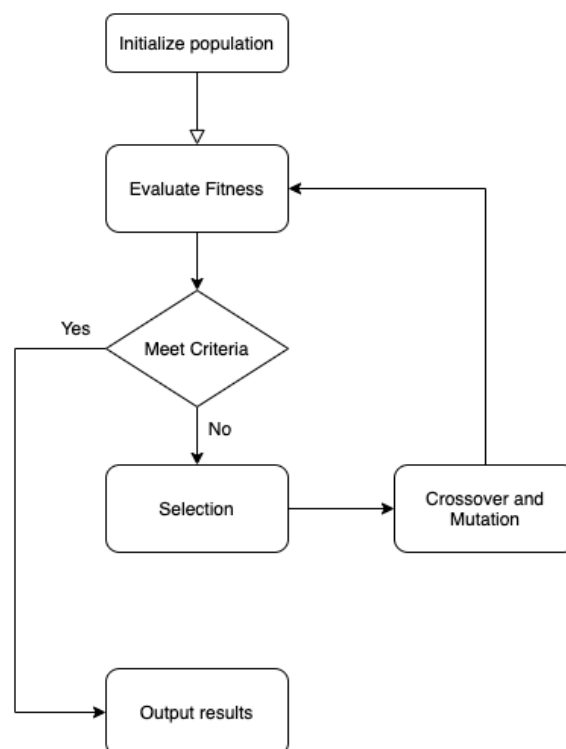


Figure 6-1: An overview of the steps in a genetic algorithm.

6-3-3 Binary genetic algorithm

In binary genetic algorithm (BGA), the search space is discretised into cells. A cell with a wind turbine inside is coded as 1, and a cell with no turbine is coded as 0. In this case, the entire search space is coded into a chromosome with a binary gene. For example, for a design space of 3 [km] x 3 [km], it can be divided into a discrete space of 3 x 3 cells with each cell representing 1 [km] x 1 [km]. In this case, it could result in 2^9 possible layouts and can be coded as a chromosome with nine binary numbers. Figure 6-2 illustrates the coding process of such a layout.

In general, BGA is easier for the study with varying number of turbines since the mutation and crossover could create more or fewer turbines within the search space. These genetic operators make the number of turbines changeable during the optimisation. However, since the search space is discrete, BGA may not be able to find the most optimal solution and the

fitness may differ based on how fine is the grid. It is possible to have a better result with a finer grid, but the computational cost increases. In general, BGA is the easiest type of GA, and if the demand for accuracy is not high, the required computational power is usually lower than other kinds of genetic algorithms.

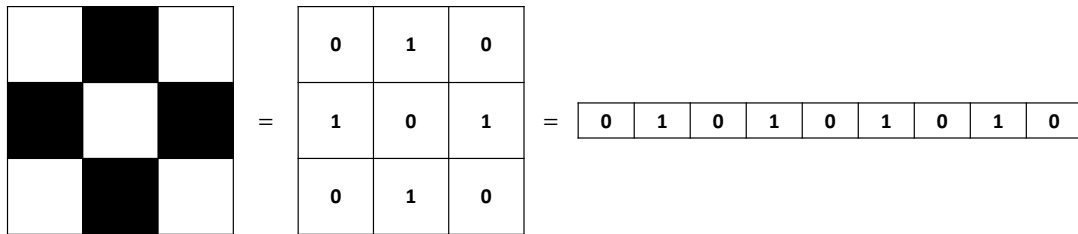


Figure 6-2: Binary representation of the chromosome from a wind farm layout. The black squares and 1's denote the cells are occupied by wind turbines. The white squares and 0's denote empty cells.

6-3-4 Real-coded genetic algorithm

In contrast, real-coded genetic algorithm (RGA) has continuous coordinates of turbines as the variant in the searching space instead of binary encoding in discretised cells. Besides this difference, other elements within the genetic algorithm are the same. The same operators are implemented to search for the optimal solution.

However, it is worth noting that RGA is easier for the study of a fixed number of turbines but difficult to simultaneously optimise the number of turbines and locations as the size of the design vector pre-defines the number of turbines. RGA can be more complicated due to its continuous coordinates, which can easily result in the violation of optimisation constraints, e.g. minimum distance between two wind turbines. This constraint is not a problem for BGA since the cell itself can already create separation for turbines. As already mentioned, the constraint of the distance between any two wind turbines follows Eq. (6-5), where it should be at least larger than two times of rotor diameter due to safety reason and most importantly, to avoid excessive wake effect.

6-3-5 Optimisation toolbox options

There are many options within the Global Optimisation toolbox in MATLAB. Based on the documentation, some options of the toolbox are explained. The following paragraphs introduce the options that are chosen to conduct wind farm layout optimisation for this research.

First, it must be emphasised that the toolbox treats an optimisation problem with integer constraints differently. When integer constraints are specified, such as is the case for BGA, the toolbox has its genetic operator options to ensure that the integer constraints are honoured. The detail of its modified setting can be found in Deep et al. [8]. When a user tries to define the options, the optimiser may provide a result that fails to meet the constraints. Therefore, the following optimisation toolbox options are mainly for RGA.

The selection option determines the way of picking parents for the next generation. The tournament selection is a commonly seen option as it picks individuals from a population to be selected for crossover operation. It involves having several tournaments for a few individuals that are randomly chosen from the population. As parents, each candidate is compared based on its fitness value against each other. The winners of each tournament are therefore select for the crossover to produce the next generation. Other than that, the remainder selection assigns parents based on a given scaled value, such as 2.3. Depending on the integer part of the scaled value, an individual is listed that many times as a parent. The rest of the parents are chosen arbitrarily besides the prescribed ones. The fractional part of the scaled value then determines the probability that a parent is selected. It is suggested by Bailleul [3] that the remainder selection can result in a faster calculation time in RGA even though this might be case dependent.

Once the selection is made, the crossover operator follows and also has its options. The crossover option determines the way that the algorithm combines the parents to form a crossover child for the next generation. The scatter crossover is based on a randomly created binary vector with the same length as the parents. The crossover selects the genes from the first parent, where the vector is an '1', and the genes from the second parent, where the vector is a '0'. The child is created based on the combination.

The last option is the mutation option, which defines the way the algorithm alters the individuals in the population to create mutation children. The uniform mutation assumes that a fraction of the gene has a probability of mutating. The selected ones are replaced by random values that are selected uniformly from the range of that gene. An overview of the chosen options for the GAs is presented in Table 6-1.

Genetic Algorithm inputs	WFLO options	
Types of GA	BGA	RGA
Design variables	Grid numbers	Turbine coordinates
Selection function	-	Remainder
Crossover function	-	Scatter
Mutation function	-	Uniform
Fitness function	AEP	AEP
Constraint tolerance	10^{-3}	10^{-3}

Table 6-1: Parameters settings for the genetic algorithms.

6-4 Deterministic optimisation

Deterministic optimisation is named deterministic because no uncertainty in the Weibull parameters and wind sector probabilities are involved. In the deterministic optimisation, the Weibull parameters are the mean values from the 50 years measurement, and so do the wind direction sector probabilities. As mentioned before, the Weibull parameters are not defined by sectors but aggregated ones for all wind sectors. The AEP is, therefore, a deterministic value.

Both kinds of GA are used to study the WFLO and are implemented in MATLAB. In the optimisation using BGA, the wind farm of 3400 x 1000 m^2 is divided into 17 x 5 square cells

(in total 85 cells of $200 \times 200 \text{ m}^2$), and the wind turbine is placed in the central point of each cell. RGA shares the same constraint dimensions of the search space as BGA, but without the constraint of grid points. The numbers of turbines for both GAs is fixed at ten turbines.

6-4-1 Optimisation with the original model

The first deterministic optimisation is the optimisation with the OM model. The wind speed is fixed at 10 [m/s] and the wind only comes from 225 [deg], which is the dominant wind direction of the investigated site as the wind rose shown in Figure 2-2. This restriction is set to test the feasibility of the optimisation algorithm with regard to the computational effort. The AEP is then computed with this set of wind speed and wind direction. There are in total of four scenarios that are studied. Besides looking into different types of genetic algorithm, the effect of an initial layout is also investigated. One set of scenarios gives an initial layout that is the grid layout, as shown in Figure 2-5 (case 1 in Chapter 4). It should be underlined that the initial layout is acting as one of the individuals in the first generation and the other individuals are self-generated by the optimiser. It is a well-designed layout by the author, which guarantees each wind turbine has a sufficient distance with other turbines to avoid excessive wake effects. On the other hand, the other set of scenarios allows the optimiser to self-generate the entire first generation for an initial layout.

The resulting optimised layouts are shown in Figure 6-3. If an initial layout is given, it is denoted as the red circles in the figures. As the dominant wind direction is from the southwest direction, most of the optimised layouts follow the general pattern of avoiding aligning the turbines in that southwest to northeast direction. This is reasonable as this alignment leads to severe wake effects.

	Ge. ¹	AEP [MWh]	Efficiency[%]	Time [min.]
Ref. layout free-stream	-	$5.4652 \cdot 10^3$	-	-
Ref. layout	-	$4.4788 \cdot 10^3$	81%	-
BGA	78	$5.4652 \cdot 10^3$	100%	2.47
RGA	55	$5.4652 \cdot 10^3$	100%	3.98
BGA Init.	66	$5.4652 \cdot 10^3$	100%	2.91
RGA Init.	62	$5.4652 \cdot 10^3$	100%	2.96

Table 6-2: Summary of the optimisation results regarding different optimisation methods with the original model. Reference (Ref.) layout is the grid layout in Figure 2-5 and is used as the initial layout for the optimiser. Init. denotes the case is given an initial layout.

A summary of the optimised results is presented in Table 6-2. According to the results, the optimised AEP are mostly the same for all four scenarios. The efficiency in the table is defined as the ratio of the wind farm produce, and the wind farm would produce if no wake effect exists, which is the AEP by the Ref. layout free-stream in the table. Based on the table, the reference layout reaches an efficiency of 81%. For this single wind speed and wind direction case, the optimiser is able to find a highly efficient layout with no wake losses. In addition, as the optimised layouts in Figure 6-3 are all different, it can be concluded that there are a lot of local optima with similar performance. This is a similar finding to Padrón et al. [42].

¹Generations until termination

Nevertheless, the highly efficient optimised layouts are caused by the simplification of having only one wind direction. The simplification makes it possible to find layouts where turbines are not in the wakes of other turbines and resulting in an 100% efficiency.

