

Modelling Power From Wind Turbines

- A Bayesian Approach -
In cooperation with Goldwind Denmark

Master's Thesis

Tobias Kallehauge and Jonas Villadsen

Aalborg University

Department of Mathematical Sciences



AALBORG UNIVERSITY
DENMARK

Copyright © Aalborg University 2020

The report is written in \LaTeX . Python 3.7.6 has been used for data processing and the figures are drawn using the Matplotlib package in Python or the Tikz package in \LaTeX .
Frontpage image source: Colourbox



AALBORG UNIVERSITY

STUDENT REPORT

Mathematical Engineering

Aalborg University

<http://www.en.aau.dk/>

Title:

Modelling Power From Wind Turbines
A Bayesian Approach

Project Period:

10th semester (spring 2020)

Project Group:

1.204c

Participants:

Tobias Kallehauge
Jonas Villadsen

Supervisors:

Ege Rubak
Jan Østergaard

Copies: 3**Page Numbers:** 95**Date of Completion:**

June 3, 2020

Abstract:

Designing wind-to-power (W2P) models is a discipline within the wind turbine industry where the produced power from wind turbine generators is modelled as a function of wind. The current regime of W2P models is data-driven where models are made with little to no regard as to the physical phenomena at play. This thesis investigates how to design a W2P model directly based on the physical processes of a wind turbine using statistical modelling of wind speed available as prior information for a specific site. The statistics are based on the method/design philosophy *Approximate Bayesian Total Error Analysis* developed for this thesis. Markov Chain Monte Carlo methods have been used to infer model parameters.

Two different models, the AW2P and NTF-AW2P model are designed and compared to existing models in a performance evaluation. The two models show similar performance but perform similarly or worse compared to existing W2P models across several error measurements, thus further research is needed to improve the models.

The content of this report is freely available, but publication (with reference) may only be pursued through agreement with the authors.

Contents

Preface	vii
1 Introduction	1
1.1 Modelling wind turbine generators	2
1.2 Methods and considerations	9
1.3 Research question	11
1.4 Methodology	12
1.5 Delimitations	13
2 Available data	15
2.1 SCADA	16
2.2 Lookup tabels	18
3 WTG modelling	21
3.1 Wind speed modelling	21
3.2 Pitch control	28
3.3 Aggregated W2P model	33
3.4 Alternative models	34
3.5 Synthetic data	37
4 Bayesian Inference	39
4.1 Bayesian total error analysis	39
4.2 Non-informative Bayesian inference priors	44
4.3 Obtaining marginal wind speed distributions	47
4.4 Power distribution	50
4.5 Posteriors and logarithm posteriors	52
4.6 Implementation	54
5 Markov Chain Monte Carlo	
Methods	59
5.1 Metropolis Algorithm	59
5.2 Experiment on synthetic data	61
6 Performance Analysis	65
6.1 Case setup: Training and test	65
6.2 Parameter estimation	67
6.3 Performance	68
6.4 Residual analysis	72
7 Discussion	75
8 Conclusion	79

9 Further Studies	81
Bibliography	83
A List of abbreviations	87
B Mathematical definitions and results	88
B.1 List of distributions	88
B.2 Results and definitions	89
C WTG engineering:	
Additional theory and results	91
C.1 Annual Energy Production and wind speed statistics	91
C.2 Estimation of \mathbf{v}_1 and \mathbf{v}_2	93
D Software	95

Preface

This Masters' thesis was written by Tobias Kallehauge and Jonas Villadsen from the 4th semester of the Masters program in Mathematical Engineering at Aalborg University.

The authors wish to thank the university supervisors Ege Rubak from the Department of Mathematical Sciences and Jan Østergaard from the Department of Electronical Systems for supervision during the project. The authors also wish to thank Goldwind Denmark who provided data for the thesis, Chief Engineer Bo Juul Pedersen and Senior Engineer Ulrich Stolz for their supervision as well as Head of HR Dorit Bech Andersen for supervising the process. Finally the American Meteorological Society is thanked for permission to show the graphic in figure 3.4.

Prerequisites for reading this thesis is undergraduate level understanding of probability theory and statistics including basic understanding of Bayesian probability theory.

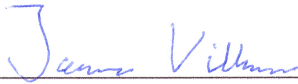
For notation, vectors and matrices are written in bold font, e.g. \mathbf{x} and \mathbf{X} . Probability density functions (pdf's) are written as p possibly with a subscript to indicate the random variable. Distinction is made between a random variable and its realisation is mostly omitted, but it should be clear from context which is being referred to. $x \sim p_x(x)$ indicates that x has pdf p_x . \log refers to the natural logarithm - i.e \log_e . Mathematical equations, statements ect. are numbered only if referred to.

Some of the information obtained through out this thesis has been obtained through consultations with Goldwind Denmark. To emphasise this, information where the only source is through consultation is marked with a "+" at the end of the relevant paragraph.

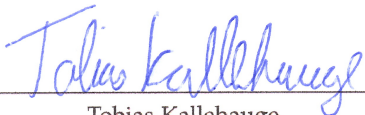
A list of the abbreviations used through out this thesis can be found in Appendix A.

The scripts developed during the process of writing the thesis are written in Python 3.7. Specific scripts will be referred through the text but are not publicly available. Access may be arranged upon request.

Aalborg University, June 3, 2020



Jonas Villadsen
jvill14@student.aau.dk



Tobias Kallehauge
tkalle15@student.aau.dk

1 | Introduction

Climate change is one of the great challenges facing humanity in the 21st century considering, that the global temperature compared to the average since 1885 has been above the average since the 1980's, and in 2018 the global temperature anomaly was 0.82 degrees Celsius [45]. Renewable sources of energy as alternatives to fossil fuel might be one of the ways to solve this challenge, but on a global scale only 14% of the worldwide energy supply came from renewable sources in 2016 [31]^[1]. Even though renewable energy sources generate a relatively small part of the global energy today, the future could see a significant increase in contribution from these sources. In a report by the *Global Commission on the Geopolitics of Energy Transformation* from 2019 [20] several prediction studies are surveyed and it is hypothesised that the demand for renewable energy sources will supersede that of fossil fuels between 2040 and 2050 [20, p. 17]. This is of course associated with a lot of uncertainty but renewable energy production is already rapidly increasing and the report lists several arguments why this is the case, including “*declining cost*” and “*technological innovation*” [20, pp. 18–23]. The report highlights wind and photovoltaic power since they are undergoing “*very rapid growth*” [20, p. 15], but these technologies also face many challenges, one of these being their uncertain nature due to being dependent on the weather.

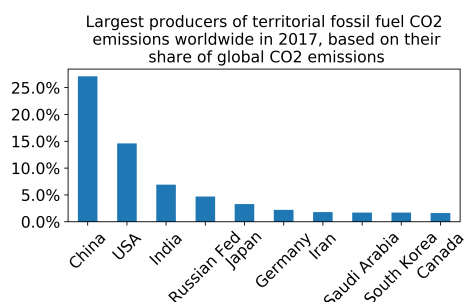


Figure 1.1: Source: [19]

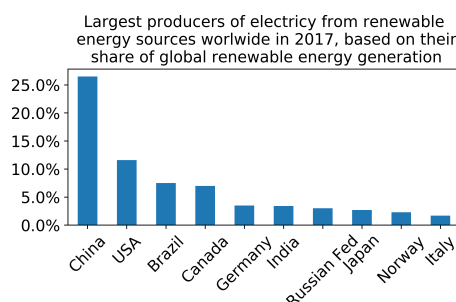


Figure 1.2: Source: [47]

China is in a unique position in regards to climate change. On the one hand China is the worlds largest emitter of carbon dioxide with 27.2% of the total carbon dioxide emissions in 2017 [19], but on the other hand they are also the worlds largest producer of renewable energy with 26.5% of the total renewable energy produced in 2017, amounting to 17.4% of domestic energy generation, according to [47, 3]^[2]. This of course has to do with the fact that China is the worlds most populous country with almost 1.4 billion citizens in 2018 [12], but it also means that China as a country has a big responsibility in help solve the climate change challenge.

^[1]The main sources being biofuel, hydroelectrics, geothermal energy, photovoltaic power, wind power and thermal energy.

^[2]Counting nuclear energy in China's energy production from renewable sources.

Even though China generates most of its renewable energy from hydroelectrics, the contribution from wind power is still significant with 5% of the total energy generated in 2018 [6]. Furthermore, there seems to be an exponential growth of power generated from wind with only 11 GWh in 2008 versus 305 GWh in 2017 and the Chinese region of Inner Mongolia generates a surprising amount of this with 19% of the total generated wind power in 2016 [32, 7]. This rapid growth is explained partly by government subsidies and declining price of installing wind turbine generators (WTG's), but many other factors, such as a huge offshore potential in China's long coastline and challenges with power grid infrastructure due to the size of the country also play an important role in the growth rate [16, 54]. Of the companies contributing to China's growth in the wind power sector, Goldwind is in the lead being the single largest manufacturer of WTG's in China, both in terms of newly installed and total installed capacity. In 2018 WTG's from Goldwind had a total wind power capacity of 49 GW - the closest contender is United Power with 20 GW wind power capacity in 2018 [9]. On an international level Goldwind primarily exports WTG's to Australia and Argentina as the largest buyers in 2018 [10]. Goldwind has a total of 7 global research and development facilities, including one in Hinnerup, Denmark, which this project is written in cooperation with.

The system of generating electric power from the wind with WTG's is complicated and models can suffer from a significant degree of uncertainty. Models for electric power are used in a variety of applications presented in the following section and advances for these could lead to advances in the WTG industry. The theme throughout this thesis will be to analyse specific physical processes behind power production in order to try to bring the uncertainty down. This will be done with a statistical approach using Bayesian methods and the models will be tested using data provided by Goldwind Denmark. The following section will explore further what the theme of the thesis is, followed by some considerations about methods and the specific focus of the thesis. The research question will summarize the considerations followed by methodology and finally, the chapter will be concluded with the delimitations of the thesis.

1.1 Modelling wind turbine generators

As mentioned, modelling a wind turbine generator is a complicated task. Many parameters influence the performance of a WTG, including design, wind flow, air composition, surrounding terrain and presence of other nearby WTG's. The amount of power generated over a period of time (measured in Watt (W)) is a simple measure of performance, but other measures, such as the mechanical stress are also important, as they concern the operational life span of the WTG. In this thesis the interaction between parameters that influence the WTG and the generated power will be examined. Why this interaction is important to companies in the WTG business, and what this thesis hopes to achieve, will be covered shortly, but initially it should be specified what is actually modelled.

As one might expect, wind speed is the most important input parameter for modelling produced power from a WTG so let $v \in \mathbb{R}$ denote the (positive) wind speed measured in m/s. Throughout the thesis v is usually associated with a degree of uncertainty and thus it can be characterised as a realisation of a random variable which will be characterised in chapter 3. The input, as will be reviewed in section 1.1.2, can be limited to only to wind speed v , but other variables such as air density and wind direction can be relevant, along with more “advanced parameters” like turbulence intensity. The available input variables for this thesis will be explained in detail in Chapter 2. In terms of output variables the generated power over a period time, $P \in \mathbb{R}$, measured in kilo Watt hours (kWh) is the variable of interest. In a simple setup with just one WTG, P is just a scalar, but it might also refer to a collection of WTG’s referred to as a wind farm (WF) with $\mathbf{P} \in \mathbb{R}^m$ where m would be the number of WTG’s in the WF. In the scalar case with input limited to wind speed, the input/output relation between v and P is modelled by a *wind to power* (W2P) model, denoted simply as $W2P : \mathbb{R} \rightarrow \mathbb{R}$ such that:

$$P = W2P(v; \theta), \quad (1.1)$$

where $\theta \in \Theta$ is the parametrisation of the W2P model and Θ is the parameter space. The W2P model in (1.1) is a simplification in different ways - e.g by assuming that there exists a deterministic relation between v and P and assuming that input is limited only to wind speed, but it is sufficient for the following considerations.

For wind turbine companies such as Goldwind, having access to accurate W2P models is valuable for several reasons. One of these reasons is the need for accurate wind turbine power curves (WTPC’s or, simply, power curve), which can be computed from a W2P model. A power curve is a function associating a certain wind speed with a certain power generation from a specific WTG typically measured in kWh. Associated with the power curve are values for three different wind speeds: *cut-in* (v_c), *rated* (v_r) and *cut-out* (v_{co}) [33, p. 77]. Cut-in is the lowest wind speed where the torque from the blades is greater than the inertia of the system and below this wind speed no significant power is generated. At the rated wind speed the WTG generates its maximum *nominal power* (P_{nom}) and above the rated speed the WTG is controlled in order to maintain its rated power. The cut-out is the maximum wind speed at which the WTG can safely operate and above this wind speed the WTG is controlled to stop. See figure 1.3 for an illustration of a power curve. In practice the power curve is not a simple function of only wind speed - see chapter 3 for more details.

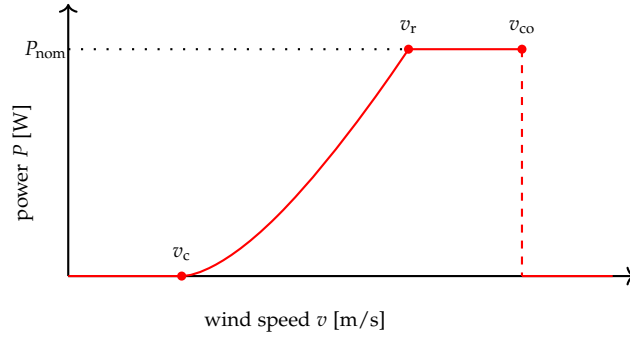


Figure 1.3: Idealised wind turbine power curve. The cut-in, rated and cut-out wind speeds, denoted v_c , v_r and v_{co} respectively, as well as the nominal power P_{nom} are marked. Note that power is displayed on the ordinate and thus the power curve is aggregated in the sense that any point of the graph is the total amount of Watts (Joules/sec) generated with a certain wind speed over a specified time period - e.g. 10 minutes.

A power curve is important to companies since it is used in the specifications for a WTG. When commissioned to install one or more WTG's at a particular site, the WTG company will deliver a power curve that specifies how much power will be generated at that location under different circumstances. Different factors affect this curve, so for a new location, engineers will model a power curve specific for that location accounting for various factors[†]. The power curve has to account for uncertainty in the model and will only be a lower boundary of modelled power - e.g. 95% - since the company can be fined if the WTG's do not produce the promised power as predicted by the contract curve[†].

Another use of a W2P model is in estimation of the *annual energy production* (AEP) [33, ch 7]. The AEP is estimated using long term weather forecasts and in practise it is associated with a measure of uncertainty [33, pp. 118–119]. Like the power curve, the AEP is used in the specifications of a WTG. If the AEP is not sufficiently accurate, it may be necessary to report an AEP much lower than the true one in order to make sure that the WTG produces at least as much energy as specified[†]. Of course, having access to accurate weather forecasts is also required in order to achieve accurate AEP estimates. Some additional details about AEP calculations and how weather forecast are generated is covered in appendix C.1.

The accuracy of the W2P model is central both in terms of modelling the contract power curve and estimation of the AEP and will be the main concern of this thesis. A calibrated W2P model could potentially be a very accurate (digital) description of a physical WTG and in recent years the concept of a *digital twin* have been used in many areas including in the wind power industry [57]. The following section will explore the concept of a digital twin.

1.1.1 Digital and physical twins

Several sources point to Dr. Michael Greives as the person introducing the concept of a *digital twin* in 2002 [22, 29, 42], although the term itself was coined by John Vickers in 2011 [23, p. 177]. By Greives own definition from his 2019 paper [23], *the Digital Twin is the information construct of the Physical Twin* [23, p. 176]. In a mathematical sense a digital

twin is a model/function calibrated to behave like a physical object which is the physical twin. There are two important distinctions separating a digital twin and a model in general. Firstly, a digital twin is always associated with a real physical object or at least a prototype in development. Secondly, a digital twin receives real-time information about the physical twin measured from sensors which is often transmitted via the internet [23, p. 186]. Having a direct link to the physical twin (referred to as the *digital thread*) allows for various exploitations of the model, such as real time evaluation of the model accuracy or even calibrating the model in real time in order to match the behaviour of the physical twin. An important factor accelerating continued use of digital twins is the expanding *internet of things* enabling more devices to be connected via the digital thread allowing for a digital twin to exist [23]. Uses of digital twins includes detecting cracks in WTG blades, aircraft life prediction models and testing parameter adjustments for production lines before applying them in real systems [57, 58, 36].

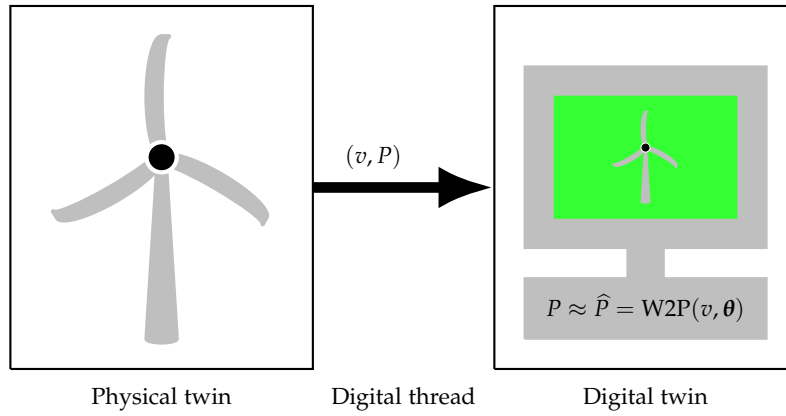


Figure 1.4: Illustration of the concept of interaction between the physical and digital twin - a WTG, with associated W2P-model is shown here. The arrow illustrates real time data sent from the physical twin measured with on-site sensors.

The W2P model can be treated as the main component of a digital twin. Here the physical twin would be a WTG or a WF where sensors measuring with some frequency are installed. This kind of data is referred to as *Supervisory Control and Data Acquisition* (SCADA) data and for every time period the package (v, P) is transmitted to a central server forming the digital thread. See figure 1.4 for an illustration of this concept.

Having a calibrated digital twin modelling a WTG can be advantageous for several reasons. For this thesis the main use is the ability to test a developed model - that is, using the acquired SCADA data on an existing WTG location as a training data for developing and calibrating a W2P-model. Other uses for a digital twin includes *monitoring* and *power prediction*. Monitoring a WTG using a digital twin can be done by continuously calibrating θ in an online setup and tracking how it changes over time and how that affects the W2P-model[†]. For example, if the power curve based on $W2P(\cdot, \theta)$ is significantly lower than earlier then this could be a sign of degradation of the WTG or malfunctioning parts. Power prediction is a somewhat a more straight forward use of a digital twin. Given a

short time weather forecast, a digital twin can give predictions of the generated power and these predictions are used extensively in the energy market [55]. Monitoring and prediction will not be investigated in this thesis, but are mentioned as they are closely linked to the concept of a digital twin.

1.1.2 Existing solutions

The concept of modelling power output from WTG's is a well studied area. Surveying the range of applied and suggested models in the literature, a useful shorthand is categorising them into two categories: *parametric models* based on a relative small number of parameters possibly using a-priori information about the system and *non-parametric models* using a relatively large number of parameters where the model is data driven determined completely from a data set. This section will explore these two categories. Other distinctions, such as *deterministic* versus *stochastic* models, are relevant and will be explored further in chapter 3 and 4.

Parametric models In general parametric models are characterised by having a finite set of parameters where the number of parameters is fixed before data is observed [21, p. 115]. Depending on point of view the models can be either deterministic or probabilistic. The simplest W2P-model is purely deterministic and based on theoretical physics. It is based on the *power equation* for wind, stating what the power (kinetic energy per second) of the wind P_{wind} over the WTG blades is [33, p. 10]:

$$P_{\text{wind}} = \dot{E} = \frac{1}{2} \rho \pi R^2 v^3, \quad (1.2)$$

where ρ is the density of the air in kg/m^3 , R is the blade length in m and v is wind speed in m/s, which is assumed to be uniform over the area of the WTG blades. Not all energy can be extracted from the wind though and a realistic model includes a power coefficient $C_P = P/P_{\text{wind}}$ where P is the actual generated power. In 1919 Albert Betz found the surprising result that even under ideal conditions the power coefficient is significantly below unity. He found that, under some assumptions, $C_P \leq 16/27$ even under idealized conditions [33, p. 19]. This result is referred to as *Betz's law* or *Betz's limit* stating that maximum power generated for a WTG is $P_{\text{max}} = \frac{16}{27} \rho \pi R^2 v^3$. A general W2P-model must then fulfil the following relation [33, p. 19]:

$$P = \frac{1}{2} \rho \pi R^2 v^3 C_P \quad \text{for } C_P \leq \frac{16}{27}. \quad (1.3)$$

Fitting (1.3) to a given WTG would require determining C_P , which is done by simple linear regression. A few things are worth mentioning about (1.3). Firstly, it is seen that the power is proportional to the square of the swept area $P \propto R^2$ and the cube of the wind speed $P \propto v^3$. Thus, installing large WTG's at windy places maximises the generated power and in this context offshore WTG's are ideal. Secondly, (1.3) is of course a simplification compared to a real physical system. Noteworthy simplifications includes the assumption of constant wind speed across the swept area, a constant air pressure, unified direction of wind flow, no turbulence and no loss of energy due to heat transfer [33]. Other than the simplifications of the physics there is also an issue of measuring the variables in a real

modelling scenario. If sensors for wind speed measurements are located behind the blades relative to the wind direction, the measured wind speed is smaller than the one applying force on the blades. This is due to a drop in wind speed which is caused by the blades themselves [33, p. 18]. In conclusion: The model (1.3) is a simple model derived from theoretical considerations and simplifications, but it lacks a more detailed consideration of the physical phenomena in play.

A large range of parametric models are described in the literature. Many of these are not based on the power equation in (1.2), but are rather *data driven* in the sense that they infer information about the W2P model solely through observation without considerations of physical effects. A survey from 2013 [39] reviewed several of these and in an accompanying paper it is found that a 5-parameter logistic model showed the best performance [40]^[3]. It is given by:

$$P(v; \theta) = d + \frac{(a - d)}{\left(1 + \left(\frac{v}{c}\right)^b\right)^g} \quad (1.4)$$

where $\theta = (a, b, c, d, g)$. The parameters have the interpretation that a is the intercept at $v = 0$, b is the asymmetry factor, c determines inflection point, d is the limit as $v \rightarrow \infty$ and g is the steepness of the function. In W2P models $a = 0$, $d = P_{\text{nom}}$ and (b, c, g) are fitted based on observed data. Figure 1.5 shows a few parametrisations of the model.

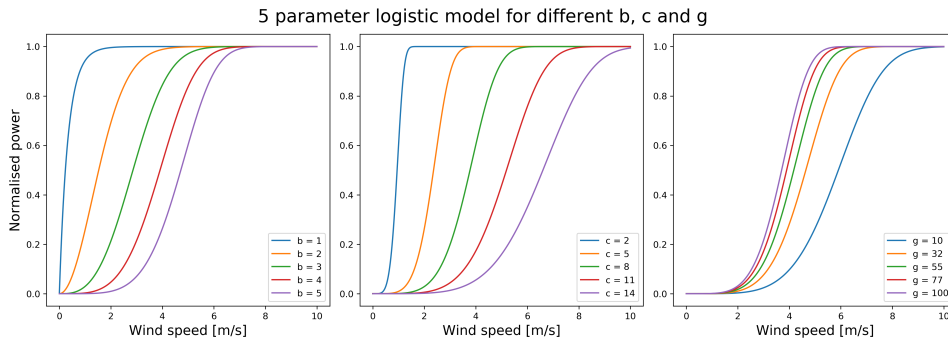


Figure 1.5: Example of 5 parameter logistic model for wind speed between 0 and 10 m/s. The default parameters are $a = 0$, $b = 5$, $c = 10$, $d = 1$, $g = 30$ and the subplots show the model with different choices of b , c and g .

^[3]Performance of parametric models including a *linearised segmented curve*, *4-parameter logistic model* and the *5-parameter logistic model* was compared to non-parametric models including a *neural network*, *data clustering model* and *data mining algorithm*. The performance was measured in as the root mean square error and mean absolute error across 5 different data sets and it was found that the 5-parameter logistic model showed the best overall performance.

Non-parametric models With non-parametric models large datasets are used in order to calibrate a model that matches input to target data. Just like a parametric model, a given instance of a non-parametric model will have a finite number of parameters, but what separates them from parametric models is the fact that the number of parameters often scales with the number of samples in the training set [21, p. 115]. So for non-parametric models, the larger the training set is, the more advanced a model can potentially be. Uses of non-parametric models for W2P models includes cubic spline interpolation, fuzzy methods, k-nearest neighbours and neural networks [39]. Goldwind currently uses a W2P model based on a neural network and this is explored further here.

Neural networks are popular in machine learning. By using neural networks, connections between more advanced variables such as turbulence intensity and output can be discovered and if the architecture is chosen appropriately and enough training data is available, then the neural network is able to learn the patterns relating input to output using multidimensional optimization methods^[4]. The model currently employed by Goldwind is also a neural network, specifically a neural network with a *feed forward* architecture. A feed forward neural network consists of a number of layers each representing an affine operation of the input followed by an element wise non-linear operation. For the input layer Goldwind takes wind speed, air density, wind shear, turbulence intensity and terrain inclination as input. The architecture itself is a 4-layer feed forward neural network which gives an input-layer, an output-layer as well as two hidden layers. The network has been calibrated to model different types of WTG's keeping the same structure. This has left Goldwind with a functional network capable of simulating the power output of a WTG with reasonable accuracy over a wide range of WTG's installed at various locations[†].

Another important model is the one set by the International Electrotechnical Commission (IEC). The IEC 61400-12-2:2013 standard, titled *Power performance of electricity-producing wind turbines based on nacelle anemometry*, specifies how power performance from a WTG should be verified [30]. It also describes how to obtain the AEP estimate previously mentioned. The standard specifies that a power curve should be fitted using the *methods of bins* where the power curve is determined as the average over certain wind speed bins [39]^[5]. As an example if bin i spans wind speeds between $v_{0,i}$ and $v_{1,i}$ then (\bar{v}_i, \bar{P}_i) is a point on the power curve where \bar{v}_i and \bar{P}_i is the average wind speed and power respectively for wind speeds observed in bin i given a set of observations. Linear interpolation can be used to get values between the bins if necessary.

In general, non-parametric models can be very accurate but they also has some downsides which is discussed in the following section.

^[4]See [5] for general theory about machine learning and neural networks.

^[5]The full IEC 61400-12-2:2013 publication have not been available to the authors, so the methods presented here is based on the description of the standard by [39].

1.2 Methods and considerations

This thesis seeks to design a W2P-model, but before stating the research question some considerations are made.

Contract W2P model The intended application for a W2P-model is relevant for the model design since it determines what prior information is available for calibration and which inputs are accessible to the model. For the sake of the discussion here, W2P-models are divided into categories: Models for *wind resource assessment* (WRA) before a WTG is physically installed and models used in *site operation* after installation. WRA is a multi-step process where a specific site is assessed for its financial potential - an important output here is the AEP estimate [33, p. 117]. Analysis for WRA can be performed using 3rd party software giving different models and statistics for a specific site [33]. windPRO [60] is one such software package and available for this thesis are a various statistics given by windPRO in some lookup tables[†]. Examples of statistics in the lookup tables are wind speed distributions and measures for turbulence - see chapter 2 for further description. A W2P model designed for a WRA framework given statistics about the site and calibrated using SCADA from existing WTG projects will be referred to as a *contract W2P model*. On the other hand a W2P-model designed for site operation after installation has access to SCADA measurements from the same site and has a wider range of applications such as monitoring and power prediction as mentioned in the previous section. A W2P model designed for this framework will be referred to as an *operational W2P model*. Figure 1.6 illustrates the two frameworks. Both an operational and contract W2P model can be thought of as digital twin of a WTG although the former having direct access to SCADA from the WTG is more closely linked to this concept.

For Goldwind designing a contract W2P-model is the main interest primarily since the model can be used to give power curves and AEP estimates. The more accurate the contract W2P-model is, the less uncertainty has to be accounted for in the power curve and AEP estimation which means that a higher percentage of the modelled power can be reported - e.g. 99% of the predicted power instead of 95%[†]. Exploiting digital twin applications like real time calibration or having time dependencies in the model is certainly also in Goldwind's interest, but it is secondary to having an accurate W2P-model for AEP estimation. Of course model accuracy could potentially be improved by incorporating e.g. time dependencies or real time calibration, but options like these are not available for a contract W2P model since power curve and AEP are delivered to the customer before the WTG is installed. Therefore the focus for this thesis will be to design the a contract W2P model and further exploitations of the digital twin will not be considered.