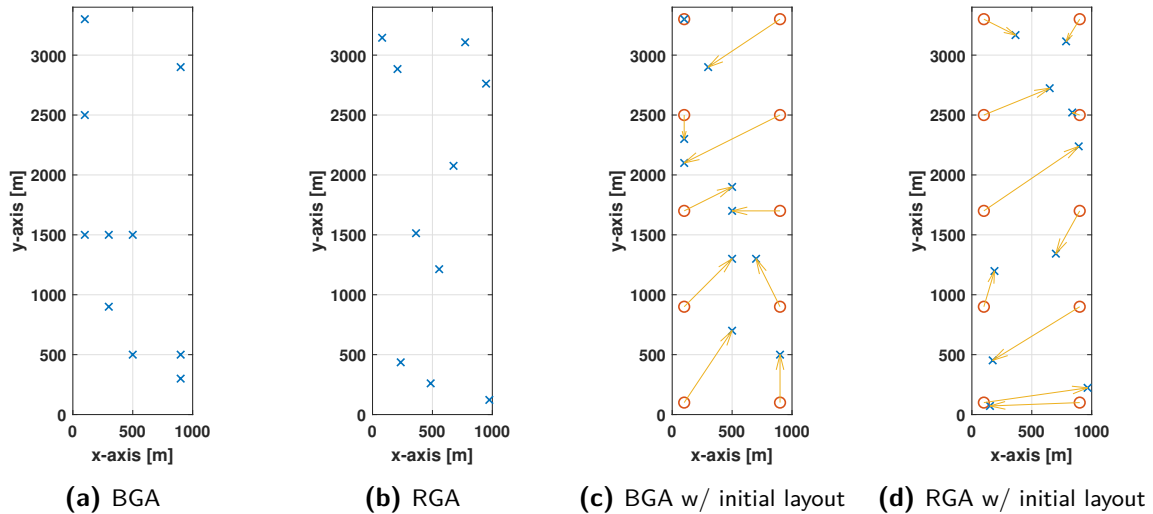


Figure 6-3: Deterministic optimisation of wind farm layout with original model. The blue crosses are the optimised wind turbine locations and the red circles are the reference initial wind farm layout, if specified.

Instead of using one wind speed and one wind direction, the research also looks into the optimisation using the entire domain of wind speed and wind direction. It is, however, resulted in an extremely long computation and was not able to continue with the computational power in hand. As the medium-fidelity OM model evaluates a wind farm layout in around 4 minutes for every evaluation, the genetic optimiser undergoes a great computational effort if a large generation and population are specified. Representative results are commonly coming from a sufficient number of generation and population, and thus, the long computational time for this case is not avoidable.

6-4-2 Optimisation with the surrogate models

It has been mentioned that optimisation with a full OM model leads to a substantial computational effort. As has been shown and verified in Chapter 3, surrogate models can help to reduce the computational effort in terms of computational time. The optimisation with the surrogate models utilises PCE model instead of the Kriging model. This is because it has been shown that the approximation using the PCE is slightly better, and the time required to construct the PCE model is also shorter.

As already mentioned in Chapter 3, there are two sampling schemes to make surrogate models. The two schemes are expected to influence the results of the optimisation. Moreover, the randomness inside the random sampling scheme, Latin hypercube sampling (LHS), has proven to add some additional uncertainty to the approximation by the surrogate models. This has been presented in Chapter 4 and it has been proposed to use pre-defined (PD) samples that are generated via Latin hypercube sampling to eliminate this additional uncertainty.

Nevertheless, it is decided to test a smaller size of generations and population of the genetic algorithms for the sake of easiness of investigation. Therefore, the number of generations and the population are both set to 20.

The differences between the two LHS (R & PD) schemes are presented in Table 6-3 and Table 6-4, where the two tables are based on whether an initial layout is given to the optimiser or not. However, the optimised results are only evaluated using the surrogate models. There may exist an approximation error between the AEP value from the surrogate model and the one from the original model. Hence, the optimised layouts are verified by the original model to see if they truly improve and are not affected by the approximation error of the surrogate models. The original model is used to calculate the true AEP of the optimised layout and the real improvement (denoted as A. Imp. in the table) is checked to see if the layout indeed improves. The original model calculates $2.2348 \cdot 10^5$ [MWh] of AEP for the reference grid layout and the improvements (denoted as Imp. in the tables) are based on this value. For cases in Table 6-3, there is no initial layout given to the optimiser, so the improvement is just a reference value to see if the optimised layout surpasses the reference grid layout.

	Ge	AEP [MWh]	Imp. [%]	A. AEP ² [MWh]	A. Imp. ³ [%]	Time [hr.]
BGA R	20	$2.1905 \cdot 10^5$	-1.98%	$2.1770 \cdot 10^5$	-2.59%	0.5
BGA PD	20	$2.1941 \cdot 10^5$	-1.82%	$2.1727 \cdot 10^5$	-2.78%	0.55
RGA R	20	$2.1970 \cdot 10^5$	-1.69%	$2.1607 \cdot 10^5$	-3.32%	0.4
RGA PD	20	$2.2136 \cdot 10^5$	-0.95%	$2.1928 \cdot 10^5$	-1.88%	0.5

Table 6-3: Summary of the optimisation results regarding different optimisation methods with surrogate models and a self-generated layout. R in the first column denotes random sampling scheme by LHS, and PD means there is a pre-defined set of LHS samples.

First, the randomness from the samples is inspected. According to the results in Table 6-3, if the model does not have a pre-defined sample set, the existing randomness could lead to a larger difference between the improvement by the surrogate models and the actual improvement by the original model. This is also a finding in Section 4-3. It was found that not having a pre-defined sample set could lead to additional uncertainty on the estimation of approximately 1%. It is also noteworthy that the randomness also confuses the optimiser. Figure 6-4 presents the convergence history of the BGA optimisation. The best fitness in Figure 6-4a does not show a step-by-step improvement as in Figure 6-4b but rather an unstable one as the generation increases. On the other hand, it is reasonable to see this unstable improvement in the mean fitness as the population of each generation is performing the genetic operators to find a better generation.

Similar to this finding, Table 6-4 shows another finding. It is found that the random LHS scheme could result in a negative improvement even when an initial reference layout is given. Figure 6-5 shows the convergence history of the BGA optimisation when an initial layout is given to the optimiser. As all methods failed to find a better layout than the given initial layout, the difference between the surrogate improvement and the actual improvement indicates the approximation error by the surrogate model. Nevertheless, the convergence history in Figure 6-5a clearly shows that the randomness from the sample set is confusing the optimiser as the initial layout is still the optimised layout after these 20 generations. This is a clear

²Actual AEP

³Actual improvement

evidence that the samples from the wind speed and wind direction matter to the accuracy of the approximation. Until now, it has been shown that using a pre-defined sample set is better for the optimiser.

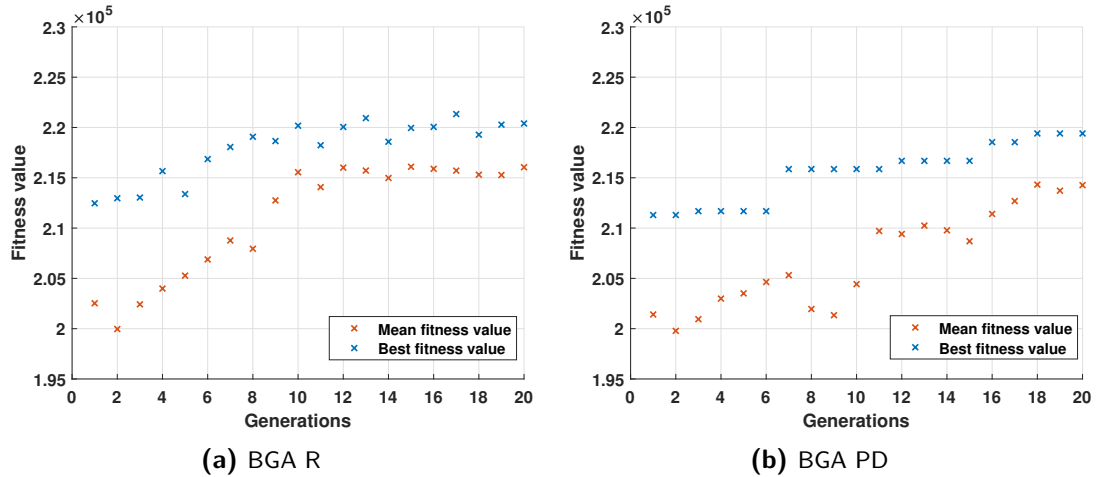


Figure 6-4: Convergence history of the BGA optimisation with the surrogate model. No initial layout is given to the optimiser.

	Ge	AEP [MWh]	Imp. [%]	A. AEP [MWh]	A. Imp. [%]	Time [hr.]
BGA R	20	$2.2157 \cdot 10^5$	-0.85%	$2.2348 \cdot 10^5$	0%	0.57
BGA PD	20	$2.2538 \cdot 10^5$	0.85%	$2.2348 \cdot 10^5$	0%	1.21
RGA R	20	$2.2412 \cdot 10^5$	0.29%	$2.2348 \cdot 10^5$	0%	0.5
RGA PD	20	$2.2538 \cdot 10^5$	0.85%	$2.2348 \cdot 10^5$	0%	0.41

Table 6-4: Summary of the optimisation results regarding different optimisation methods with surrogate models and a given initial layout. R in the first column denotes random sampling scheme by LHS, and PD means there is a pre-defined set of LHS samples.

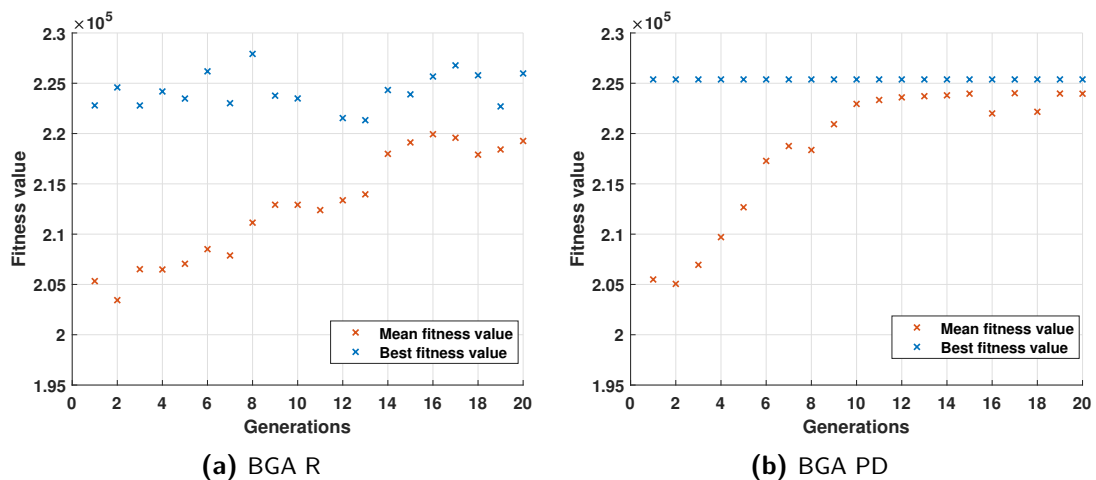


Figure 6-5: Convergence history of the BGA optimisation with the surrogate model. An initial layout is given to the optimiser.