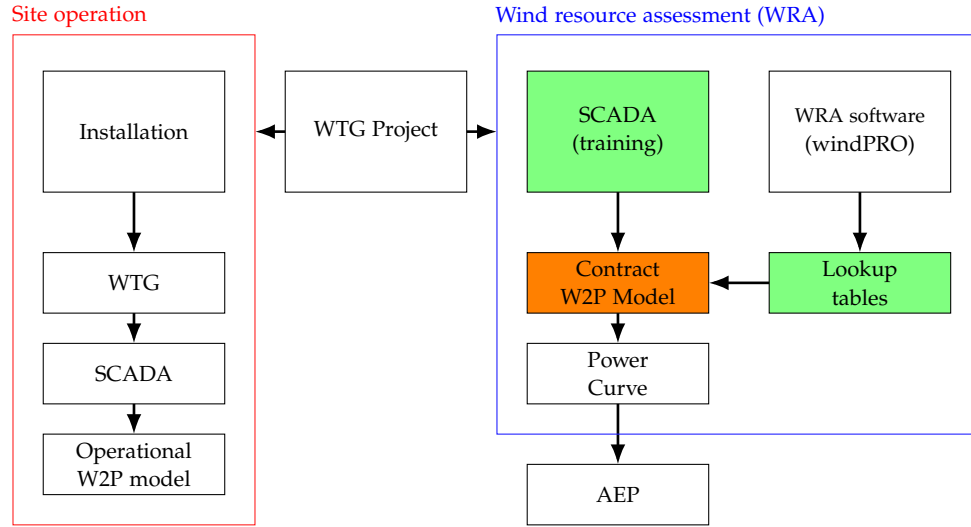


Figure 1.6: System diagram for a WTG project. The orange box is the focus of this thesis and the green boxes contains the data supplied by Goldwind Denmark.

Methods for designing W2P model The methods discussed for modelling power generation from WTG's have various levels of complexity both in terms of calibrating and evaluating the models of which the non-parametric methods are potentially the most computationally complex models in both regards. The modelling capabilities of non-parametric methods and neural networks can be unquestionable in particular when enough data is available and since the solution currently employed by Goldwind is also a neural network a natural conclusion could be that the design of a W2P model should be based on a neural network as well. Based on a few considerations though, this thesis will consider parametric models.

The point that the current solution employed by Goldwind is a non-parametric model might not be an argument for using these here. Even though it would be natural to follow the current design regime, this could also simply lead to similar models to what is currently being used without much insight or model accuracy gained. Secondly, and more importantly, this thesis seeks to gain insight into the complex interactions between input variables such as wind speed, turbulence, wind shear etc. and for this a non-parametric model is not the obvious choice. Non-parametric models, like neural networks, can achieve very precise models but they do so by treating the system as a black box model and then tries to learn patterns from a large data set. A calibrated non-parametric model can be precise at mapping input to output but how the input parameters interact to produce the output can be difficult to understand, in particular for models with high number of parameters. Parametric models on the other hand are often easier to interpret and can be designed based on the physical understanding of how the system works, possibly incorporating statistical properties of various variables like the ones available in the lookup tables. Additionally, parametric models might be able to generalise better to different WTG types, terrains, etc. than non-parametric by using the statistics for the specific site. Some non-parametric models like neural networks might also be able to use the statistics as inputs,

but it may be difficult to learn the dependencies.

There is certainly a conflict between the goal of designing the most accurate W2P model, which is maybe most easily achieved by non-parametric models, and the goal of gaining insight into the system, which is most easily achieved using parametric models, but for this thesis the latter goal is deemed more important. Therefore, this thesis will avoid using non-parametric methods and instead focus on incorporating physical and statistical knowledge of the system in direct way using a priori information of WTG's and utilizing a parametric model.

Model uncertainties It was briefly mentioned that the sensor measuring wind speed is inaccurate if located behind the blades, which is the case for the available SCADA data. Of course, the relation between measured wind speed and the one acting on the blades could be modelled, but there might be a lot of uncertainty in the model. In general, the problem of input uncertainty is a known issue to Goldwind Denmark in particular in regards to the variables in the lookup tables like turbulence intensity of the wind speed which are stochastic in nature - see chapter 2. If the wind speed v follows some probability distribution then:

$$v \sim p_v(v; \beta_v),$$

where p_v is the probability density function (pdf) parametrised by β_v . If the properties of p_v are known, the uncertainties of the W2P model can be analysed and possibly used to increase to accuracy of the model. Therefore this thesis will also seek to analyse the uncertainties of the input variables and use them directly in the modelling of the W2P model. The uncertainties will be analysed using *Bayesian methods* where prior information about a specific WTG can be incorporated into the model and used for calibration of parameters.

1.3 Research question

Climate change is a serious issue and shifting from fossil fuel to renewable energy sources like wind power could be a way to combat the problem. One area of research is modelling power generation from WTG's where better models are needed in order to make accurate power curves, estimates of AEP, predicting short term power generation and monitoring. This thesis seeks to design such - specifically a contract W2P model which is calibrated for a particular wind farm prior to installation, in a wind resource assessment framework. Parametric models will be used with the aim of modelling power production based on the physical understanding of processes that affect this using the statistics about the site given through the lookup tables. The available input data is known to be associated with a significant degree of uncertainty and the W2P model should take this into account using Bayesian methods. This is summarized in the research question for this thesis:

How is the power generation from a wind turbine generator modelled using parametric models accounting for uncertainties using Bayesian methods?

In order to answer this, the following study questions are considered:

1. Which physical processes should be modelled in order to characterise the W2P model?
2. How should the model itself be designed?
3. How can uncertainty be taken into account?
4. How is the model calibrated?
5. What are the statistical properties of the model, and how can these be used to analyse it?
6. How does the model perform and is there a performance gain compared to existing models?

1.4 Methodology

The thesis revolves around designing a contract W2P model given the inputs shown in the block diagram in figure 1.6. Figure 1.7 shows which elements will be used to design the block for the contract W2P model. The chapters for the thesis are laid out as follows.

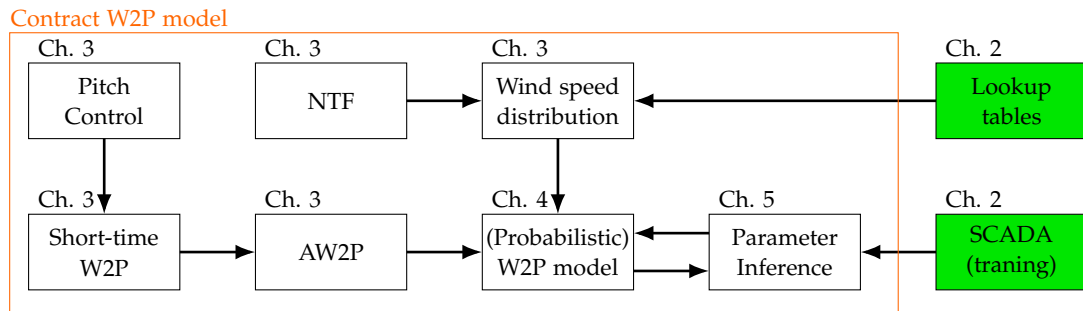


Figure 1.7: Contract W2P model unfolded

Initially chapter 2 describes the lookup tables and available SCADA data in detail. Then chapter 3 models wind speed probability distributions under two different assumptions using statistics given in the lookup tables. For one of the assumptions the relation between measured wind speed and the wind speed acting on the WTG blades is modelled using a nacelle transfer function (NTF) which is also described in chapter 3. The chapter then goes on to develop an aggregated W2P (AW2P) model, which is a model for power over 10 minutes based on a short time W2P model. The short-time W2P model is based on theory about pitch control which is how power production is controlled for the type of WTG modelled here. In chapter 4, theory about Bayesian methods is described which is then used to combine the (deterministic) AW2P model and wind speed distributions into two different *probabilistic W2P* models. The last part of the block is parameter inference which is covered in chapter 5 using the *Metropolis Algorithm*. The remaining chapters are chapter 6 where the performance of the developed methods is compared to existing methods, chapter 7 is the discussion, chapter 8 the conclusion and finally chapter 9 suggests further

studies. Appendix A is a list of used abbreviations, appendix B are various mathematical definitions and results, appendix C are additional WTG engineering aspects and additional modelling considerations, lastly, appendix D describes a few selected scripts.

1.5 Delimitations

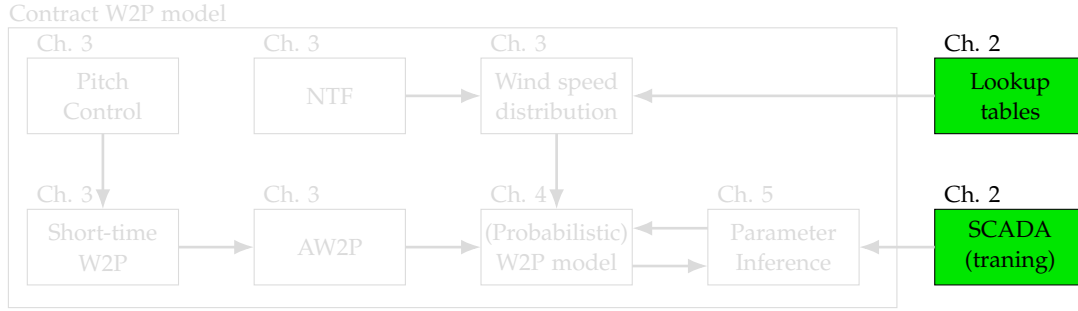
It was mentioned that the variables available for input to the W2P model are known to be associated with a degree of uncertainty. This might also be the case for the measured power output P although it is a direct measurement and therefore more reliable than e.g. the wind speed. Therefore P is treated as the true value throughout the thesis.

The nominal power for a WTG, P_{nom} , is specified for a particular WTG model, but it was found that the observed nominal power in the used data and the one according to the specifications showed disagreement. This discrepancy can be explained e.g. by changes in manufacturing or calibration differences. Effects like these will not be modelled so the effective P_{nom} will be inferred from data. Data from two WF's are available and P_{nom} is computed as the median power for observed wind speeds above the rated wind speed separate for each wind farm - i.e. one separate P_{nom} for each wind farm.

Finally, the physics of WTG's is a well studied area based on Newtonian mechanics and fluid dynamics to name a few. Theory about the physics of WTG's will be used in the thesis, but only to the extent that it helps answering the research and study questions. Thus some of the more subtle mechanics of WTG's are left out.

2 | Available data

This chapter describes the available data provided by Goldwind Denmark which consists of SCADA data and lookup tables as highlighted in the figure below.



The provided data are from two wind farms placed in South America and Asia with 32 WTG's in each. The wind farm in South America is coastal whereas the wind farm in Asia is mainland. SCADA measurements are available for the South American wind farm in the time period 22-03-2018 to 11-09-2019 and for the Asian wind farm in the time period 01-01-2017 to 31-12-2018. The relative position for WTG's at both sites can be seen in figure 2.1.

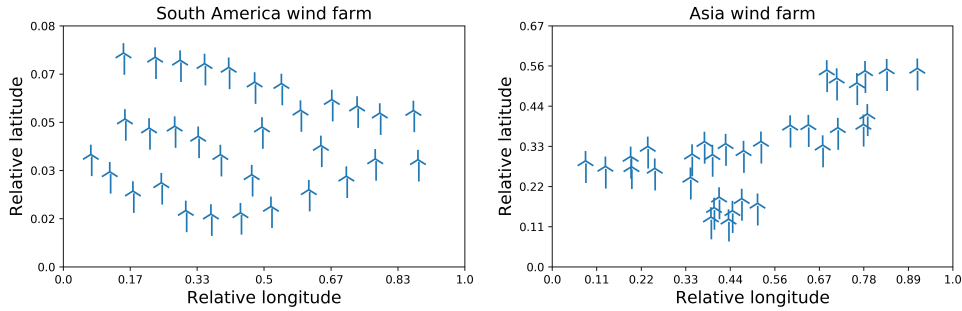


Figure 2.1: Relative position for the South American and Asian wind farm. The blue markers are the positions of individual WTG's

The WTG's in both wind farms are the model GW 121/2500 with the following specifications: Nominal power of 2500 kWh, a rotor hub height of 90 m, cut-in wind speed v_c of 2.8 m/s, rated wind speed v_r of 9.3 m/s and cut-out wind speed v_{co} of 22 m/s [13]. Before introducing the data, some WTG terminology is shown in figure 2.2.

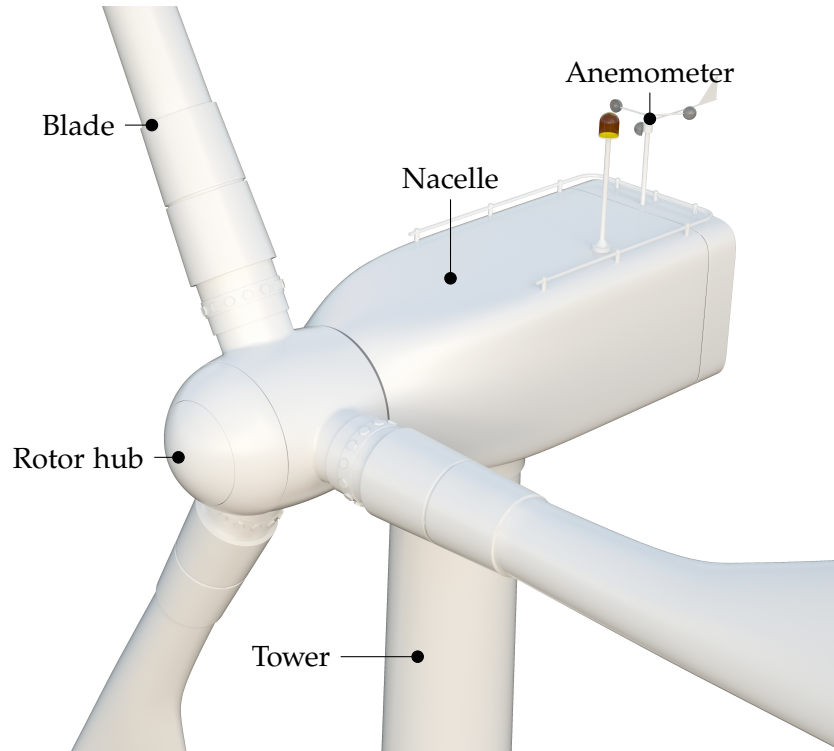


Figure 2.2: Illustration of WTG terminology. Image source: Colourbox

As observed in the figure, the blades are the three arms connecting to the rotor hub. The rotor hub is the head of the wind turbine and the collective of blades and rotor hub is referred to as the rotor system. The nacelle is the housing of all the internal components e.g. motor, gears and electric components. On the nacelle is the anemometer measuring wind speed and direction. Below the nacelle is the tower. [33, ch. 9].

2.1 SCADA

SCADA is a system of sensors taking different measurements at some sampling frequency, logging the measurements which is used for monitoring and control of a WTG [33, p. 231]. Each WTG in the described wind farms is equipped with a SCADA system and measurements are available in 10 minute resolution. The provided SCADA contains many measurements for each 10 minute period - in some cases several hundreds - but only a few of these are relevant here. The relevant SCADA data includes:

- 10-minute average horizontal wind speed measured by an *anemometer* in m/s. The wind speed is measured behind the rotor system at the nacelle as shown in figure 2.2.
- 10-minute average horizontal wind direction (WD) measured by a *wind vane* in degrees with respect to north.
- 10-minute average temperature measured in Celsius.
- 10-minute total generated power measured in kWh rounded to the nearest whole kWh.
- Status variables for each 10 minute period indicating the status of a WTG in different ways. The status variables that are of interest here are variables that indicate when a WTG is either malfunctioning or not operating. Some examples of these are: *Manual shut-down* and *Maintenance*.

An example of how SCADA is structured is shown in table 2.1

WF	WTG nr.	Time	WS [m/s]	Temp. [°C]	WD [°]	P [kWh]
Asia	1	2017-01-01 00:00	5.00	21.57	173.23	134
		2017-01-01 00:20	4.53	21.69	172.61	102
		2017-01-01 00:30	3.71	21.56	185.60	71
		2017-01-01 00:40	3.36	21.52	179.69	52
		2017-01-01 00:50	3.63	21.42	178.76	59

Table 2.1: Example of SCADA data for the Asia wind farm. Empty rows refer to repeated values. The shown values are based on the available data but random noise has been added due to confidentiality.

To visualise the SCADA data figure 2.3 and 2.4 show histograms of wind speed and wind direction from the South American and Asian wind farm.

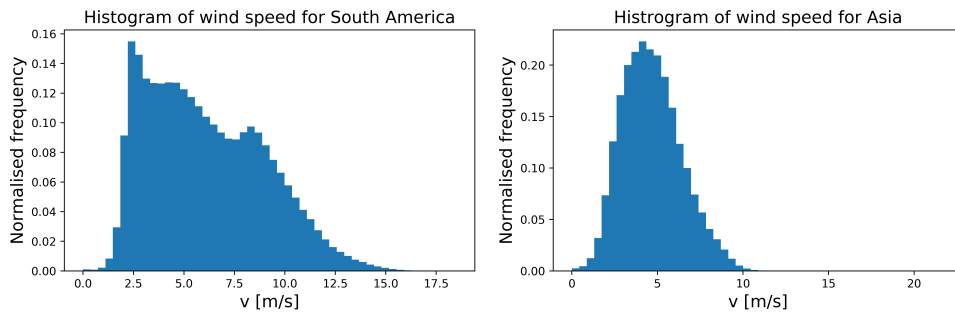


Figure 2.3

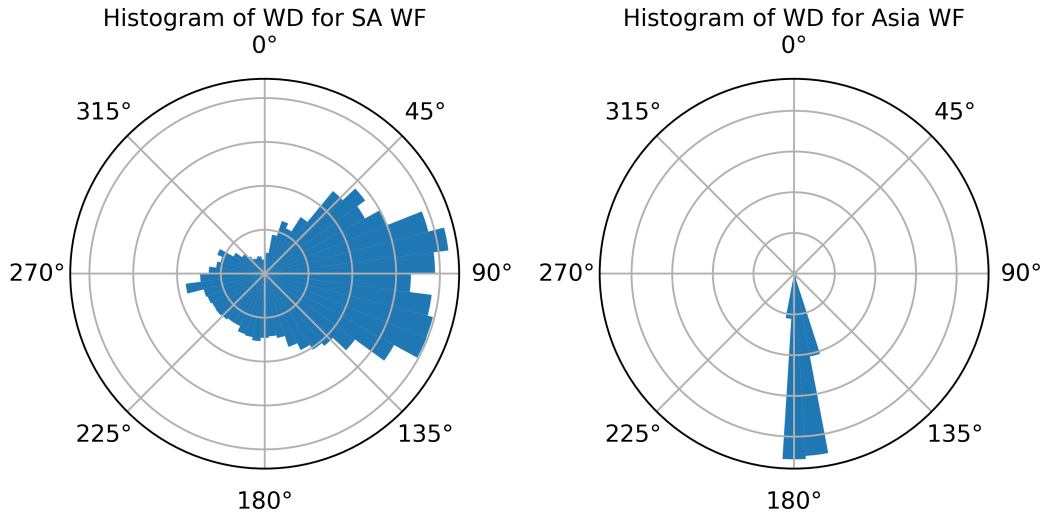


Figure 2.4

The distribution of wind speed is seen as right-skewed for both wind farms, however the histogram for South America seems multimodal where the histogram for Asia is unimodal. Higher wind speeds are more frequent for the South America wind farm which aligns well with the fact that it is a coastal wind farm - e.g. because of flat terrain. Figure 2.4 show that the prevailing wind directions are from East and South for the Southern American and Asian wind farm, respectively, with the prevailing wind direction being more dominant in the Asian wind farm. Statistics for the wind speed is described further in section C.1.

2.2 Lookup tabels

The available lookup tables are based on calculations from windPRO - a 3rd-party wind resource assessment software package. It provides general statistics for each WTG in the two wind farms. The relevant statistics are described here.

Atmospheric condition The lookup tables gives the estimated mean air density $\bar{\rho}$, mean temperature \bar{T} and mean air pressure \bar{p} at hub height for each WTG. Air pressure and temperature are related through the ideal gas law for a volume of gas [33, p. 42]

$$pV = nRT, \quad (2.1)$$

where p is the pressure in Pa, V is the volume in m^3 , n is number of moles, R is the gas constant in $\text{m}^3 \cdot \text{Pa} / (\text{K} \cdot \text{mol})$ and T is the temperature in Kelvin (K). Air density is a measure of mass-per-unit volume. When measured in mol/m^3 it is given by the number of moles per unit volume:

$$\rho = \frac{n}{V} \Rightarrow \rho = \frac{p}{RT} \quad (2.2)$$

Using that the molecular mass of dry air is $M = 28.9644 \text{ g/mol}$ and the gas constant $R = 8.31432 \text{ J/(mol} \cdot \text{K)}$ the air density in SI units is [33, p. 42]:

$$\rho = \frac{p}{RT} \frac{M}{1000} \left[\frac{\text{kg}}{\text{m}^3} \right] \quad (2.3)$$

Here it is seen that the density is given indirectly by pressure and temperature. Theoretically air density also depends on humidity, but the humidity is not available in the data so this is not considered here.

Turbulence intensity In general, turbulence is a measure of the variability in wind speed. One measure is *turbulence intensity*, which is defined as the ratio between the standard deviation and the mean of wind speed over intervals of 10 minute [33, p. 105]^[1]

$$\text{TI} = \frac{\sigma_v}{\bar{v}}. \quad (2.4)$$

The lookup tables provide different statistics for the turbulence intensity at different wind speeds. The relevant statistics provided for this thesis are the mean and standard deviation for TI - i.e. σ_{TI} and μ_{TI} . These statistics are given for each WTG with values for 12 different wind direction sectors separated by 30° and with unique values for wind speeds from 3 m/s to 30 m/s with a resolution of 1 m/s. An example of a TI statistics is shown in table 2.2.

WTG nr	Sector	Statistics	3 m/s	4 m/s	...	30 m/s
1	N	μ_{TI}	0.140	0.115	...	0.092
		σ_{TI}	0.059	0.043	...	0.013
	NNW	μ_{TI}	0.151	0.124	...	0.039
		σ_{TI}	0.053	0.045	...	0.012
	:	μ_{TI}	:	:	...	:
		σ_{TI}	:	:	...	:
32	N	μ_{TI}	0.169	0.125	...	0.125
		σ_{TI}	0.046	0.040	...	0.025
	NNW	μ_{TI}	0.164	0.134	...	0.074
		σ_{TI}	0.046	0.040	...	0.021
	:	μ_{TI}	:	:	...	:
		σ_{TI}	:	:	...	:

Table 2.2: Example of available TI statistics. The statistics shown are based on the lookup table from the South American wind farm with random noise added due to confidentiality.

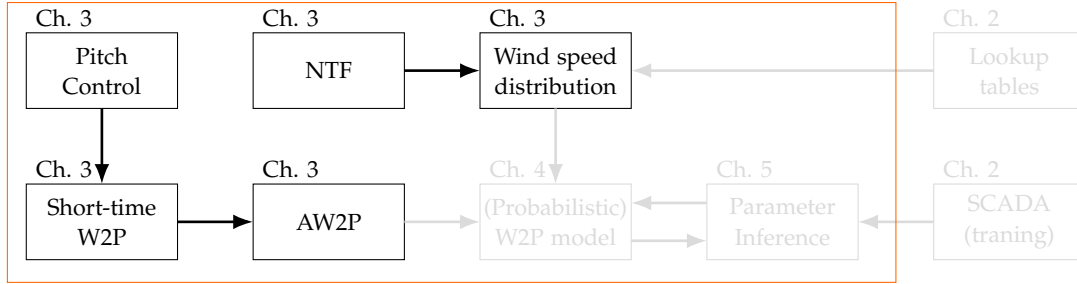
μ_{TI} and σ_{TI} decreases with wind speed in the lookup tables. This implies that more variation is observed for lower wind speeds relative to the mean wind speed.

^[1]Turbulence intensity can be interpreted as a measure of spread defined similarly to the *coefficient of variation* which is $100 \cdot \sigma / \mu$ for a random variable with mean μ and standard deviation σ [15, p. 78].

3 | WTG modelling

Various existing W2P models were discussed in section 1.1.2. This chapter will further explore how to model the produced power from a WTG specifically with the available data presented in chapter 2 in mind. Initially statistical models for wind speed is introduced followed by the theory behind *pitch control*. A W2P based on pitch control is presented along with some existing models and lastly, how to generate synthetic data is described. This is illustrated in the highlighted boxes below.

Contract W2P model



3.1 Wind speed modelling

SCADA provides 10 minutes nacelle wind speed averages denoted \bar{v} . These measurements are based on 1 Hz anemometer readings and the nacelle wind speed over the 10 minutes should follow some distribution with the mean approximately equal to the sample mean \bar{v} . The lookup tables also provide statistics about the wind speed standard deviation within 10 minutes indirectly through the turbulence intensity. As described in section 2.2 the turbulence intensity is defined as $TI = \sigma_v / \bar{v}$, and the lookup tables provide $E[TI] = \mu_{TI}$ and $\text{std}[TI] = \sigma_{TI}$ for each WTG's under different wind speeds and wind directions. Conditioning on \bar{v} gives the mean and variance for σ_v :

$$\mu_{\sigma_v | \bar{v}} = E[\sigma_v | \bar{v}] = E[\bar{v} TI | \bar{v}] = \bar{v} E[TI | \bar{v}] = \bar{v} \mu_{TI | \bar{v}} \quad (3.1)$$

$$\sigma_{\sigma_v | \bar{v}}^2 = \text{Var}[\sigma_v | \bar{v}] = \text{Var}[\bar{v} TI | \bar{v}] = \bar{v}^2 \text{Var}[TI | \bar{v}] = \bar{v}^2 (\text{std}[TI | \bar{v}])^2 = \bar{v}^2 \sigma_{TI | \bar{v}}^2 \quad (3.2)$$

As described in section C.1, the 10-minute average wind speeds \bar{v} follows a Weibull distribution with statistics available in the lookup tables. The models in this thesis, though, require statistics for shorter time periods - e.g. 1 second. Let v_t denote the wind speed at time t where t is a discrete variable and Δt is the period between consecutive samples e.g. v_t and v_{t+1} . As a simplification, the wind speed is assumed constant within periods of Δt . In chapter 4, Bayesian statistics are used to model the produced power and for this, the conditional distribution for v_t given \bar{v} is needed. Several effects complicate this:

1. **Temporal dependency** In WTG engineering it is common practice to work with wind speeds as 10-minute averages based on an assumption of a *spectral gap* for this

time period which simplifies some modelling consideration. However in general, the wind speed spectrum is non-white which means that wind speed features some temporal dependence - see section 3.1.3.

2. **Spatial dependency** In a wind farm, the wind speeds that act on each WTG are expected to be dependent due to proximity. Additionally, *wake effects*, which are reduced wind speeds and increased turbulence from other upwind WTG's, also affect the spatial dependencies. Given the 10-minute average \bar{v} for each WTG, the spatial dependencies should be somewhat lower but it might still have an effect on the distribution.
3. **Blade shadowing** The 10-minute average \bar{v} , as described in section 2.1, is based on nacelle measurements taken behind the blades. If the WTG is rotated towards the wind, the nacelle measurements are lowered compared to the effective wind speed acting on the blades due to wake effects from the blades and other parts of the WTG. The wind in front of the blades is referred to as the *free wind speed* denoted v_f and the nacelle measured wind speed is denoted v_{nc} . See section 3.1.2 for further description.
4. **Point measurement vs. vector field** The free wind speed v_f is different for every point in the swept area. Additionally, the direction might have more than the horizontal component. A vector field where each point in space has a 3-dimensional wind vector characterises this.

Modelling all the above mentioned effects is beyond the scope of this thesis, so some simplifying assumptions are made. Firstly, spatial dependency is simplified by assuming conditional independence on the following form: Let $v_t(\mathbf{x}_i)$ and $v_t(\mathbf{x}_j)$ be the wind speed of two different WTG's in a wind farm at location \mathbf{x}_i and \mathbf{x}_j . Given the average wind speed \bar{v}_i and \bar{v}_j measured at both WTG's and t occurs within that timespan then independence is assumed:

$$\begin{aligned} p(v_t(\mathbf{x}_i), v_t(\mathbf{x}_j) | \bar{v}_i, \bar{v}_j) &= p(v_t(\mathbf{x}_i) | \bar{v}_i, \bar{v}_j) p(v_t(\mathbf{x}_j) | \bar{v}_i, \bar{v}_j) \\ &= p(v_t(\mathbf{x}_i) | \bar{v}_i) p(v_t(\mathbf{x}_j) | \bar{v}_j) \quad \forall i \neq j, \end{aligned}$$

where the subscripts in the pdf p is omitted. Secondly, in regards to describing the wind as a vector field, it is assumed that the wind in the swept area of the WTG is a uniform field with only a horizontal component. In regards to temporal dependency, this thesis assumes conditional temporal independence in the following sense: If v_t and $v_{t'}$ are observed within the same 10 minute period with mean wind speed \bar{v} and additional statistics β_v - e.g. standard deviation - then:

$$p(v_t, v_{t'} | \bar{v}, \beta_v) = p(v_t | \bar{v}, \beta_v) p(v_{t'} | \bar{v}, \beta_v) \quad \forall t \neq t',$$

Independence between v_t and $v_{t'}$ is also assumed when t and t' occur within different 10 minute periods given the statistics for each period. The advantage of assuming independence is a simpler model easing theoretical considerations, although a W2P model based on this might be less precise. To mitigate this, section 3.1.3 will describe some results for wind speed temporal dependencies, but using them is beyond the scope here.