Next, the research compares the random sampling scheme with the linearly spaced samples (LSS). Table 6-5 presents the optimised results from the optimisation and the verification by the original model. As the differences between the improvement from the surrogate models and the actual improvement are large, it considers that LSS is not a good choice of the sampling scheme. This could lead back to Chapter 3, where it was mentioned that LSS is not able to accurately sample enough conditions, and it results in a poor approximation.

	Ge	AEP [MWh]	Imp. [%]	A. AEP [MWh]	A. Imp. [%]	Time [hr.]
BGA LSS	20	$2.3811 \cdot 10^5$	6.55%	$2.1462 \cdot 10^5$	-3.96%	0.51
RGA LSS	20	$2.4126 \cdot 10^5$	7.94%	$2.1756 \cdot 10^5$	-2.65%	0.47
BGA LSS Init.	20	$2.4303 \cdot 10^5$	8.76%	$2.1851 \cdot 10^5$	-2.22%	0.54
RGA LSS Init.	20	$2.3183 \cdot 10^5$	8.21%	$2.2033 \cdot 10^5$	-1.41%	0.41

Table 6-5: Summary of the optimisation results regarding different optimisation methods with surrogate models using fixed linearly spaced samples. Init. denotes the case is given an initial layout.

Therefore, it has been verified that using a pre-defined sample set is a better approach to conduct optimisation, along with using RGA. Next, once the algorithm and the methods are clarified and decided, larger generations and population are applied to the optimiser. The number of generations and population is set to 100 in this case. The optimised layouts are again verified by the original model to see if they truly improve and are not affected by the approximation error of the surrogate models.

According to Table 6-6, the differences between the verification of OM model and the PCE model on the optimised layout are small and can be regarded as the approximation error of the surrogate model. The optimisation with a self-generated layout fails to find a better layout than the one that is hand-picked by the author. Nevertheless, it still verifies that the optimiser can find a better layout if more generations are used. This is concluded as the difference to the reference layout is only 0.3%, and it is considered that the optimiser starts from a layout that was generated randomly. On the other hand, the optimiser also does not find a layout with great improvement in AEP when an initial layout is given. It is possible that the initial layout could have been superior to most of the layouts in the genetic population, making the optimiser fail to find a better one. The optimised layouts are shown in Figure 6-6. In general, the cases that start with an initial reference layout could possibly lead to a better optimal result than the ones that start from a random layout. However, starting from a random layout (self-generated layout by the optimiser) makes it possible to explore more of the design space and therefore lead to different layouts instead of ones that are similar to the reference layout.

Method	AEP improvement w/ self-generated layout		AEP improvement w/ initial layout	
	PCE	OM	PCE	OM
RGA	0.77%	-0.30%	1.31%	0.14%

Table 6-6: The improvement in AEP of the optimised layouts using RGA with respect to their starting layout.

So far, several findings on the optimisation algorithm can be concluded based on the optimisations. From the theory of genetic algorithms, BGA is faster than RGA to reach a certain

generation, which could mean less computational power is needed for BGA. This is not yet observed from the results by this point as the generations are still small. In contrast, RGA is able to provide a better improvement than BGA with a limited number of generations. This is because the search domain for RGA is not limited to the pre-defined grid points as in BGA. These findings will be rechecked to see if they still hold when the optimisation is conducted with surrogate models and UQ models.

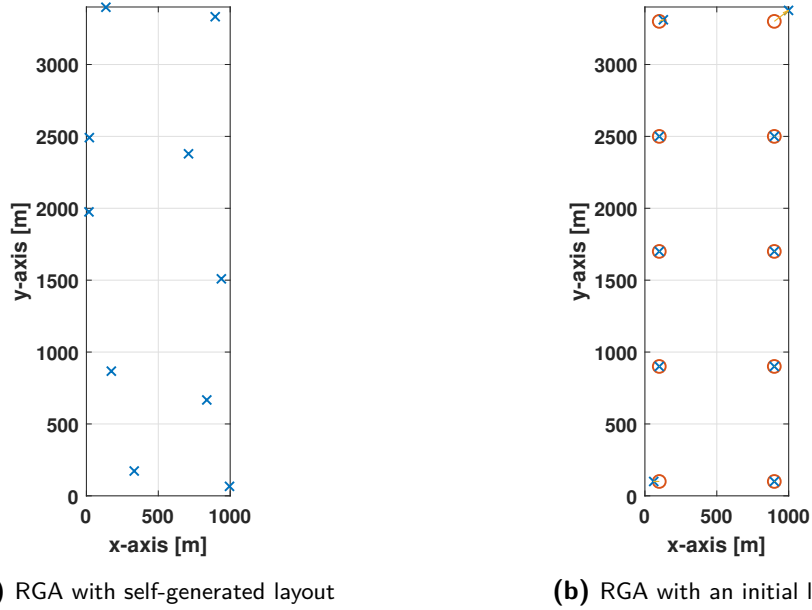


Figure 6-6: Deterministic optimisation of wind farm layout with surrogate models. The blue crosses are the optimised wind turbine locations and the red circles are the reference initial wind farm layout, if specified.

6-5 Optimisation under uncertainty

Different from the deterministic optimisation, the optimisation objective is now a function of turbine locations, \hat{X} , and uncertain variables, ξ . Here, the uncertain variables are the ones that are mentioned in Chapter 5, namely the Weibull parameters and the wind sector probabilities. The probability distributions have been given in Table 5-1.

The optimisation under uncertainty problem can be expressed in the mathematical notation as:

$$\text{maximise} \quad P_{90} AEP(\hat{X}, \xi) \quad (6-7a)$$

$$\text{subject to} \quad S_{i,j} \geq 2D \quad \forall i, j = 1, \dots, 10 \quad \forall i \neq j \quad (6-7b)$$

$$0 \leq x_i \leq 1000 \quad \forall i, j = 1, \dots, 10 \quad (6-7c)$$

$$0 \leq y_i \leq 3400 \quad \forall i, j = 1, \dots, 10 \quad (6-7d)$$

Traditionally, the optimisation under uncertainty is conducted by the brute force Monte Carlo method. This approach might be feasible for a simple objective model but not for a

high-fidelity model or large simulation. As shown in Figure 1-1b, the concept of OUU is that a nested loop exists to compute and form a statistics of response output for the optimisation process. Even though the wind farm power model here is not a high-fidelity model like CFD codes, a high computational time is required to evaluate the wind farm power with a fine interval in wind speed and wind direction. Consequently, to overcome this issue, the optimisation under uncertainty uses the developed UQ model, which consists of two surrogate models.

6-5-1 Optimisation with the UQ model

It has been shown and verified in Chapter 5 that the use of surrogate models in uncertainty quantification leads to a reduction in computational time without scarifying too much of the estimation accuracy. Among all of the combinations of the surrogate model on wind farm power output model and on the ability to propagate the uncertainty inputs, PCE-PCE stands out as it provides a reasonable estimation on P50, P90 value of AEP, and standard deviation. The UQ model consists of two surrogate models. The first surrogate model corresponds to the one in Section 6-4-2, and the second one is the surrogate model uses to propagate the inter-year uncertainty to compute the statistics of AEP response. Furthermore, OM-MC in Chapter 5 is treated as a comparison, which is based on a medium-fidelity original wind farm model and propagates 100,000 Monte Carlo samples to compute the AEP response.

Similarly, both BGA and RGA are implemented as the optimiser. The optimised wind farm layout from the optimisation is presented in Figure 6-7. For the two cases with self-generated layouts, the algorithms shift the wind turbine locations to the boundary to avoid wake effects. On the other hand, the two cases with a given initial layout sort of stick at a local extremum (i.e., the initial reference layout). A summary of the optimised results is reported in Table 6-7. In the table, PD refers to the pre-defined samples in wind speed and wind direction. The samples for the uncertainty inputs are random samples via Monte Carlo sampling. Despite the use of PCE model, it is not possible to use the analytical expression of mean and variance from PCE model as the output response is not necessarily a normal distribution, as shown in Figure 5-7a. The P50 value is not equal to the mean of the response in this case. The OM-MC model is used to verify that the optimised layout is indeed better in the P90 value of AEP. Furthermore, based on the Table 6-7, it is seen that the optimisation with BGA takes around 17 hours to complete 100 generations while RGA requires more than 20 hours. However, the longer computational time might be worthy as the design space for RGA is not constrained to the pre-defined grid points, and the resulting P90 AEP is possibly higher than the one from BGA.

	Iter.	P90 [MWh]	Improv. [%]	A. P90 [MWh]	A. Improv. [%]	Time [hr.]
OM-MC	-	$1.9623 \cdot 10^5$	-	-	-	-
BGA PD	100	$1.9666 \cdot 10^5$	0.22%	$1.9393 \cdot 10^5$	-1.17%	16.34
RGA PD	100	$1.9646 \cdot 10^5$	0.12%	$1.9558 \cdot 10^5$	-0.33%	22.71
BGA Init. PD	100	$1.9642 \cdot 10^5$	0.10%	$1.9628 \cdot 10^5$	0.02%	16.75
RGA Init. PD	100	$1.9874 \cdot 10^5$	1.28%	$1.9580 \cdot 10^5$	-0.22%	22.37

Table 6-7: Summary of the optimisation results regarding different optimisation methods with UQ models. Init. denotes the case is given an initial layout.