Finally, blade shadowing will be modelled by considering the *nacelle transfer function* that

gives a relation between free and nacelle wind speed. Modelling blade shadowing is a way of modelling input uncertainty considering \bar{v} as the measured input by an average over v_{nc} and the free wind speed v_f as the true input. As presented in the research question, this thesis seeks to analyse the effect of uncertainty therefore two models will be presented: One model accounting for uncertainty through blade shadowing and one that does not. These models can then be compared analysing the accuracy of each. The models are based on different assumptions about wind speed distributions which are presented in the remaining section. The assumptions will be based on existing literature but the modelling of wind speed distribution as a whole is developed for this thesis.

Some notation should be clarified. v will generally refer to the free wind speed v_f if not otherwise specified. When needed, the distinction between free wind speed v_f and nacelle wind speed v_{nc} is made explicit. The time discrete index t is sometimes omitted when not necessary in which case v refers to an arbitrary v_t .

3.1.1 Wind speed assuming independence

Given average \bar{v} , standard deviation σ_v and assuming temporal independence, v_t is assumed uncorrelated Gaussian with:

$$v_t | \bar{v}, \sigma_v \sim \mathcal{N}(\bar{v}, \sigma_v^2) \quad (3.3)$$

The Gaussian assumption is not well documented in the literature perhaps because the distribution of $v_t | \bar{v}, \sigma_v$ assuming temporal independence is not well studied in general as it is not a common setup. The assumption of (3.3) was suggested in consultation with Goldwind, and despite the flaw of allowing v_t to be negative with some probability, it might be a reasonable approach.

To fully characterise v_t , the distribution for σ_v should also be specified. Despite σ_v being positive, a common assumption is a Gaussian distribution according to [17, p. 51]. Under this assumption, and using the statistic given in the lookup tables, it was found that $P(\sigma_v < 0)$ was significant in particular for low wind speeds where the mean standard deviation is generally lower^[1]. Negative standard deviations have no meaning and cause some numerical issues in the later models so the assumption by [17] is altered slightly here to a folded Gaussian distribution. The folded Gaussian distribution corresponds to the distribution of $Y = |X|$ where X is Gaussian with mean and variance μ and σ^2 . Here the folded Gaussian distribution is denoted $Y \sim \mathcal{N}_+(\mu, \sigma^2)$. A folded Gaussian is approximately Gaussian in particular for large (positive) means or small variances so the assumption of folded Gaussian distribution is mostly similar to assuming a Gaussian distribution of σ_v . See appendix B.1 for the pdf of a folded Gaussian. Another approach would be to use a truncated Gaussian [25] where the bell shape is retained even for inputs close to 0. This would result in a density similar to the Gaussian, when considering values of μ_{σ_v} close to 0, whereas the folded Gaussian becomes flat near 0, but nonetheless the folded Gaussian distribution will be used since it is easier to simulate.

^[1] Assuming Gaussian distribution for σ_v yields $P(\sigma_v < 0) \approx 2.4\%$ for wind speeds between 2 and 4 m/s on average with the used lookup tables. In one case $P(\sigma_v < 0) = 48\%$ occurring at 2 m/s.

Now consider $\mathbf{v} \in \mathbb{R}^N$ as the vector of wind speeds within a period of 10 minutes sampled with time period Δt . The SCADA data provides an average of \mathbf{v} and σ_v stays constant within this period. The distribution of \mathbf{v} now becomes a multivariate normal distribution with diagonal variance covariance matrix and folded Gaussian standard deviation:

$$\mathbf{v} | \bar{v}, \sigma_v \sim \mathcal{N}(\bar{v} \mathbf{1}_N, \sigma_v^2 \mathbf{I}_{N \times N}) \quad \text{and} \quad \sigma_v | \bar{v} \sim \mathcal{N}_+(\mu_{\sigma_v}, \sigma_{\sigma_v}^2) \quad (3.4)$$

Figure 3.1 shows two a histogram of the $\mathbf{v} | \bar{v}$ obtained through Monte Carlo simulation.

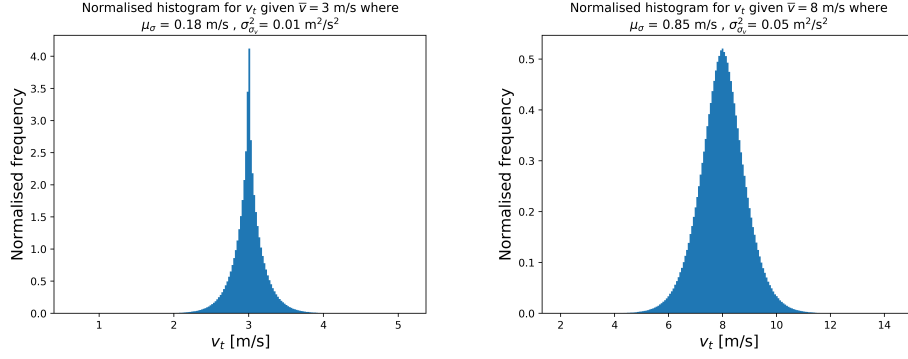


Figure 3.1: Simulation $\mathbf{v} | \bar{v}$ with lookup statistic from a WTG in the South America wind farm. Only a single element $v_t \in \mathbf{v}$ is shown here. The normalised histogram shown is obtained through Monte Carlo simulation according to (3.4).

3.1.2 Nacelle transfer functions

Recall that wind speed measurements v_{nc} are taken using an anemometer positioned at the nacelle of a WTG (behind the blades). The free wind speed v_f refers to the wind speed in front of the WTG which was previously assumed uniform across the swept area. Figure 3.2 illustrates this setup.

In WTG engineering, the relation between v_{nc} and v_f is often characterised through a nacelle transfer function (NTF) which is simply a function that links the two variables. Noteworthy literature studying NTF's includes: [37] proposing a new standard for establishing NTF's by bin averages, [43] studying the effects of turbulence on NTFs, [28] using the bin average method to establish NTF's and [52] using data from the same wind farm as [28] to calibrate NTF's over a period of 6 years to detect degradation. Finally, [11] thoroughly investigates sources of uncertainty in the NTF's, although this is done using data from *spinner anemometers* rather than *cup anemometers* which the data used in this thesis is measured with. All of the above mentioned articles have in common that they study NTF's by comparing measurements from a WTG with measurements from a nearby *meteorological tower* (met-tower). A met-tower is a tall pole usually between 50 and 120 m tall equipped with sensors at different heights in order to measure quantities like wind speed, wind direction, temperature, barometric pressure, humidity [33, pp. 82–83]. As an example, [43] compares measurements from a WTG with measurements from a met-tower distanced $2 \times$ blade diameter away from the WTG. Of course, measurements taken at a met-tower in proximity to a WTG is not a true measurement of the free wind speed, but it is a good

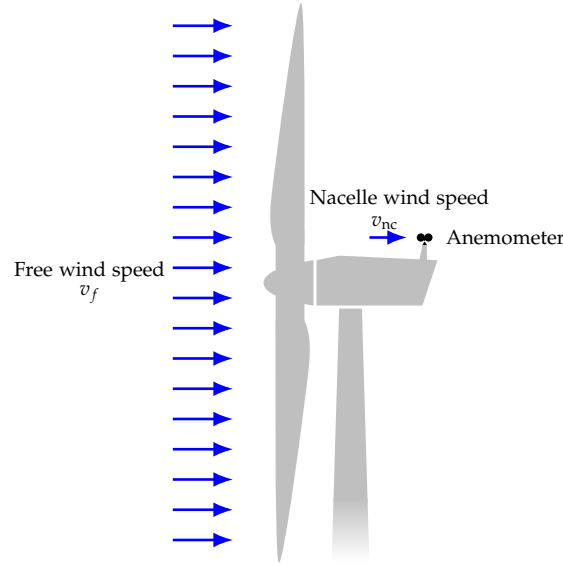


Figure 3.2: Free wind speed v_f in front of the WTG compared to the measured nacelle wind speed v_{nc} using a cup anemometer.

approximation or at least a better approximation than the nacelle measured wind speed.

Ignoring uncertainties for the moment, the sources [43, 28, 52] suggest that there is a linear relation between v_{nc} and v_f for lower wind speeds than around 14 m/s [43]:

$$v_f = \text{NTF}(v_{nc}; \boldsymbol{\eta}) = \eta_1 v_{nc} + \eta_2 \quad \text{for } v_f \leq 14 \text{ m/s},$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2)$ is the parametrisation of the NTF. For higher wind speeds than about 14 m/s [43] finds that the relation becomes non-linear, but for the sake of this thesis a linear relation can be assumed since wind speeds much above the rated wind speed $v_r = 9.3$ m/s are not directly modelled. If the parameter $\boldsymbol{\eta}$ is known the nacelle measured wind speeds can simply be converted through the NTF, but for a given WTG this requires installation of met-tower followed by a calibration. A reasonable assumption would be that if $\boldsymbol{\eta}$ has been calibrated to a specific type of WTG, then $\boldsymbol{\eta}$ also applies for a WTG of the same type at a different location given similar terrain. Still, one would expect $\boldsymbol{\eta}$ to vary from WTG to WTG, therefore $\boldsymbol{\eta}$ is seen here as a stochastic variable. To get insight into the distribution of $\boldsymbol{\eta}$, figure 3.3 shows the estimated $\boldsymbol{\eta}$ from a few articles.

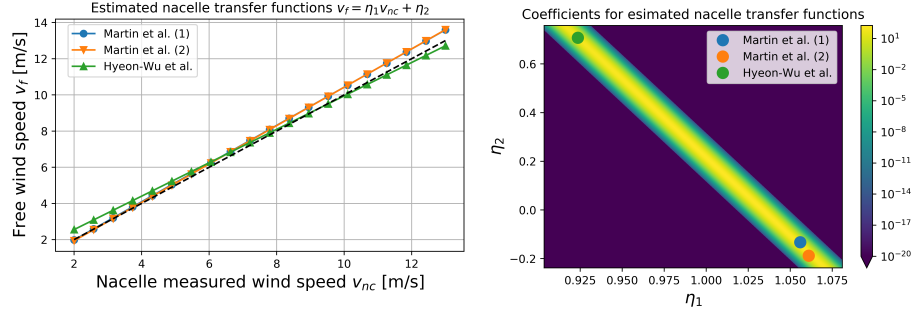


Figure 3.3: Nacelle transfer functions and parameters. [43] (*Martin et al.*) estimates two functions and [28] (*Hyeon-Wu et al.*) estimates a single. The colors in the right figure are the pdf values for a multivariate Gaussian distribution with parameters estimated from the three points giving means $\mu_{\eta_1} \approx 1.013$ and $\mu_{\eta_2} \approx 0.129$, variances $\sigma_{\eta_1}^2 \approx 6.078 \cdot 10^{-3}$ and $\sigma_{\eta_2}^2 \approx 0.252$ and covariance $\sigma_{\eta_1, \eta_2}^2 \approx -3.91 \cdot 10^{-2}$.

Figure 3.3 shows that $v_f > v_{nc}$ for most wind speeds with some disagreement for higher values. It also suggests that η_1 is negatively correlated with η_2 so a possible distribution for $\boldsymbol{\eta}$ is a multivariate Gaussian with mean $\boldsymbol{\mu}_\eta$ and variance covariance matrix $\boldsymbol{\Sigma}_\eta$ which is also shown in the figure. Guessing the type distribution from just three samples is questionable at best, but it might be a reasonable assumption for a prior distribution of $\boldsymbol{\eta}$ and due to the lack of evidence from publicly available estimated NTF's, this will be assumed. Estimating the prior distribution for $\boldsymbol{\eta}$ could be required in different situations e.g. when met-tower measurements are not available at a particular location or for modelling before a WTG is installed. A suggested for best practise for doing this is to use data from the same type of WTG with similar terrain and wind conditions. Unfortunately this thesis lacks this information so in order to illustrate the effects of modelling the NTF, the values for $(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta)$ mentioned below figure 3.3 will be used.

Uncertainties in the NTF should also be considered. Uncertainties come from random behaviour in the NTF and also from measurement uncertainties using anemometers in general. [33, pp. 92–95] mentions a few different sources of error when measuring wind speed from an anemometer - e.g. sensor calibration uncertainty, over speeding^[2], mounting effects etc [33, ch. 6]. In [11] the standard deviation, specifically when estimating free wind speed based on anemometer readings, was found to be between about 0.1 and 0.6 m/s depending on the free wind speed, but as mentioned this was done using spinner anemometers which are located on the rotor hub in front of the blades. In regards to distribution, [33, p. 92] assumes additive Gaussian errors in wind speed measurements, so this will also be used here. In conclusion, the relation between v_f and v_{nc} can be characterised through the NTF with additive Gaussian error:

$$v_f = \text{NTF}(v_{nc}; \boldsymbol{\eta}) + \xi = \eta_1 v_{nc} + \eta_2 + \xi \quad \text{where} \quad \xi \sim \mathcal{N}(0, \sigma_\xi^2)$$

For the models used in chapter 4, the distribution of σ_ξ^2 is also required. Similar to (3.4), it will be assumed that σ_ξ follows a folded normal distribution with mean $\mu_{\sigma_\xi}^2$ and standard

^[2]The anemometer might lack behind the current wind speed.

deviation $\sigma_{\sigma_{\zeta}}$. Again, the best practise would be to estimate these parameters from measurements based on the same type of WTG, but these parameters are not available and it will be assumed that $\mu_{\sigma_{\zeta}} = 0.4$ m/s and $\sigma_{\sigma_{\zeta}} = 0.2$ m/s which resembles the findings in [11].

3.1.3 Wind speed power spectrum

In regards to temporal correlation in wind speeds, one heavily cited article is [26] from 1956, which establishes the notion of a *spectral gap* for certain frequencies of measured wind speeds. In the article frequencies between $1.94 \cdot 10^{-7}$ and 0.25 Hz which corresponds to time periods of approximately 60 days and 4 seconds were analysed using estimates for the *power spectral density* (PSD) - see appendix B.2 for definition. The original graphic is shown in figure 3.4.

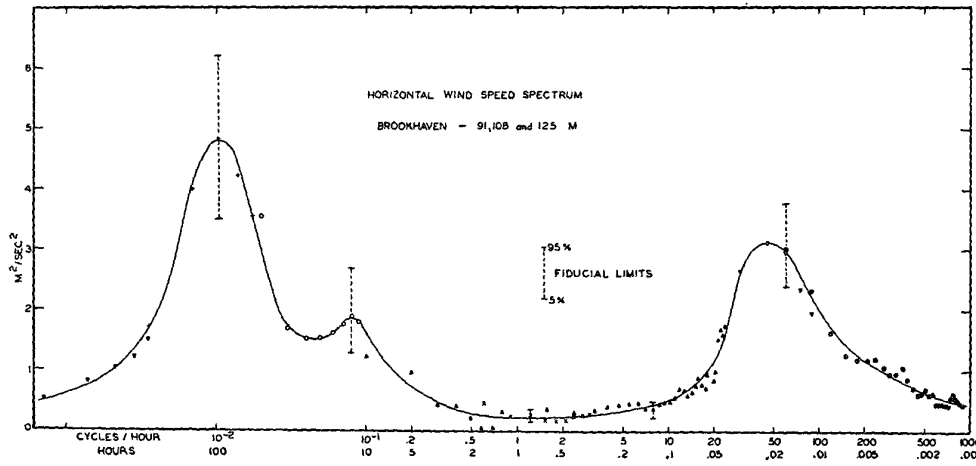


Figure 3.4: Estimated wind speed PSD in m^2/s^2 based on measurements from a meteorological tower at Brookhaven National Laboratory in the US between 1955-1956. The spectrum is pieced together from 8 separate spectra based on data sets with different sampling frequencies. The abscissa shows the frequency in cycles/hour in the first row and the time period in hours in the second row. Source: [26, p. 161] - Reprinted with permission from the American Meteorological Society

Figure 3.4 shows 3 distinct peaks occurring at time periods of approximately 4 days, 12 hours and 1 minute. In [26] it is explained that the power at 4 days is caused by fluctuations of large pressure systems in the scale of 1000 kilometers or more across. The smaller peak at 12 hours is disregarded as it is only found in some of the used data sets although it is mentioned that a 1 day spectral peak was expected. Finally, the peak at 1 minute is explained by *mechanical turbulence* caused by wind flow over irregular terrains such as hills or man-made objects like WTG's [4] as well as *convective turbulence* caused by storms and in particular thunderstorms [2]. Importantly the article argues that for time periods between 1 hour to 6 minutes "there is no physical process could support wind-speed fluctuations in this frequency range" [26, p. 164] which is the spectral gap. The spectral gap is exploited in the wind industry by modelling 10 minute averages in the middle of the spectral gap,

where the energy of the wind fluctuations is considered negligible [14]. From a fluid flow perspective, fluctuations are characterised by *eddy kinetic energy* which is the component of the fluid flow that represents fluctuation from the mean [14, 48]. By modelling wind speed as 10 minute averages these phenomena has negligible effect.

For the sake of wind speed dependence the fact that the spectrum is clearly not from a white process means that and some auto-correlation exists. A model accounting for this based directly on the PSD is a mixture of sinusoids according to the following [53]:

$$v_t = \sum_{k=1}^q [U_{k,1} \sin(2\pi f_k t) + U_{k,2} \cos(2\pi f_k t)] \quad \text{for } t \in \mathbb{Z},$$

where $U_{k,1}, U_{k,2}$ are zero mean independent random variables for $k = 1, \dots, q$, q is the number of frequencies and ω_k is the frequency index in cycles per unit time - e.g. if $f_k = 1$ the k^{th} sinusoid makes a cycle every unit time [53, p. 167]. An approximate time series can then be constructed based on an estimated PDS - e.g using the peaks in figure 3.4. Despite this it has been chosen not to model like this. Firstly modelling the dependencies would complicate the statistical models presented in chapter 4, which is beyond the scope here. Secondly, since the modelled power will later be limited to a 10 minute total, modelling the dependence in wind speed for small lags may not affect this too much. How exactly dependencies in wind speed would affect the 10-minute total power is left as an open question. Furthermore, the data provided by Goldwind Denmark is limited to the 10-minute average so data required to analyse short-time dependencies have not been available.

3.2 Pitch control

The power generated from a WTG at a given wind speed highly depends on the aerodynamics of the blades at different wind speeds. In section 1.1.2, a simple model for the power was given as:

$$P = \frac{1}{2} C_P \rho \pi R^2 v^3 \quad \text{for } C_P \leq \frac{16}{27},$$

but this kind of model is only accurate for a limited range of wind speeds when C_P is fixed. A more accurate model considers C_P as a function of the momentary wind conditions and the aerodynamics of the particular WTG. A common way to parametrise C_P is by considering the *tip speed ratio* (TSR) ψ and the pitch angle θ . The pitch angle is illustrated in figure 3.5 and the tip speed ratio is defined as:

$$\psi = \frac{\omega R}{v},$$

where ω is the rotational speed of the WTG blades in radians/sec^[3], R is the blade length in meters and v is the wind speed in m/s.

^[3]If the blades do a full rotation every T seconds the rotational speed is $\omega = 2\pi f = 2\pi/T$, where $f = 1/T$ is the frequency in Hertz.

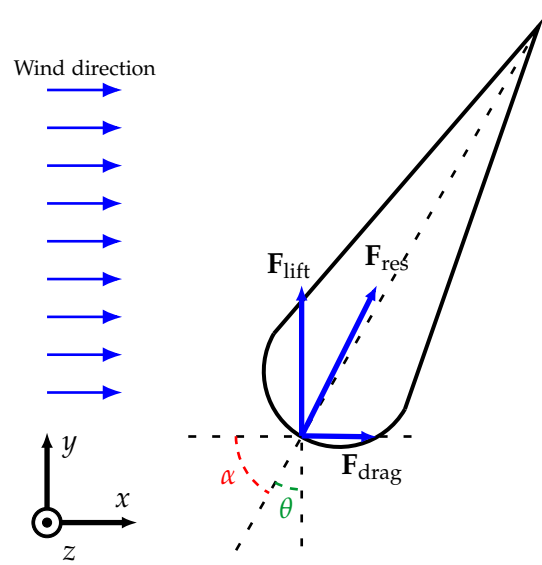


Figure 3.5: Profile of WTG blade at distance r from the rotor hub with rotation in the y - z plane and wind direction along the x -axis. The z -axis is towards the viewer. α is the *angle of attack* and θ is the *pitch angle*. Lift force F_{lift} and drag force F_{drag} are respectively perpendicular and parallel to the wind direction while the resultant aerodynamic drag force F_{res} is the addition of the two [33, p. 49]. To keep an optimal angle of attack, the pitch changes along the blade starting from almost parallel to the x -axis at the root and decreasing along the blade - see more details in [33, pp. 49–53].

WTG blades are designed to achieve maximum lift and minimum drag along the full length of the blades and these forces are highly dependent on the speed and direction of the wind - see figure 3.5 for illustrations of lift and drag. The theory behind this is quite extensive - see [33, ch. 4-5] - but for the purpose of this thesis it suffices to state that the power coefficient can be considered as a function of TSR and pitch angle. A few methods are available for modelling C_p , ranging from ones that directly models the physical behaviour to ones that are empirically fitted through observed data - see [24, sec. 2.1] for further details. Figure 3.6 illustrates an empirical approach. By computing the TSR using sensor data the pitch angle that maximises C_p is obtained. Thus, pitch control is about pitching the blades to obtain maximum C_p under different turbulence conditions. In practise the pitch can be controlled via eg. a PID-controller [61].

Pitch controlled WTG's tends to follow a power curve similar to the one illustrated in figure 1.3. The major benefit of pitch controlled WTG's is their ability to maintain nominal power for higher wind speeds, which is achieved by increasing the pitch when wind speeds approach the rated wind speed [33, pp. 75–76]. For comparison a *stall-regulated* WTG has a fixed pitch and, for higher wind speeds where the TSR tends to be lower, less efficiency is achieved according to figure 3.6.

A realistic W2P model considers the aerodynamics of the WTG and the effects from pitch control. Modelling this directly requires modelling the moment to moment interactions between varying wind speeds, turbulence conditions and pitching. This approach would perhaps require a short-time statistical wind speed model, high frequency sensor measure-

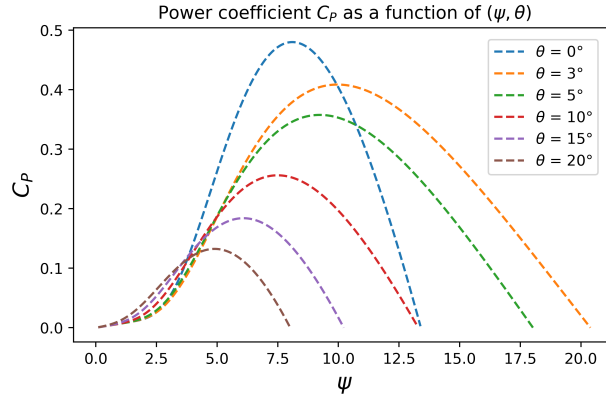


Figure 3.6: Relation between power coefficient C_P , tip speed ratio ψ and pitch angle θ ^[4]. The relation plotted here is modelled via an empirical formula with 5 parameters which can be fitted to a particular WTG depending on design and turbulence conditions. The maximum $C_P \approx 0.48$ occurs at $\theta = 0$ and $\psi \approx 8.1$. The parameters here are obtained from [24, p. 43] and do not necessarily agree with the type of WTG used in this thesis.

ments for calibration, as well as knowledge about the applied pitch control algorithm. For this thesis though, the W2P model is always the total power over a 10 minute period and it is therefore argued that direct modelling of short time interactions is an overcomplication. Instead, the effects from turbulence and pitch control will be modelled indirectly by assuming a fixed model depending only on wind speed. This model will mimic the expected effects from pitch control and any short time deviations from this are assumed to average out over the 10 minute period. As described in appendix C, power curves and AEP estimations are based on 10 minute models, so having a W2P model that operates on the same time scale is a benefit. The next section describes this model in details.

3.2.1 Pitch Control Wind Speed Model

To model the 10 minute average effects of pitch control on the power coefficient without knowledge about the tip speed ratio, pitch and pitch algorithm, a model between the power coefficient C_P and wind speed v is established. The model for C_P is develop for this thesis and a few assumptions underlie this, drawing inspiration from a literature study [38, 44], data exploration and consultations with Goldwind Denmark. As introduced in section 1.1.2 the power coefficient is defined as the ratio:

$$C_P = \frac{P}{P_{\text{wind}}}, \quad (3.5)$$

where P is the power produced from a WTG and P_{wind} is the theoretical kinetic energy in the wind given by $0.5\rho\pi R^2 v^3$. To characterise C_P as a function of v the following assump-

^[4]Note the distinction between *local* and *global* pitch. Local pitch refers to the cross sectional pitch, which is illustrated in figure 3.5 and changes along the blade while global pitch refers to pitching the entire blade independent of local pitch angles. The pitch angle referred to in figure 3.6 is the global pitch.

tions are made:

1. **Constant wind conditions:** For some windspeed, the wind conditions and resulting turbulence behaviour is always the same independent of wind direction. Note that the assumption about wind direction neglects *wake effects* which are reduced wind speeds and increased turbulence from other upwind WTG's often present in a wind farm.
2. **Constant rotational speed:** For some windspeed the rotational speed of the blades is always the same which means that the tip speed ratio is fully determined by the wind speed. This assumption is clearly not realistic since the inertia of the blades will cause some delay of rotational speed with changing wind speed.
3. **No delay in control:** Pitch control is instantaneous. In practise some delay in control will occur, but this is neglected.
4. **Constant control:** It is assumed that the blades are always pitched at the same angle for a particular measured wind speed.

With these assumptions C_P can be modelled as a function of only v and although the assumptions are quite naive, the averaging effect over 10 minutes should make them more realistic. To model the power coefficient a scatterplot of observed C_P is inspected - see figure 3.7. The relation between v and C_P in figure 3.7 features some distinct regions depending on v . The general behaviour is explained by the effect of pitch control on C_P :

- For wind speeds between 0 and $v_c = 2.8$ m/s the energy of the wind is not enough to overcome the inertia of the system and C_P is effectively 0.
- From cut-in to medium wind speeds (about 4 m/s) the rotation of the blades is low causing the tip speed ratio to be low as well. Therefore the blades are pitched at a higher angle than 0 to achieve larger C_P in this region [33, p. 76].
- From about 4 m/s to 8 m/s the pitch angle is zero and C_P is at its max which is also seen in figure 3.6.
- For 8 m/s and above the wind speed begins to dominate the tip speed ratio again causing it to be lowered and the pitch angle is increased to counteract this [33, p. 76]. In this region C_P is lowered until the rated wind speed v_r is reached.
- At v_r and above, pitching is continually increased and C_P is approximately inverse proportional to v^3 , this is seen since $C_P(v) \propto \frac{P_{nom}}{v^3}$, for $v > v_r$. The effect on the produced power is a flattening and in principle the nominal power P_{nom} should be produced for $v > v_r$ for pitch controlled WTG's.

These effects are partly seen in the SCADA data - see figure 3.9. It is observed that the pitch angle is regulated up to roughly v_c after which the pitch angle approximately becomes 0[°] and lastly above v_r the pitch angle is seen to increase.

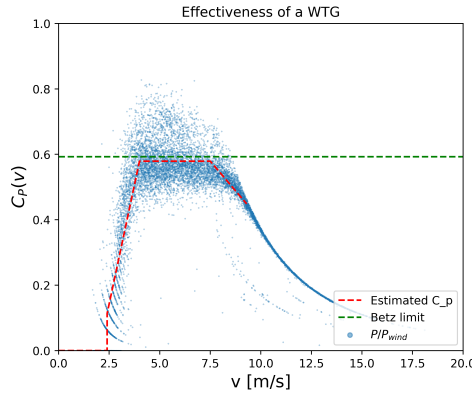


Figure 3.7: Relation between power coefficient C_p and wind speed v ^[5]. The dashed red line follows the piecewise linear function described by (3.6). The estimated C_{\max} is 0.57 in the interval $[4.5, 7.5]$.

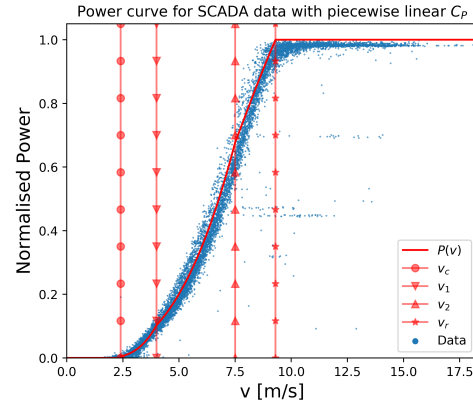


Figure 3.8: Power curve parametrised by (3.7) and (3.6)

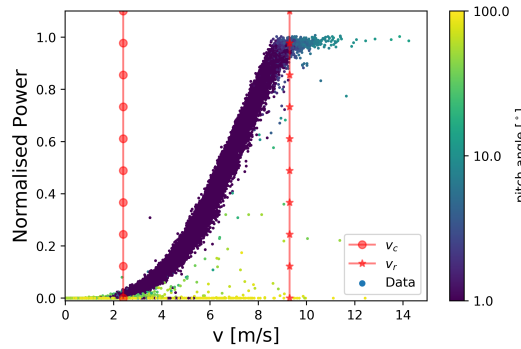


Figure 3.9: Relation between the wind speed, power production and pitch angle for raw data.