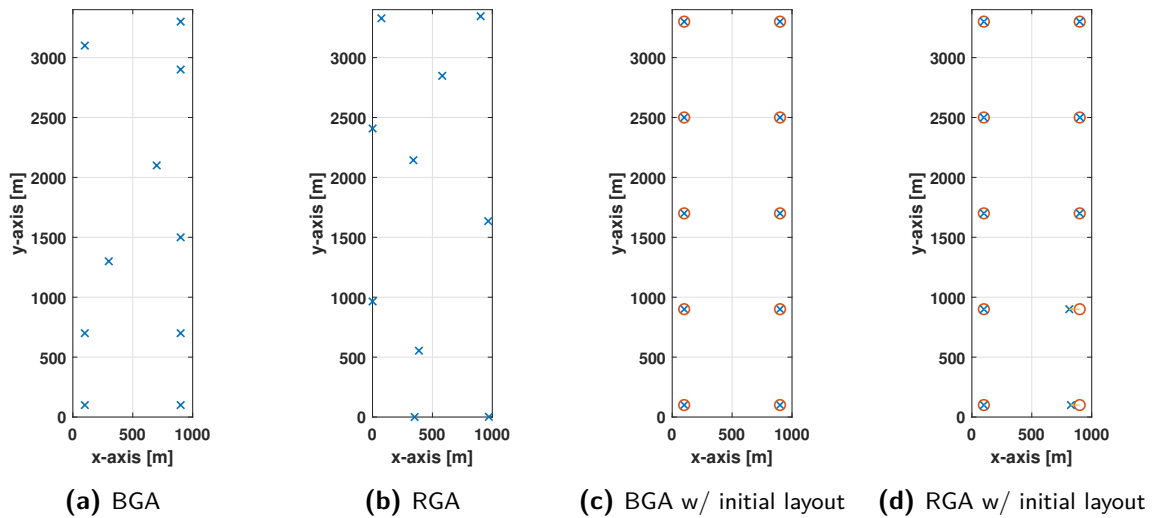


Figure 6-7: Optimisation under uncertainty of wind farm layout with UQ model. The blue crosses are the optimised wind turbine locations and the red circles are the reference initial wind farm layout, if specified.

In order to show the advantage of optimisation under uncertainty, the optimised layout in Figure 6-7d is investigated to study the difference with the deterministic optimisation. The deterministic OM model computes the AEP of the optimised layout by $2.2291 \cdot 10^5$ [MWh]. On the other hand, the UQ model with an OM-MC method gives a P50 value of $2.2272 \cdot 10^5$ [MWh] and a P90 value of $1.9580 \cdot 10^5$ [MWh] for the wind farm AEP. The probabilistic analysis from Figure 6-8 suggests that the AEP calculated by the deterministic model has only 49.5% of chance for the AEP to be larger than $2.2291 \cdot 10^5$ [MWh], which is a lower chance than the P50 value.

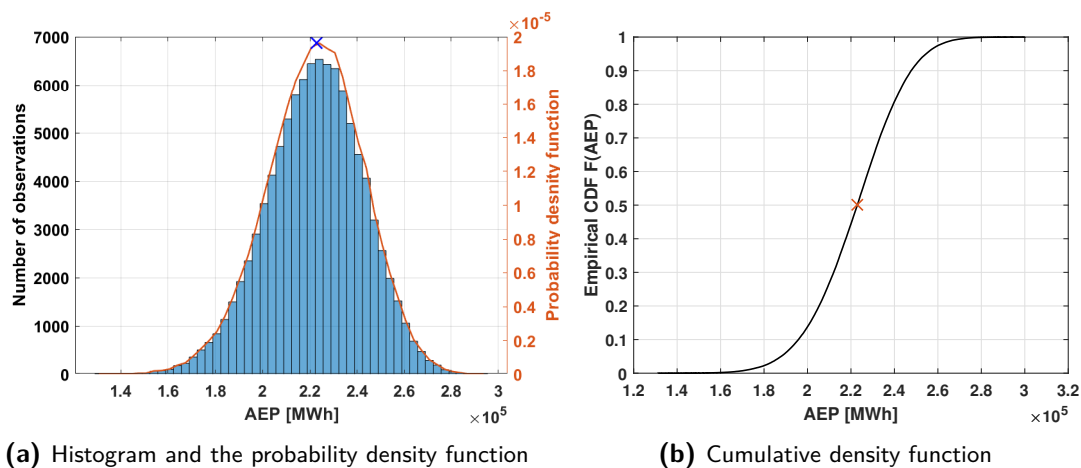


Figure 6-8: AEP probability distribution of the optimised layout compared to the deterministic model. The crosses denote the AEP calculated by the deterministic OM model.

Further, two optimised layouts from deterministic optimisation and optimisation under uncertainty are investigated with their robustness against inter-year variation in wind resource.

These optimised layouts are shown in Figure 6-6a, and Figure 6-7b and they are both optimised by RGA with no initial layout given to the optimiser. The size of the population and the number of generation for the optimiser are the same for both cases. As stated in Eq. (6-7), the OUU optimiser looks for the P90 value of AEP. The optimised layout from the deterministic optimisation is then subjected to the UQ model to have a valid comparison. The P90 value of AEP from both layout are given in Table 6-8. It can be concluded that the layout from the optimisation under uncertainty possesses a higher P90 value, which could mean that it is more robust to the inter-year variation.

	Deterministic Optimisation	Optimisation under Uncertainty
P90 AEP [MWh]	$1.9537 \cdot 10^5$	$1.9548 \cdot 10^5$

Table 6-8: Comparison of P90 value of AEP on the optimised layout with deterministic optimisation and optimisation under uncertainty using RGA.

From the above findings, they indicate optimisation under uncertainty is likely to give a more robust layout as it looks for a more conservative objective value, P90 of AEP, and can give a higher value of the objective value. It is, therefore, a necessary procedure when the uncertainty variables in wind resource are considered.

Nevertheless, some findings need to be pointed out. First, it is possible that the initial layout that was hand-picked by the author may be too good in terms of AEP. As the distances are more than six rotor diameters in that layout, it already lowers the wake effects to some certain extent. This further causes the optimiser to linger around this local optimum. After these optimisations, insights into the wind farm layout optimisation based on the genetic algorithm can be concluded. BGA has an advantage in computational time. RGA has the advantage of possibly giving better results. The two genetic algorithms are easy to implement and work well with optimisation under uncertainty in this research. The time of performing optimisation under uncertainty is reduced by the surrogate UQ model and is now similar to the deterministic optimisation with the medium-fidelity model. However, it provides a layout that may be more robust to the inter-year variation as indicated by the P90 value of AEP. Despite this, it must be stressed that these conclusions are based on the cases investigated in this research, and it might be case sensitive. Also, there are many more algorithms for the optimisation that could engage with the topic of optimisation under uncertainty.

6-6 Summary of the chapter

In this chapter, the optimisation algorithm, genetic algorithm, is explained along with its two variation types. A various number of options exist in the MATLAB toolbox environment. However, it should be noted that the options chosen are heavily case dependent. In deterministic optimisation, it failed to optimise the wind farm layout with a complete wind farm model due to the massive computational effort. As an alternative approach, a single wind speed and wind direction are used to optimise the layout. The results show that the genetic optimiser can indeed improve the layout into a more efficient one. Nevertheless, it shows the need for using the surrogate model in order to tackle the computational issue. The surrogate models have been proven the ability to reduce the computational time but at the cost of

a small approximation error. As for the optimisation under uncertainty, the surrogate UQ model has shown its ability to reduce the number of uncertain samples needed in the AEP evaluation. This reduction further leads to less optimisation time. In the end, the two types of the genetic algorithm have shown their ability to optimise the wind farm layout despite that they stuck at local extrema in some cases. This can be because the initial hand-picked layout by the author is already performing well in terms of AEP.

Conclusion & Recommendations

7-1 Conclusions

The purpose of this thesis research is to examine an optimisation under uncertainty problem, WFLO, by implementing surrogate models in the wind farm performance calculation and uncertainty quantification on the inter-year variation within the wind resource. PCE and Kriging surrogate models on the wind farm power model are introduced in Chapter 3. Both surrogate models perform well for the one-dimensional case of wind speed but deviate for the case of wind direction. It has been found that the PCE model struggles with the approximation of the power output in the case of wind direction. It is unable to capture the wake effects in certain wind directions. In contrast, Kriging is good at capturing the wake effect in this case.

Nonetheless, when looking into a two-dimensional case, where both wind speed and wind direction are taken as surrogate input variables, the two surrogate models reach a satisfactory accuracy on the estimation of AEP. The errors of AEP between the estimation and the one from the original model lie within 2% for both models. Besides the accuracy of the estimation using the surrogate models, the time required to construct such models is also measured. A two-dimensional PCE model of wind farm power needs 25% less time to build than a two-dimensional Kriging model. In order to be sure that the constructed surrogate models work correctly on different types of layout, an investigation on 7 user-defined wind farm layouts is performed. Based on the results in Chapter 4, both PCE and Kriging wind farm power model are slightly more incapable of wake-intense layouts in terms of approximation. However, the error is still within 2.5% to the original medium-fidelity model.

Many research has suggested to conduct uncertainty quantification in energy and cost analysis of a wind farm. As the objective parameters (e.g., AEP and LCOE) become statistical responses due to the effect of uncertain sources, the evaluation on the risk in financial aspect and the uncertainty quantification becomes essential. Traditionally, the uncertainties are propagated by the Monte Carlo method. This research has proposed and verified the use of surrogate models as a replacement of the computationally expensive Monte Carlo method.

According to the results, the combination of PCE to model wind farm power output, and again PCE to propagate inter-year uncertainty in wind conditions has a good estimation on P50, P90 of the AEP and the standard deviation. The number of samples from the inter-year uncertain inputs are reduced to 100 from 100,000 with an error around 1%.

Regarding the wind farm layout optimisation, both binary and real-coded genetic algorithms are used. Before implementing them to conduct optimisation under uncertainty, both genetic algorithms are verified with the original and surrogate wind farm power model, and they indeed provide a better layout in various cases. It is found that the optimised layout depends on the initial layout of the optimiser. If an initial layout is given, the optimiser appears to linger around this layout as it could be possible that the given layout is already a good design. In contrast, a self-generated layout by the optimiser is able to reach a similar performance of that well-designed layout. However, the improvement in AEP over the reference layout is not significant in most of the investigated cases. As for the optimisation under uncertainty, the surrogate UQ model reduces 98% computational time from a model with the combination of the medium-fidelity original model for wind farm power output, and Monte Carlo method that propagates the inter-year uncertainty in wind conditions. This also contributes to a faster optimisation of OUU. In the end, the optimised layout from the optimisation under uncertainty provides a layout that may be more robust to the inter-year variation as indicated by a higher P90 value of AEP than the layout from the deterministic optimisation.