Based on these observation and the assumptions made, the following model for C_p is suggested:

$$C_p(v) = \begin{cases} 0 & \text{if } v \leq v_c \\ (v - v_1)a_1 + C_{\max} & \text{if } v_c < v \leq v_1 \\ C_{\max} & \text{if } v_1 < v \leq v_2 \\ (v - v_2)a_2 + C_{\max} & \text{if } v_2 < v \leq v_r \end{cases} \quad (3.6)$$

where the maximal efficiency is achieved between v_1 and v_2 . An example of (3.6) is shown

^[5]It is observed that a large portion of observed $C_p(v)$ in figure 3.8 is located above the theoretical upper limit $C_{\max} = 16/27$ introduced as Betz limit in section 1.1.2. General measurement error could explain this and in particular the effect of wind speed reduction behind the rotor system, introduced in section 3.1.2, explains this observation well.

in figure 3.7, where (a_1, a_2, C_{\max}) have been found using a least squares estimate, the wind speed parameters are $v_c = 2.8$ m/s, $v_r = 9.3$ m/s from the WTG specifications and $v_1 = 4$ m/s, $v_2 = 8$ m/s are qualified guesses. With the model in (3.6) it is also possible to model the power curve as a function of only v through the W2P model:

$$W2P(v; \theta) = \begin{cases} \frac{1}{2} \rho \pi r^2 v^3 C_p(v; \theta) & \text{if } v \leq v_r \\ P_{\text{nom}} & \text{else} \end{cases}, \quad (3.7)$$

where $\theta = (a_1, a_2, C_{\max})$. In principle, the W2P model in (3.7) also depends on the air density ρ . The results shown throughout the thesis uses ρ as input, but the dependence on ρ is omitted in the W2P model to ease the notation^[6]. Additionally, the wind speed parameters v_1 and v_2 could be considered parameters for the W2P model since these are not known from specification. This has not been chosen for two reasons. Firstly because a rough approximation of v_1 and v_2 can quite easily be obtained from data, and secondly because it was found that estimating these with the Metropolis Algorithm, which is presented in chapter 5, the parameter estimates tended to fluctuate much more between iterations than if using the rough approximations necessitating more iterations in the algorithm. In appendix C.2 the values are found to be $v_1 \approx 3.87$ m/s and $v_2 \approx 7.93$ m/s which is used for the remaining thesis. It is assumed that θ is the same for WTG's of the same type independent of the specific site. Since all WTG's used for this thesis are of the same type it will only be necessary to estimate one θ .

An example of the power curve using (3.7) can be seen in figure 3.8. Based on figure 3.7 and 3.8 the models in (3.6) and (3.7) seem reasonably accurate upon first inspection. More advanced models of C_p could be suggested, but for the sake of simplicity a piecewise linear model has been chosen.

To get the average effect of pitch control over 10 minutes the next section will introduce a framework in which (3.6) and (3.7) are applied only over a short timespan and then 10 minute total power is analysed.

3.3 Aggregated W2P model

An *aggregated wind to power* (AW2P) model, which is a model of total power over 10 minutes, is now be presented. Let wind speed v_t be assumed constant within intervals of period Δt [s] like in section 3.1 such that $v_t = v(\tau)|_{\tau=t\Delta t}$ where τ is a continuous variable and t is discrete. Since the $v(\tau)$ is constant over a period of Δt the total power over that period becomes proportional to the power at time e.g $\tau = t\Delta$. Additionally it is assumed that the W2P model suffers from a structural error which is modelled by additive independent Gaussian noise with total variance σ_ζ^2 over Δt :

$$P_t = \int_{t\Delta t}^{(t+1)\Delta t} P(v(\tau)) d\tau = \Delta t P(v_{t\Delta t}) = \Delta t W2P(v_t; \theta) + \zeta_t, \quad t = 0, 1, \dots, N-1, \\ \zeta_t \sim \mathcal{N}(0, \sigma_\zeta^2)$$

^[6]Air density can be calculated using (2.3) with the mean pressure \bar{p} from the lookup tables and the temperature readings in the SCADA data.

where $N = 600 \text{ s}/\Delta t$ is the number of discrete values within a 10 minute period. $\Delta t = 1$ second have been used throughout the thesis giving $N = 600$. Now the total power over 10 minutes, which is the quantity of interest, is denoted \bar{P} and is obtained through the AW2P model again using that P_t is constant over periods of Δt :^[7]

$$\bar{P} = \int_0^{10\text{min}} P(v_\tau) d\tau = \sum_{t=1}^N P_t = \sum_{t=1}^N \Delta t \text{W2P}(v_t; \theta) + \sum_{t=1}^N \zeta_t \triangleq \text{AW2P}(\mathbf{v}; \theta) + \underbrace{\sum_{t=1}^N \zeta_t}_{\bar{\zeta}}, \quad (3.8)$$

where $\mathbf{v} \in \mathbb{R}^N$ is a vector of wind speeds and $\bar{\zeta}$ is also Gaussian distributed with zero mean and variance $\sigma_{\bar{\zeta}}^2 = N\sigma_{\zeta}^2$. \bar{P} is measured in W which can be converted into kWh by dividing with $1000 \times 60 \times 60$.

As a final note, \bar{P} in (3.8) refers to the power over 10 minutes in some arbitrary interval. In a general setup \bar{P}_n refers to \bar{P} for a particular time interval n . Distinctions between different WTG's and WF's could also be denoted with additional subscripts but these will not be needed.

3.4 Alternative models

As presented in chapter 1, existing W2P models in literature include the 5-parameter logistic (logistic) model and the method of bins (bin) model which only take wind speed as input. The AW2P model also takes air density as input, so in order to make the existing methods comparable these are extended to include air density as a parameter as well.

Based on the power in the wind, (1.3) gives that power is proportional to air density - i.e. $P \propto \rho$. Now let f be a function of wind speed given by the logistic or bins model with parameters θ , then air density is accounted for using:

$$P(v, \rho) = f(v; \theta) \rho$$

As described in section 2.2 the air density can be estimated as a function of temperature T and air pressure p . Temperature is given as the 10 minute average T in the SCADA data and the overall mean air pressure \bar{p} is given by the lookup tables for each WTG - thus $\rho = \rho(T, \bar{p})$. Additionally, a W2P model should generalise to WTG's with different nominal power P_{nom} , so to account for this a normalised function $f_n(\cdot) = f(\cdot)/P_{\text{nom}}$ is used in the model:

$$P(v, T, \bar{p}; \theta) = P_{\text{nom}} f_n(v; \theta) \rho(T, \bar{p}). \quad (3.9)$$

Given an observation (P, v, T, \bar{p}) , f_n should be fitted through θ to the target variable denoted:

$$\tilde{P} = \frac{P}{P_{\text{nom}} \cdot \rho(T, \bar{p})} \quad (3.10)$$

The method to fit f is different for the two models.

^[7]Note that \bar{P} is not at average as the notation might suggest.

3.4.1 5-parameter logistic model

The logistic model for \tilde{P} is given by:

$$f_n(v; \theta) = d + \frac{(a - d)}{\left(1 + \left(\frac{v}{c}\right)^b\right)^g}$$

$\theta = (a, b, c, d, e, g)$ is found by minimizing the mean squared error with terms of the form $(\tilde{P} - f_n(v; \theta))^2$ using the Nelder-Mead method implemented in the SciPy Optimize library in Python [50] - see [18] for further description. The logistic model is fitted using the training data described in section 6.1 with wind speeds between 1.8 m/s and 10 m/s. Results are shown in figure 3.10 and 3.11.

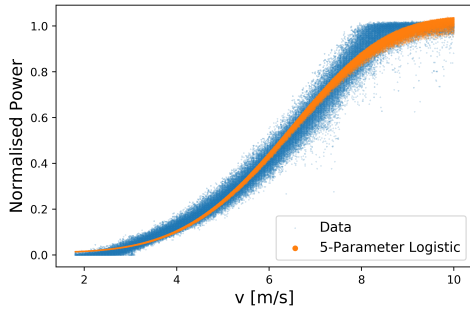


Figure 3.10: Training data and fitted five-parametric logistic power curve with $\theta = (9.1 \cdot 10^{-3}, 4.3, 1.3 \cdot 10^2, 0.86, 2.8 \cdot 10^5)$.

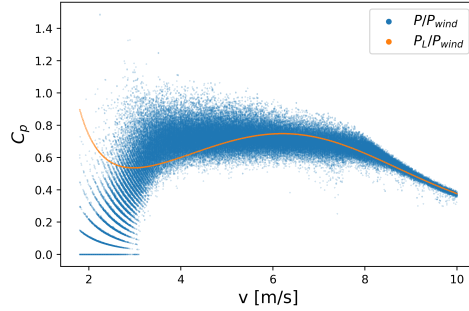


Figure 3.11: $C_p = P/P_{\text{wind}}$ for different wind speeds for SCADA data and the logistic model. P_L is the estimated power by the logistic model given by (3.9)

From visual inspection it is observed that the logistic model tends to either underestimate or overestimate the produced power for a given wind speed. For low wind speeds the logistic model approaches $a = 9.1 \cdot 10^{-3} > 0$ which causes the differences between the observed and modelled C_p seen in figure 3.11.

3.4.2 Method of bins

As described in section 1.1.2, the bin model is fitted by computing the average of wind speed and power within wind speed bins of a chosen width. Given a set of N_{dat} observations $\{(v_n, \tilde{P}_n)\}_{n=1}^{N_{\text{dat}}}$ these are grouped into subsets based on which wind speed bin v_n fall into. Denote these subsets $\{(v_{n,i}, \tilde{P}_{n,i})\}_{n=1}^{N_i}$ for each wind bin i of size N_i where $v_{n,i}$ fall in bin i . The average wind speed \bar{v}_i and \bar{P}_i within bins are then calculated according to:

$$\bar{v}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} v_{n,i} \quad \text{and} \quad \bar{P}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} \tilde{P}_{n,i}$$

Given the averages, f_n is tabulated and linear interpolation is used to get values the tabulated points. Here the bin width is chosen to be $\Delta v = 0.05$ m/s giving 164 wind bins between 1.8 m/s and 10 m/s. The training data described in section 6.1 is used to compute the averages. A normalised power curve and the associated coefficient of power as a function of wind speed can be seen in figure 3.12 and 3.13.

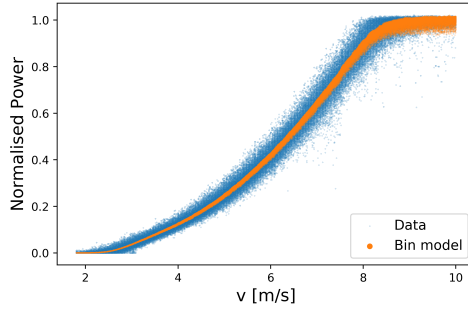


Figure 3.12: SCADA data and a the fitted bin model.

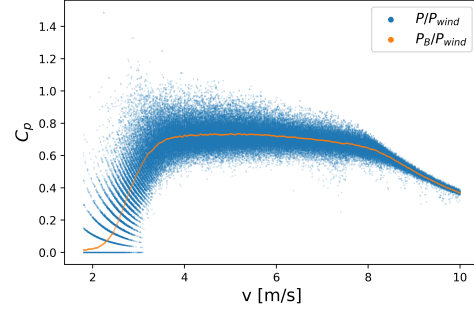


Figure 3.13: $C_p = P/P_{\text{wind}}$ for different wind speeds for SCADA data and the bin model. P_B is the estimated power by the bin model given by (3.9)

It is observed that the model fits quite well, in the sense that the power curve and C_p is centralized through the data. But this method might not generalise as well to the test data. This was also assessed in [39] where the bin model was critiqued for not accounting for the local wind conditions - e.g. turbulence - when applied at different sites.

3.5 Synthetic data

Synthetic data is later used to tabulate some wind speed distributions and is also used to generate data for a controlled test of parameter inference. The algorithm to generate synthetic data through simulation is described here.

For a 10 minute period, synthetic data contains total power \bar{P} and average measured wind speed \bar{v} . These are obtained by first simulating the free wind speed \mathbf{v}_f which is constant over periods of Δt with N discrete wind speeds for the 10 minutes. \mathbf{v}_f is simulated according to (3.4) denoting the mean wind speed μ_v . Total power is then given by the AW2P model of the simulated wind speeds with additive structural error ζ according to (3.8). If it assumed that the measured wind speed suffers from input uncertainty modelled by the NTF, then the average measured wind speed \bar{v} is modelled according to section 3.1.2, otherwise μ_v is assumed known and \bar{v} is set to this. The algorithm for simulation of data is presented in algorithm 1.

Algorithm 1 Simulating free wind speed with mean μ_v

Input : wind speed standard deviation statistics $(\mu_{\sigma_v}, \sigma_{\sigma_v}^2)$, NTF parameter statistics (μ_η, Σ_η) , NTF standard deviation statistics $(\mu_{\sigma_\xi}, \sigma_{\sigma_\xi}^2)$, W2P parameters θ , structural error statistics σ_ζ^2 and whether to include NTF or not.

for $n = 1, \dots, N_{\text{sim}}$ **do**

Simulate $\sigma_v \sim \mathcal{N}_+(\mu_{\sigma_v}, \sigma_{\sigma_v}^2)$ (WS std.)

Simulate $\mathbf{v}_f \sim \mathcal{N}(\mu_v \mathbf{1}_N, \sigma_v^2 \mathbf{I}_{N \times N})$ (Free WS)

Simulate $\zeta \sim \mathcal{N}(\mathbf{0}_N, \sigma_\zeta^2 \mathbf{I}_{N \times N})$ (Structural error)

$\mathbf{P} \leftarrow \text{W2P}(\mathbf{v}_f; \theta) + \zeta$ (W2P model + structural error)

$\bar{P}_n \leftarrow \sum_{t=1}^N P_t$ (Aggregated power)

if NTF **then**

Simulate $\sigma_\xi \sim \mathcal{N}_+(\mu_{\sigma_\xi}, \sigma_{\sigma_\xi}^2)$ (NTF structural error std.)

Simulate $\xi \sim \mathcal{N}(\mathbf{0}_N, \sigma_\xi^2 \mathbf{I}_{N \times N})$ (Structural NTF error)

Simulate $\eta \sim \mathcal{N}(\mu_\eta, \Sigma_\eta)$ (NTF parameters)

$\bar{v}_n \leftarrow \frac{1}{N} \sum_{t=1}^N \left(\frac{v_{f,t} - \eta_2 - \xi_t}{\eta_1} \right)$ (Average WS through NTF)

else

$\bar{v}_n \leftarrow \mu_v$ (Known mean WS)

return $\{\bar{v}_n, \bar{P}_n\}_{n=1}^{N_{\text{sim}}}$

To get a representative sample of synthetic data, algorithm 1 is run multiple times for different μ_v - e.g $\mu_v \sim \mathcal{U}(v_{\min}, v_{\max})$ for some v_{\min} and v_{\max} . The algorithm is implemented in `sim_data.py` - see appendix D. Note that the values μ_{σ_v} and σ_{σ_v} are given in the lookup tables, the parameters and statistics for the NTF are based on those found in section 3.1.2. θ and σ_ζ are chosen based on the setup. An example of synthetic data with and without the NTF is seen in figure 3.14 and 3.15.