7-2 Recommendations for future work

Even though the research objectives and questions have been achieved and answered, several approaches have been kept as simple as possible to obtain the results due to the time limit of this research. It, therefore, considers some recommendations for future work. These recommendations can be divided into three categories, as each of them corresponds to an essential part of this research. The first part is about surrogate modelling techniques. The second part considers the uncertainty quantification approach, whereas the third reflects on the recommendation on the optimisation of wind farm layout.

Surrogate modelling

To begin with, the first surrogate model is based on the wind farm power model. It is also possible to use a more advance and accurate wake model than the Jensen wake model. It will be interesting to see how the two surrogate models perform with different wake models. There are also more options for surrogate models, such as low-rank tensor approximations and support vector machines that could possibly yield better approximations. Lastly, for the surrogate model on the wind farm power output, it is recommended to perform the same study on a wind farm with a larger scale or even on a real-life wind farm to verify the effectiveness instead of a simplified case as studied in this research.

Uncertainty quantification

The uncertain variables in this research are the yearly Weibull parameters and the wind direction sector probabilities. Extensions on these uncertain variables are possible as the

Weibull parameters can be defined per wind direction sector instead of using an aggregated one for all. This could lead to an improvement in the AEP estimation. As for the wind sector probabilities, future work is to characterise them properly. In the current approach, the dominant wind sector is left out to ensure the sum of all wind sector can reach a probability of one. The consequence of this approach is still unknown and may lead to results that are not fully representative.

Further, it is possible to add more uncertain variables into the calculation of AEP. For example, turbulence intensity and the availability of the wind turbines could be additional variables. Nonetheless, with additional variables, the dimension of the problem becomes large. In this case, sensitivity analysis and gradient analysis could be in hand to avoid turning the problem into an infeasible one.

Optimisation

The optimisation under uncertainty problem is a challenging application which can further improve the robustness of an optimised result. It is suggested to experiment with other surrogate UQ models than the combination of PCE to model wind farm power output, and again PCE to propagate inter-year uncertainty in wind condition as it might lead to different optimised results. Further, a straightforward extension of the wind farm layout optimisation would be to consider a different optimisation objective. For example, LCOE seems to be a more realistic objective function for the wind farm layout optimisation. The number of generations and population size for the genetic algorithm could also be increased to search for a better result. Another extension could be using a different optimisation algorithm, which might be a better choice than the genetic algorithm regarding computational efficiency. Last but not least, it is suggested to run the wind farm optimisation with actual wind farm data. For example, the initial layout can be based on existing wind turbine coordinates while optimising the wind farm layout using the site condition data.

Appendix A

Surrogate model flowchart

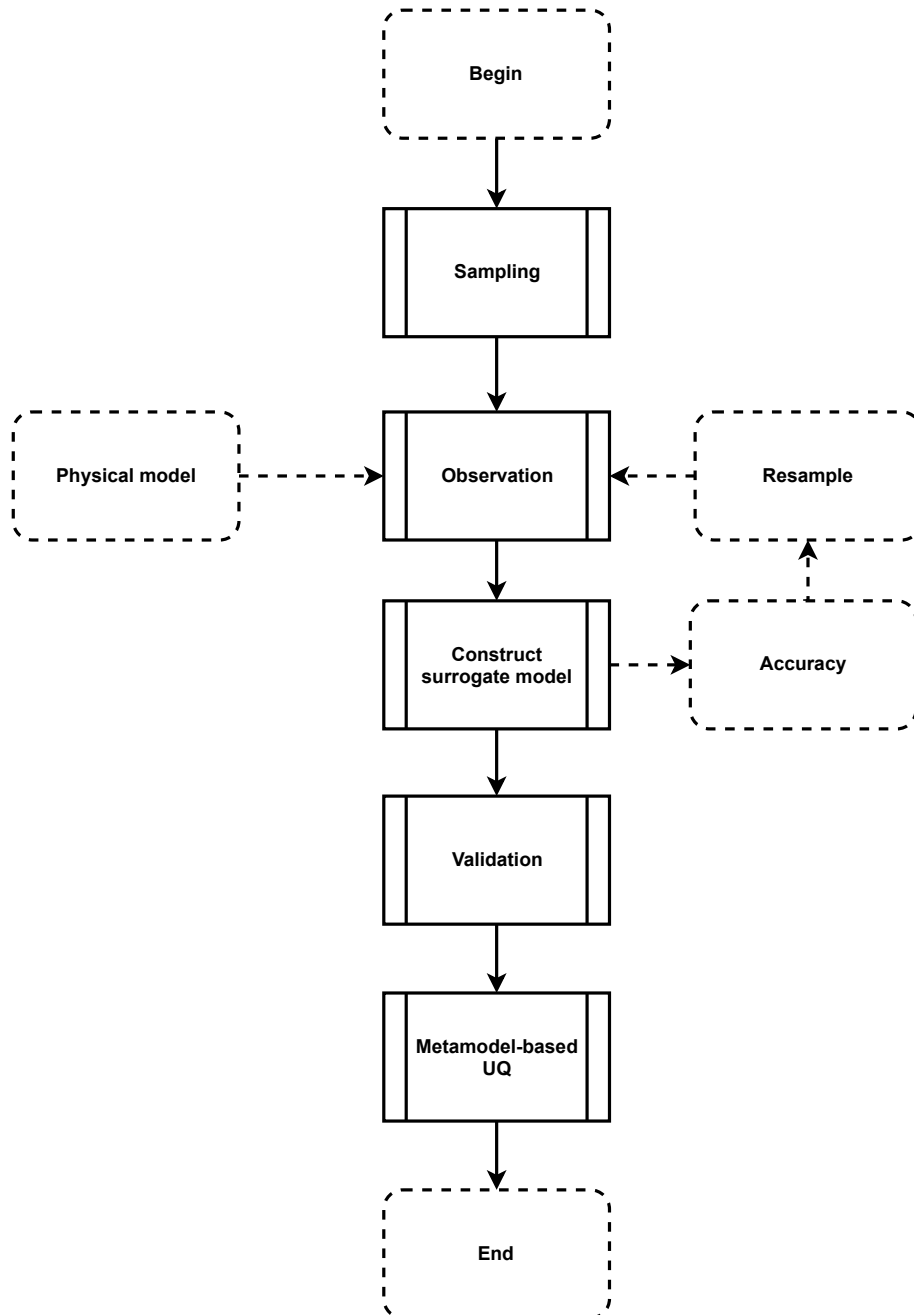


Figure A-1: Flowchart of constructing a surrogate model. The step of Metamodel-based UQ is only for the use of surrogating an UQ process.

Kriging modelling parameters

As the Kriging model requires more effort on deciding its parameters set-up, a study on these parameters are conducted. The comparisons are based on the percentage error between the approximation by the Kriging models and the original model with respect to a different number of samples on wind speed and wind direction.

The first comparison is on the correlation functions R , namely linear, exponential, and Gaussian correlation function. The mathematical expression of each correlation function has been shown in Table 3-3. The second comparison is on the estimation of hyperparameters θ . Within each comparison, the trend of the Kriging model varies from using constant ordinary Kriging to universal Kriging with different degree of polynomials. The results are presented in Figure B-1, Figure B-2, and Figure B-3.

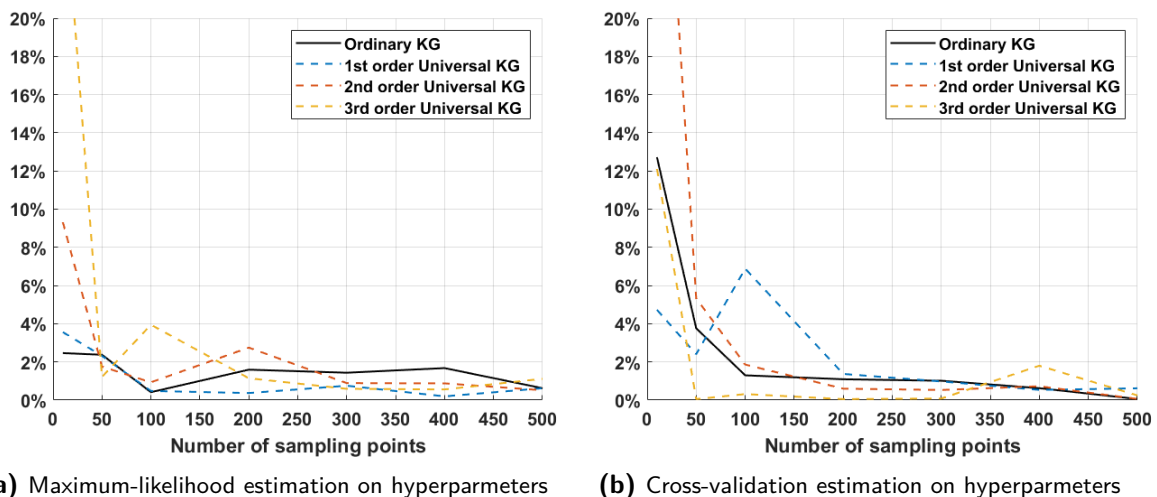


Figure B-1: Kriging models based on the linear correlation function.

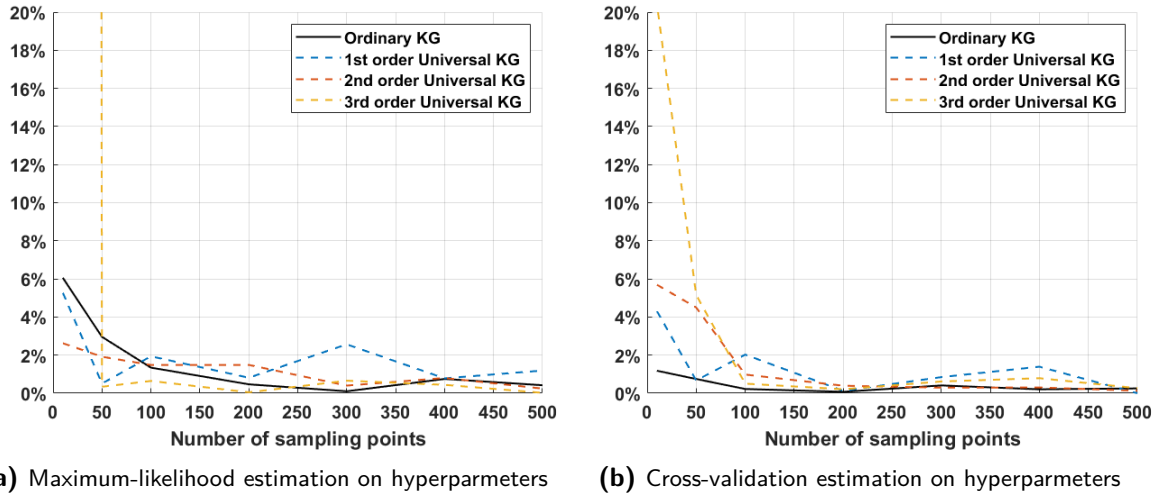


Figure B-2: Kriging models based on the exponential correlation function.

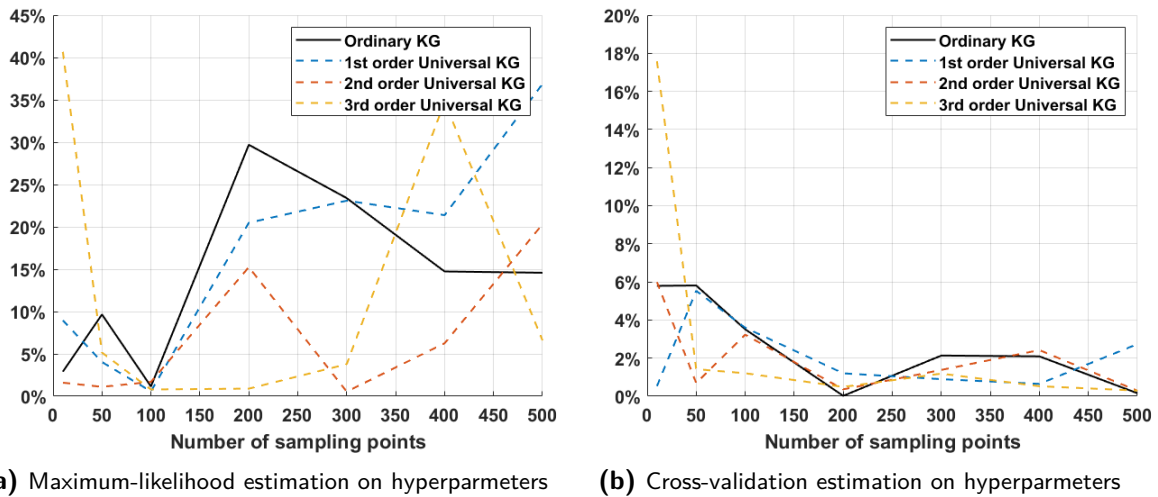


Figure B-3: Kriging models based on the Gaussian correlation function.

From the comparisons in all figures, the percentage error of the approximation by the Gaussian correlation function is the worst. As the correlation function indicates the influence of nearby points to the approximation points, it can say that the Gaussian correlation function does not describe the case well. On the other hands, both linear and exponential correlation functions are more accurate compared to the Gaussian one. There is no obvious point to conclude which correlation function is better. In the end, this research uses exponential correlation with a cross-validation estimation on hyperparameters. It is because, in a similar wind farm research of Gkoutis [16], the author recommends using exponential correlation function as it is the most typical choice for most Kriging modelling problems.

Nevertheless, it should be noted that this study of parameters is heavily case dependent. The conclusion may be suitable for this research only.

Appendix C

**WAsP wind farm report with wake
model output**

'2x5 Layouts' wind farm

Summary results

Parameter	Total	Average	Minimum	Maximum
Net AEP [GWh]	214.842	21.484	21.098	21.962
Gross AEP [GWh]	227.004	22.700	22.640	22.760
Wake loss [%]	5.36	-	-	-

Site results

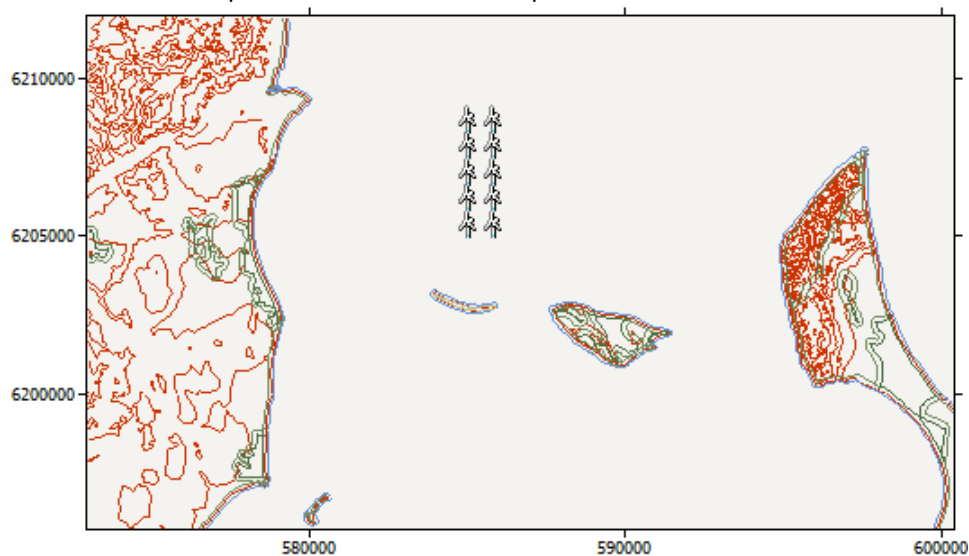
Site	Location [m]	Turbine	Elevation [m a.s.l.]	Height [m a.g.l.]	Net AEP [GWh]	Wake loss [%]
Turbine 001	(585100, 6205000)	NREL 5MW	0	90	21.962	3.0
Turbine 002	(585900, 6205000)	NREL 5MW	0	90	21.953	3.43
Turbine 003	(585100, 6205900)	NREL 5MW	0	90	21.530	4.9
Turbine 004	(585900, 6205900)	NREL 5MW	0	90	21.374	5.98
Turbine 005	(585100, 6206700)	NREL 5MW	0	90	21.385	5.6
Turbine 006	(585900, 6206700)	NREL 5MW	0	90	21.127	7.12
Turbine 007	(585100, 6207500)	NREL 5MW	0	90	21.390	5.61
Turbine 008	(585900, 6207500)	NREL 5MW	0	90	21.098	7.3
Turbine 009	(585100, 6208300)	NREL 5MW	0	90	21.644	4.55
Turbine 010	(585900, 6208300)	NREL 5MW	0	90	21.380	6.06

Site wind climates

Site	Location [m]	Height [m a.g.l.]	A [m/s]	k	U [m/s]	E [W/m ²]	RIX [%]	dRIX [%]
Turbine 001	(585100, 6205000)	90	10.3	2.22	9.15	811	0.0	N/A
Turbine 002	(585900, 6205000)	90	10.4	2.22	9.18	820	0.0	N/A
Turbine 003	(585100, 6205900)	90	10.3	2.22	9.15	811	0.0	N/A

Turbine 004	(585900, 6205900)	90	10.4	2.22	9.18	819	0.0	N/A
Turbine 005	(585100, 6206700)	90	10.3	2.23	9.15	810	0.0	N/A
Turbine 006	(585900, 6206700)	90	10.4	2.22	9.18	820	0.0	N/A
Turbine 007	(585100, 6207500)	90	10.3	2.23	9.15	810	0.0	N/A
Turbine 008	(585900, 6207500)	90	10.4	2.23	9.18	819	0.0	N/A
Turbine 009	(585100, 6208300)	90	10.3	2.23	9.15	811	0.0	N/A
Turbine 010	(585900, 6208300)	90	10.4	2.23	9.18	819	0.0	N/A

The wind farm lies in a map called 'Tunoe elevation map'.



The wind farm is in a project called 'Project 1'

A wind atlas called 'IJmuiden_offshore' was used to calculate the predicted wind climates

Calculation of annual output for '2x5 Layouts'

Decay constants: 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04

Sector 1 (0°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine 001	8.1	2.24	6.06	7.21	932.382	691.398	74.15

Turbine 002	8.1	2.24	6.06	7.21	932.287	691.271	74.15
Turbine 003	8.1	2.24	6.06	7.21	931.860	686.326	73.65
Turbine 004	8.1	2.24	6.06	7.21	931.655	686.165	73.65
Turbine 005	8.1	2.24	6.06	7.21	931.763	707.779	75.96
Turbine 006	8.1	2.24	6.06	7.21	931.733	708.765	76.07
Turbine 007	8.1	2.24	6.05	7.21	931.345	746.316	80.13
Turbine 008	8.1	2.24	6.06	7.21	931.301	746.051	80.11
Turbine 009	8.1	2.24	6.05	7.21	931.290	931.290	100.0
Turbine 010	8.1	2.24	6.05	7.21	930.977	930.977	100.0
Sector 1 total	-	-	-	-	9316.593	7526.338	80.78

Sector 2 (30°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine 001	8.5	2.83	3.46	7.55	573.826	455.495	79.38
Turbine 002	8.5	2.83	3.45	7.55	573.739	573.739	100.0
Turbine 003	8.5	2.83	3.45	7.55	573.462	457.981	79.86
Turbine 004	8.5	2.83	3.45	7.55	573.283	573.283	100.0
Turbine 005	8.5	2.83	3.45	7.55	573.394	479.361	83.6
Turbine 006	8.5	2.83	3.45	7.55	573.275	573.275	100.0
Turbine 007	8.5	2.83	3.45	7.55	573.018	530.384	92.56
Turbine 008	8.5	2.83	3.45	7.55	573.014	573.014	100.0
Turbine 009	8.5	2.83	3.45	7.55	572.954	572.954	100.0
Turbine 010	8.5	2.83	3.45	7.55	572.756	572.756	100.0
Sector 2 total	-	-	-	-	5732.722	5362.243	93.54

Sector 3 (60°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine 001	9.2	3.04	5.98	8.23	1185.106	1165.267	98.33
Turbine 002	9.2	3.04	5.98	8.23	1185.346	1185.346	100.0
Turbine 003	9.2	3.04	5.97	8.23	1184.197	1103.065	93.15
Turbine 004	9.2	3.04	5.97	8.23	1184.305	1184.305	100.0

Turbine 005	9.2	3.04	5.97	8.23	1183.610	1102.512	93.15
Turbine 006	9.2	3.04	5.97	8.23	1184.016	1184.016	100.0
Turbine 007	9.2	3.04	5.97	8.23	1182.640	1101.598	93.15
Turbine 008	9.2	3.04	5.97	8.23	1183.354	1183.354	100.0
Turbine 009	9.2	3.04	5.96	8.23	1182.253	1182.253	100.0
Turbine 010	9.2	3.04	5.97	8.23	1182.703	1182.703	100.0
Sector 3 total	-	-	-	-	11837.532	11574.419	97.78

Sector 4 (90°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine 001	10.3	2.94	7.74	9.20	1863.696	1562.477	83.84
Turbine 002	10.3	2.94	7.74	9.19	1860.007	1860.007	100.0
Turbine 003	10.3	2.94	7.75	9.21	1867.595	1566.027	83.85
Turbine 004	10.3	2.94	7.74	9.20	1865.273	1865.273	100.0
Turbine 005	10.3	2.94	7.75	9.22	1871.488	1569.559	83.87
Turbine 006	10.3	2.94	7.75	9.21	1870.306	1870.306	100.0
Turbine 007	10.3	2.94	7.75	9.23	1874.164	1572.146	83.89
Turbine 008	10.3	2.94	7.75	9.23	1875.280	1875.280	100.0
Turbine 009	10.3	2.94	7.75	9.23	1877.167	1574.992	83.9
Turbine 010	10.3	2.94	7.75	9.23	1878.094	1878.094	100.0
Sector 4 total	-	-	-	-	18703.071	17194.164	91.93

Sector 5 (120°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine 001	9.0	2.43	5.24	7.98	975.470	975.470	100.0
Turbine 002	9.0	2.44	5.24	8.02	984.931	984.931	100.0
Turbine 003	9.1	2.44	5.25	8.06	995.849	981.040	98.51
Turbine 004	9.1	2.44	5.27	8.08	1003.252	1003.252	100.0
Turbine 005	9.1	2.44	5.27	8.09	1005.596	943.824	93.86
Turbine 006	9.1	2.44	5.26	8.10	1006.538	1006.538	100.0
Turbine 007	9.1	2.44	5.28	8.10	1009.042	947.147	93.87

Turbine 008	9.1	2.44	5.26	8.10	1005.455	1005.455	100.0
Turbine 009	9.1	2.44	5.26	8.10	1008.138	946.364	93.87
Turbine 010	9.1	2.44	5.26	8.09	1005.329	1005.329	100.0
Sector 5 total	-	-	-	-	9999.601	9799.350	98.0

Sector 6 (150°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine 001	10.1	2.42	5.18	8.99	1169.271	1169.271	100.0
Turbine 002	10.1	2.42	5.19	8.94	1162.883	1162.883	100.0
Turbine 003	10.1	2.42	5.18	8.95	1162.023	1076.695	92.66
Turbine 004	10.0	2.43	5.19	8.89	1153.026	1153.026	100.0
Turbine 005	10.1	2.43	5.18	8.93	1160.310	1048.536	90.37
Turbine 006	10.0	2.43	5.19	8.90	1156.103	1156.103	100.0
Turbine 007	10.1	2.43	5.18	8.94	1161.694	1001.086	86.17
Turbine 008	10.1	2.43	5.19	8.93	1159.642	1159.642	100.0
Turbine 009	10.1	2.43	5.18	8.95	1163.823	1002.354	86.13
Turbine 010	10.1	2.43	5.18	8.95	1162.766	1162.766	100.0
Sector 6 total	-	-	-	-	11611.542	11092.363	95.53

Sector 7 (180°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine 001	11.8	2.71	9.15	10.48	2574.685	2574.685	100.0
Turbine 002	11.8	2.71	9.16	10.48	2575.803	2575.803	100.0
Turbine 003	11.8	2.71	9.15	10.48	2573.685	2307.404	89.65
Turbine 004	11.8	2.71	9.15	10.48	2574.366	2308.000	89.65
Turbine 005	11.8	2.71	9.15	10.48	2573.578	2179.918	84.7
Turbine 006	11.8	2.71	9.15	10.48	2574.798	2180.933	84.7
Turbine 007	11.8	2.71	9.15	10.48	2573.237	2134.514	82.95
Turbine 008	11.8	2.71	9.15	10.48	2574.354	2135.408	82.95
Turbine 009	11.8	2.71	9.15	10.49	2574.163	2068.593	80.36
Turbine 010	11.8	2.71	9.15	10.48	2573.755	2068.201	80.36

Sector 7 total	-	-	-	-	25742.422	22533.460	87.53
----------------	---	---	---	---	-----------	-----------	-------

Sector 8 (210°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine 001	12.2	2.40	14.11	10.79	3947.183	3947.183	100.0
Turbine 002	12.2	2.40	14.08	10.83	3950.302	3950.302	100.0
Turbine 003	12.1	2.39	14.14	10.76	3937.080	3937.080	100.0
Turbine 004	12.2	2.40	14.10	10.80	3945.469	3741.293	94.83
Turbine 005	12.1	2.39	14.14	10.75	3935.685	3935.685	100.0
Turbine 006	12.2	2.40	14.10	10.80	3945.171	3678.574	93.24
Turbine 007	12.1	2.39	14.16	10.72	3929.363	3929.363	100.0
Turbine 008	12.1	2.40	14.12	10.77	3938.869	3554.518	90.24
Turbine 009	12.1	2.39	14.16	10.71	3928.021	3928.021	100.0
Turbine 010	12.1	2.39	14.14	10.74	3932.951	3545.099	90.14
Sector 8 total	-	-	-	-	39390.093	38147.117	96.84

Sector 9 (240°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine 001	11.3	2.22	16.20	10.04	4111.796	4111.796	100.0
Turbine 002	11.4	2.23	16.21	10.12	4153.426	4153.426	100.0
Turbine 003	11.3	2.23	16.19	10.05	4117.167	4117.167	100.0
Turbine 004	11.4	2.23	16.20	10.13	4153.530	4114.836	99.07
Turbine 005	11.4	2.23	16.19	10.07	4123.355	4123.355	100.0
Turbine 006	11.4	2.23	16.20	10.14	4156.108	3989.281	95.99
Turbine 007	11.4	2.23	16.18	10.08	4127.839	4127.839	100.0
Turbine 008	11.5	2.23	16.20	10.15	4160.954	3994.247	95.99
Turbine 009	11.4	2.23	16.18	10.07	4120.696	4120.696	100.0
Turbine 010	11.4	2.23	16.19	10.14	4153.414	3986.616	95.98
Sector 9 total	-	-	-	-	41378.284	40839.258	98.7

Sector 10 (270°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine	10.3	2.08	10.67	9.11	2375.028	2375.028	100.0

001							
Turbine	10.4	2.08	10.68	9.18	2399.250	2079.532	86.67
002							
Turbine	10.3	2.08	10.67	9.10	2370.207	2370.207	100.0
003							
Turbine	10.3	2.08	10.67	9.16	2392.322	2071.905	86.61
004							
Turbine	10.2	2.08	10.66	9.07	2361.059	2361.059	100.0
005							
Turbine	10.3	2.08	10.67	9.14	2383.943	2064.076	86.58
006							
Turbine	10.2	2.08	10.66	9.06	2353.582	2353.582	100.0
007							
Turbine	10.3	2.08	10.66	9.13	2379.395	2059.153	86.54
008							
Turbine	10.2	2.08	10.66	9.02	2341.584	2341.584	100.0
009							
Turbine	10.3	2.08	10.66	9.11	2374.095	2054.079	86.52
010							
Sector 10	-	-	-	-	23730.465	22130.205	93.26
total							

Sector 11 (300°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine	9.0	2.15	8.79	7.99	1627.456	1627.456	100.0
001							
Turbine	9.1	2.15	8.79	8.05	1647.224	1624.795	98.64
002							
Turbine	9.0	2.15	8.78	7.97	1620.819	1620.819	100.0
003							
Turbine	9.1	2.15	8.79	8.06	1651.108	1557.842	94.35
004							
Turbine	9.0	2.15	8.78	7.99	1627.888	1627.888	100.0
005							
Turbine	9.1	2.15	8.79	8.08	1659.006	1565.458	94.36
006							
Turbine	9.1	2.15	8.78	8.03	1640.486	1640.486	100.0
007							
Turbine	9.2	2.15	8.79	8.12	1671.174	1577.324	94.38
008							
Turbine	9.2	2.15	8.78	8.12	1668.591	1668.591	100.0
009							
Turbine	9.2	2.15	8.79	8.17	1687.364	1687.364	100.0
010							
Sector 11	-	-	-	-	16501.115	16198.021	98.16
total							

Sector 12 (330°)

Turbine	A [m/s]	k	Freq. [%]	U [m/s]	MWh (free)	MWh (park)	Eff. [%]
Turbine	8.8	1.91	7.41	7.80	1306.135	1306.135	100.0
001							
Turbine	8.8	1.91	7.41	7.81	1306.754	1110.670	84.99
002							
Turbine	8.8	1.91	7.41	7.81	1305.985	1305.985	100.0
003							
Turbine	8.8	1.91	7.41	7.81	1306.112	1114.412	85.32
004							

Turbine 005	8.8	1.91	7.40	7.81	1305.723	1305.723	100.0
Turbine 006	8.8	1.91	7.41	7.81	1306.352	1149.791	88.02
Turbine 007	8.8	1.91	7.40	7.81	1305.591	1305.591	100.0
Turbine 008	8.8	1.91	7.41	7.81	1306.187	1234.425	94.51
Turbine 009	8.8	1.91	7.40	7.81	1306.045	1306.045	100.0
Turbine 010	8.8	1.91	7.40	7.81	1305.947	1305.947	100.0
Sector 12 total	-	-	-	-	13060.830	12444.724	95.28

All Sectors

Turbine	Location [m]	Gross AEP [MWh]	Net AEP [MWh]	Efficiency [%]
Turbine 001	(585100, 6205000)	22642.033	21961.660	97.0
Turbine 002	(585900, 6205000)	22731.952	21952.705	96.57
Turbine 003	(585100, 6205900)	22639.929	21529.795	95.1
Turbine 004	(585900, 6205900)	22733.701	21373.590	94.02
Turbine 005	(585100, 6206700)	22653.452	21385.202	94.4
Turbine 006	(585900, 6206700)	22747.347	21127.117	92.88
Turbine 007	(585100, 6207500)	22662.001	21390.051	94.39
Turbine 008	(585900, 6207500)	22758.982	21097.871	92.7
Turbine 009	(585100, 6208300)	22674.723	21643.737	95.45
Turbine 010	(585900, 6208300)	22760.151	21379.930	93.94
Wind farm	-	227004.269	214841.655	94.64

Bibliography

- [1] S. Afanasyeva, J. Saari, S. Kukkonen, J. Partanen, and O. Pyrhönen. Optimization of wind farm design taking into account uncertainty in input parameters. In *Proceedings of the European Wind Energy Conference and Exhibition*, pages 1–10, 2013.
- [2] S. Afanasyeva, J. Saari, M. Kalkofen, J. Partanen, and O. Pyrhönen. Technical, economic and uncertainty modelling of a wind farm project. *Energy Conversion and Management*, 107:22–33, 2016.
- [3] W. Bailleul. Using polynomial chaos expansion for wind energy. 2018.
- [4] D. Bigoni. Uncertainty quantification with applications to engineering problems. 2015.
- [5] G. Blatman and B. Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of computational Physics*, 230(6):2345–2367, 2011.
- [6] S. Chowdhury, J. Zhang, A. Messac, and L. Castillo. Unrestricted wind farm layout optimization (uwflo): Investigating key factors influencing the maximum power generation. *Renewable Energy*, 38(1):16–30, 2012.
- [7] A. Clifton, A. Smith, and M. Fields. Wind plant preconstruction energy estimates. current practice and opportunities. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2016.
- [8] K. Deep, K. P. Singh, M. L. Kansal, and C. Mohan. A real coded genetic algorithm for solving integer and mixed integer optimization problems. *Applied Mathematics and Computation*, 212(2):505–518, 2009.
- [9] M. Eldred and J. Burkardt. Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification. In *47th AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition*, page 976, 2009.
- [10] C. Engelen. The nonlinear effect of combining uncertainties on the energy yield of an offshore wind farm. Master’s thesis, Technical University Delft, the Netherlands, 2015.

- [11] I. Farrance and R. Frenkel. Uncertainty in measurement: a review of monte carlo simulation using microsoft excel for the calculation of uncertainties through functional relationships, including uncertainties in empirically derived constants. *The Clinical Biochemist Reviews*, 35(1):37, 2014.
- [12] J. Feng and W. Z. Shen. Modelling wind for wind farm layout optimization using joint distribution of wind speed and wind direction. *Energies*, 8(4):3075–3092, 2015.
- [13] S. Frandsen. Uncertainty on wind turbine power curve measurements. In *13th British Wind Energy Association Conference Wind Energy and the Environment*, pages 169–174. Mechanical Engineering Publications Limited, 1991.
- [14] M. Gaumond, P.-E. Réthoré, A. Bechmann, S. Ott, G. C. Larsen, A. P. Diaz, and K. S. Hansen. Benchmarking of wind turbine wake models in large offshore windfarms. In *The Science of Making Torque from Wind 2012: 4th scientific conference*, 2012.
- [15] A. Giunta, M. Eldred, and J. Castro. Uncertainty quantification using response surface approximations. In *9th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability*, pages 26–28. Citeseer, 2004.
- [16] K. Gkoutis. Layout optimization methods for offshore wind farms using gaussian processes. 2018.
- [17] J. S. Gonzalez, M. B. Payan, and J. M. Riquelme-Santos. Optimization of wind farm turbine layout including decision making under risk. *IEEE Systems Journal*, 6(1):94–102, 2011.
- [18] F. González-Longatt, P. Wall, and V. Terzija. Wake effect in wind farm performance: Steady-state and dynamic behavior. *Renewable Energy*, 39(1):329–338, 2012.
- [19] M. A. Hariri-Ardebili. Risk, reliability, resilience (r3) and beyond in dam engineering: A state-of-the-art review. *International journal of disaster risk reduction*, 31:806–831, 2018.
- [20] J. F. Herbert-Acero, O. Probst, P.-E. Réthoré, G. C. Larsen, and K. K. Castillo-Villar. A review of methodological approaches for the design and optimization of wind farms. *Energies*, 7(11):6930–7016, 2014.
- [21] N. O. Jensen. *A note on wind generator interaction*. Risø National Laboratory, 1983.
- [22] H. John. Holland. genetic algorithms. *Scientific american*, 267(1):44–50, 1992.
- [23] J. Jonkman, S. Butterfield, W. Musial, and G. Scott. Definition of a 5-mw reference wind turbine for offshore system development. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2009.
- [24] I. Katic, J. Højstrup, and N. O. Jensen. A simple model for cluster efficiency. In *European wind energy association conference and exhibition*, pages 407–410. A. Raguzzi, 1987.
- [25] S. Kawai and K. Shimoyama. Kriging-model-based uncertainty quantification in computational fluid dynamics. In *32nd AIAA Applied Aerodynamics Conference*, page 2737, 2014.

- [26] D. G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139, 1951.
- [27] M. Lackner, A. Rogers, and J. Manwell. Uncertainty analysis in wind resource assessment and wind energy production estimation. In *45th AIAA Aerospace Sciences Meeting and Exhibit*, page 1222, 2007.
- [28] M. A. Lackner and C. N. Elkinton. An analytical framework for offshore wind farm layout optimization. *Wind Engineering*, 31(1):17–31, 2007.
- [29] G. C. Larsen. *A simple wake calculation procedure*. Risø National Laboratory, 1988.
- [30] C. Lataniotis, S. Marelli, and B. Sudret. The gaussian process modelling module in uqlab. *arXiv preprint arXiv:1709.09382*, 2017.
- [31] C. Lataniotis, D. Wicaksono, S. Marelli, and B. Sudret. UQLab user manual – Kriging (Gaussian process modeling). Technical report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2019. Report n UQLab-V1.3-105.
- [32] P. Lissaman. Energy effectiveness of arbitrary arrays of wind turbines. *Journal of Energy*, 3(6):323–328, 1979.
- [33] S. Marelli and B. Sudret. UQLab: a framework for uncertainty quantification in MATLAB. In *Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014)*, Liverpool, United Kingdom, 2014.
- [34] S. Marelli and B. Sudret. UQLab user manual – Polynomial chaos expansions. Technical report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2019. Report n UQLab-V1.3-104.
- [35] T. A. McCourt, S. Hurter, B. Lawson, F. Zhou, B. Thompson, S. Tyson, and D. Donovan. Uncertainty quantification of coal seam gas production prediction using polynomial chaos. *Journal of Petroleum Science and Engineering*, 157:1148–1159, 2017.
- [36] G. Mosetti, C. Poloni, and B. Diviacco. Optimization of wind turbine positioning in large windfarms by means of a genetic algorithm. *Journal of Wind Engineering and Industrial Aerodynamics*, 51(1):105–116, 1994.
- [37] N. Moskalenko, K. Rudion, and A. Orths. Study of wake effects for offshore wind farm planning. In *2010 Modern Electric Power Systems*, pages 1–7. IEEE, 2010.
- [38] M. Moustapha, B. Sudret, J.-M. Bourinet, and B. Guillaume. Quantile-based optimization under uncertainties using adaptive kriging surrogate models. *Structural and multidisciplinary optimization*, 54(6):1403–1421, 2016.
- [39] V. Murali. Inventory and modelling of different objective functions and their impact on optimal design of an offshore wind farm. Master’s thesis, Technical University Delft, the Netherlands, 2018.
- [40] J. Murcia, P.-E. Réthoré, A. Natarajan, and J. D. Sørensen. How many model evaluations are required to predict the aep of a wind power plant? In *Journal of Physics: Conference Series*, volume 625, page 012030. IOP Publishing, 2015.

- [41] J. M. Murphy, D. M. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430(7001):768–772, 2004.
- [42] A. Padrón, A. P. Stanley, J. Thomas, J. Alonso, and A. Ning. Polynomial chaos for the computation of annual energy production in wind farm layout optimization. In *Journal of Physics: Conference Series*, volume 753, page 032021. IOP Publishing, 2016.
- [43] A. S. Padrón, J. Thomas, J. J. Alonso, et al. Polynomial chaos to efficiently compute the annual energy production in wind farm layout optimization. *Wind Energy Science*, 4(2):211–231, 2019.
- [44] B. Pérez, R. Mínguez, and R. Guanche. Offshore wind farm layout optimization using mathematical programming techniques. *Renewable energy*, 53:389–399, 2013.
- [45] E. Quaeghebeur. Robust wind farm layout optimization using pseudo-gradients. In *The 11th International Symposium on Imprecise Probability*, 2019.
- [46] S. Rodrigues. A multi-objective optimization framework for the design of offshore wind farms. *PhD. dissertation, Delft University of Technol., Delft, Netherlands*, 2016.
- [47] T. J. Santner, B. J. Williams, W. I. Notz, and B. J. Williams. *The design and analysis of computer experiments*, volume 1. Springer, 2003.
- [48] R. Schöbi, S. Marelli, and B. Sudret. UQLab user manual – Polynomial chaos Kriging. Technical report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2019. Report n UQLab-V1.3-109.
- [49] F. Seim. Validating kinematic wake models in complex terrain using cfd. Master’s thesis, Norwegian University of Life Sciences, Ås, 2015.
- [50] R. Shakoor, M. Y. Hassan, A. Raheem, and Y.-K. Wu. Wake effect modeling: A review of wind farm layout optimization using jensen s model. *Renewable and Sustainable Energy Reviews*, 58:1048–1059, 2016.
- [51] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, 93(7):964–979, 2008.
- [52] B. Sudret, S. Marelli, and J. Wiart. Surrogate models for uncertainty quantification: An overview. In *2017 11th European conference on antennas and propagation (EUCAP)*, pages 793–797. IEEE, 2017.
- [53] A. Tesauro, P.-E. Réthoré, and G. C. Larsen. State of the art of wind farm optimization. In *EWEA 2012-European Wind Energy Conference & Exhibition*. European Wind Energy Association (EWEA), 2012.
- [54] E. Torre, S. Marelli, and B. Sudret. UQLab user manual – Statistical inference. Technical report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2019. Report n UQLab-V1.3-114.
- [55] A. M. Update. Global wind report. *Global Wind Energy Council*, 2019.

-
- [56] L. van den Bos and B. Sanderse. Uncertainty quantification for wind energy applications. *Scientific Computing [SC]*, (SC-1701), 2017.
- [57] D. R. VanLuvanee. Investigation of observed and modeled wake effects at horns rev using windpro. *Technical University of Denmark Department of Mechanical Engineering*, 2006.
- [58] V. N. Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999, 1999.
- [59] L. Vermeer, J. N. Sørensen, and A. Crespo. Wind turbine wake aerodynamics. *Progress in aerospace sciences*, 39(6-7):467–510, 2003.
- [60] A. Vestas Wind Systems. General specification v82-1.65 mw-1 speed. 2008. <https://www.edprnorthamerica.com/wp-content/uploads/2014/04/V82.pdf>, [Accessed: Feb 2020].
- [61] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.
- [62] X. Wu, C. Wang, and T. Kozłowski. Kriging-based surrogate models for uncertainty quantification and sensitivity analysis. In *Proceedings of the MC-2017, International Conference on Mathematics Computational Methods Applied to Nuclear Science Engineering*, 2017.