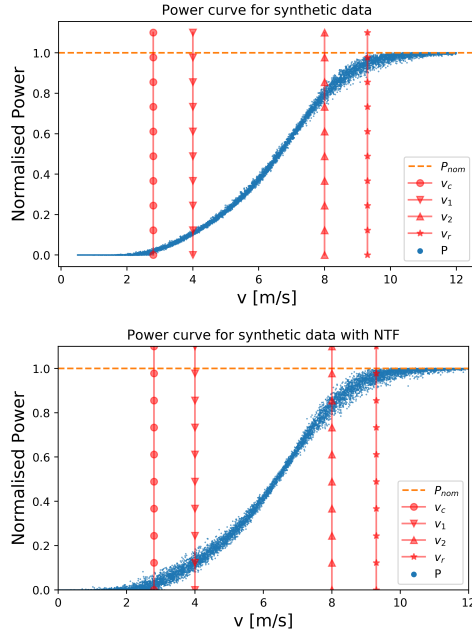


Figure 3.14: Power curve of synthetic data parametrised by (3.7) and (3.6).

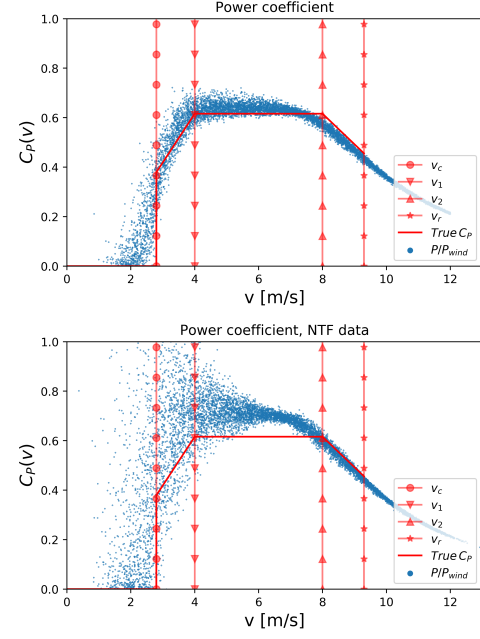
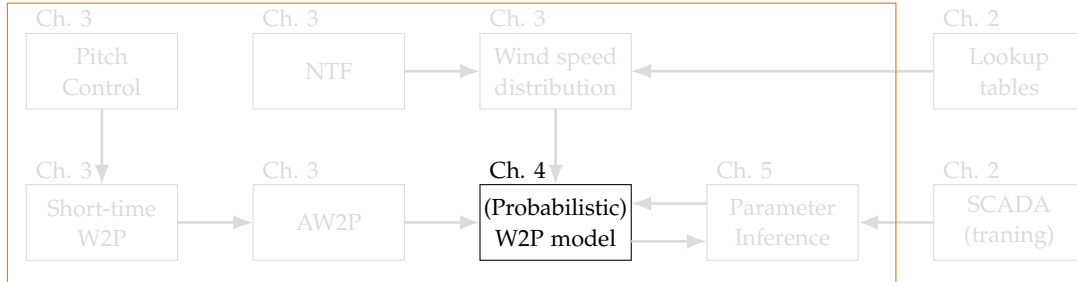


Figure 3.15: Relation between power coefficient C_p and wind speed v . The solid red line follows the piecewise linear function described by (3.6).

4 | Bayesian Inference

In chapter 3, different wind speed distributions were considered and the AW2P-model was presented. The AW2P-model gives the total power over 10 minutes, but only a single measurement, the average wind speed \bar{v} , is available so the model suffers from input uncertainty in the sense that the time-varying free wind speed is unobserved. Additionally, input uncertainty also refers to wind speed measurements suffering from uncertainty in general due to measurement errors and structural errors such as blade shadowing as discussed in section 3.1. Handling input uncertainty using Bayesian theory was discussed by [34] and [27] in the context of rainfall-runoff models using *Bayesian Total Error Analysis* (BETAE). This thesis proposes a similar method to analyse the distribution of the aggregated power \bar{P} , which leads to a method for inferring model parameters. The proposed method is similar to BETAE and will be dubbed *Approximate Bayesian Total Error Analysis* (A-BETAE). A-BETAE offers modelling flexibility and low dimensional parameters compared to the models in [34], but at the cost of using approximate distributions. The models based on A-BETAE use the AW2P model to model power directly based on the probabilistic modelling of wind speed giving probabilistic modelling of power which, to the extent of the authors' knowledge, is a novel approach. The focus of this chapter is highlighted in the figure below.

Contract W2P model



Initially, this chapter will introduce the general framework for handling input uncertainty in section 4.1, where A-BETAE is also presented. Section 4.2 discusses how to choose parameter priors in a Bayesian framework. For A-BETAE, distributions for wind speed are required thus section 4.3 presents methods to obtain these. Section 4.4 then discusses how to obtain an approximate distribution for the power \bar{P} followed by the deviation of the posterior distributions to be used in parameter inference in section 4.5. Finally, section 4.6 discusses some details about implementing A-BETAE.

4.1 Bayesian total error analysis

The following paragraph is inspired by [27]. Consider a statistical process modelled with an observed input x , a model f parametrised by θ and an additive random error ζ , such

that the output is $y = f(x; \theta) + \zeta$ - scalar values are assumed for ease of notation. This would be a common setup for modelling many processes, but some processes feature additional sources of error. In particular, if input and/or output are not directly observed additional models are needed to characterise the process. To do this consider an *uncertainty framework* where it is assumed that there exist some *true process* which maps the input x to the output y . In practise, the true process is not observed and the input x is observed with some *input error*, the model f is corrupted by a *structural error* and the output is observed with some *output error* [27]. Assuming that the input error δ , structural error ζ and output error ε , commonly referred to as *latent* parameters, are all additive the process is characterised by:

$$\begin{aligned}\tilde{x} &= x + \delta \quad (\text{input error}) \\ y &= f(x; \theta) + \zeta \quad (\text{structural error}) \\ \tilde{y} &= y + \varepsilon \quad (\text{output error})\end{aligned}$$

\tilde{x} and \tilde{y} are the observed input and output respectively. Figure 4.1 illustrates this concept. In Bayesian total error analysis models of this type are studied specifically with the aim

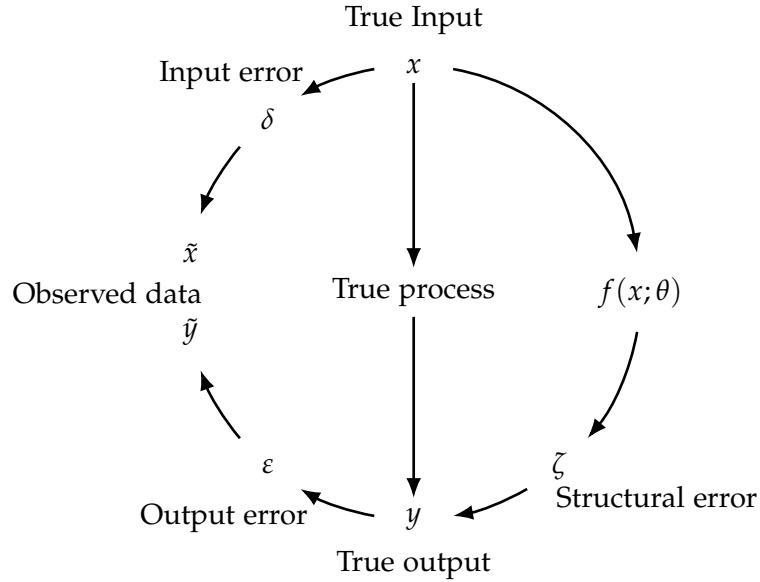


Figure 4.1: Uncertainty framework. Inspiration : [27]

of estimating the parameter θ . Power production from WTG's can be modelled in such a uncertainty framework where input error characterises the relation between nacelle measured wind speed and free wind speed. Structural error characterises processes that affect the power production but are not modelled in $f(x; \theta)$. In the delimitations it was assumed that the produced power is directly observed, so output error is not applied here.

Given a set of observed data points $\mathcal{D} = \{\tilde{x}_n, \tilde{y}_n\}_{n=1}^N$, an objective function to estimate

θ can be formulated via Bayes formula for conditional pdf's [5, p. 22]:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}, \quad (4.1)$$

where $p(\theta|\mathcal{D})$ is the *posterior* distribution, $p(\mathcal{D}|\theta)$ is the *likelihood/observation model* and $p(\theta)$ is the *prior* distribution^[1]. Since the denominator in (4.1) is invariant under changes to θ it is sufficient to consider

$$\underbrace{p(\theta|\mathcal{D})}_{\text{posterior}} \propto \underbrace{p(\mathcal{D}|\theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

for optimization. Obtaining an analytical expression for the posterior can be trivial in some cases, but, in the described uncertainty framework with various sources of error, it is generally more challenging. Furthermore, the models presented later in the chapter have the added complication such that the distributions for δ, ζ and ε feature additional parameters requiring modelling of prior distributions for these. To formalise this, let $p_\delta(\delta|\beta_\delta)$ and $p_\zeta(\zeta|\beta_\zeta)$ be known conditional distributions for the input and structural error where β_δ and β_ζ are additional nuisance parameters^[2] which follow known prior distributions $p(\beta_\delta)$ and $p(\beta_\zeta)$ - output error is omitted here. The process is now described as:

$$\tilde{x} = x + \delta \quad \text{where} \quad \delta|\beta_\delta \sim p_\delta(\delta|\beta_\delta) \quad \text{and} \quad \beta_\delta \sim p(\beta_\delta), \quad (4.2)$$

$$y = f(x; \theta) + \zeta \quad \text{where} \quad \zeta|\beta_\zeta \sim p_\zeta(\zeta|\beta_\zeta) \quad \text{and} \quad \beta_\zeta \sim p(\beta_\zeta), \quad (4.3)$$

Obtaining an analytical expression for the posterior of θ given (\tilde{x}, y) may be difficult since the likelihood is now affected by multiple random variables. A few methods may be used to solve this.

Marginalisation In some cases, it might be possible to integrate out the nuisance parameters through marginalisation. In the case of (4.2) consider first the joint distribution of $x, \beta_\delta | \tilde{x}$. The chain rule of conditional probability^[3] gives that:

$$p(x, \beta_\delta | \tilde{x}) = p(x | \tilde{x}, \beta_\delta) p(\beta_\delta | \tilde{x})$$

Using that $x = \tilde{x} - \delta$ and corollary B.2.1 in the appendix about linear transformations of random variables the pdf for $x | \tilde{x}$ is:

$$p(x | \tilde{x}, \beta_\delta) = p_\delta(\tilde{x} - x | \tilde{x}, \beta_\delta) p(\beta_\delta | \tilde{x}).$$

where p_δ is specifically the pdf for δ . The distribution for $x | \tilde{x}$ is obtained through marginalisation by integrating over the domain of δ - say Ω_δ :

$$p(x | \tilde{x}) = \int_{\Omega_\delta} p(x | \tilde{x}, \beta_\delta) d\beta_\delta = \int_{\Omega_\delta} p_\delta(\tilde{x} - x | \tilde{x}, \beta_\delta) p(\beta_\delta | \tilde{x}) d\beta_\delta \quad (4.4)$$

Integrals like this are studied in the context of *hierarchical models* which offer solutions for some specific distributions. In the context of hierarchical models $p(x | \tilde{x}, \beta_\delta)$ is considered

^[1]Note that the subscripts of the pdf's will be omitted unless needed to ease the notation.

^[2]A nuisance parameter refers to a parameter not of interest.

^[3] $p(a, b) = p(a|b)p(b)$

the *first stage model* and β_δ is a random effect described by the *second stage model* $p(\beta_\delta)$ [41, p. 200]. Examples of two stage hierarchical models where the random effect can be integrated out to obtain analytical solutions include: $X|\mu$ following a Poisson distribution with parameter μ following a Gamma distribution with known parameters [41, p. 227] and $Y|\sigma^2$ following a Gaussian distribution with known mean and $1/\sigma^2$ following a Gamma distribution with known parameters [41, p. 241]. One can try to build models in such a way that marginalisation is possible, but this gives less modelling flexibility so it may not always be desired.

Obtaining the likelihood Different techniques are available to obtain the likelihood. If the distribution for $x|\tilde{x}$ is known the distribution for $f(x;\theta)|\theta, \tilde{x}$ is given by the general change of variable theorem for scalar distributions stated in appendix B.2.1. As an example say that f is just a scalar operation multiplying with θ then $f(x;\theta)|\tilde{x}, \theta$ has pdf:

$$\frac{1}{|\theta|} p_{x|\tilde{x}}\left(\frac{x}{\theta}|\tilde{x}, \theta\right)$$

Obtaining the likelihood then requires combining the distribution of $f(x;\theta)|\tilde{x}, \theta$ with the distribution of the structural error. If the structural error is additive using *moment-generating function* might yield an analytical expression for the likelihood by exploiting that the moment-generating function of two independent random variables added together is a product of the moment-generating functions for each of the random variables [46, p. 239]^[4]. The moment generating functions for each random variable can be found by table lookup for known distributions, but obtaining the pdf for the combined distribution by the inverse transformation may be more challenging.

The methods mentioned above are by no means an exhaustive list of ways to obtain the likelihood, but it gives some insight and it highlights that finding the likelihood is generally a non-trivial task, in particular if distributions for the various uncertainties are chosen based on physical considerations rather than choosing distributions that enables analytical solutions to be found. To help this issue [34] suggests an alternative approach which is to obtain an analytical expression for the posterior by including the latent parameters only requiring that the various conditional distributions are known analytically. For the case of the setup in (4.2) and (4.3), [34] then suggests to include δ, β_δ and β_ζ in the posterior so that the posterior based on a single observation (\tilde{x}, y) is:

$$p(\theta, \delta, \beta_\delta, \beta_\zeta | \tilde{x}, y) \propto p(y | \tilde{x}, \theta, \delta, \beta_\delta, \beta_\zeta) p(\theta, \delta, \beta_\delta, \beta_\zeta | \tilde{x})$$

$y = f(x;\theta) + \zeta$ so using corollary B.2.1 again gives the pdf for the likelihood:

$$p(y | \tilde{x}, \theta, \delta, \beta_\delta, \beta_\zeta) = p_\zeta(y - \underbrace{f(\tilde{x} - \delta; \theta)}_x | \tilde{x}, \theta, \delta, \beta_\zeta),$$

The parameters can be assumed mutually independent other than the dependence of δ and β_δ . If the parameters are also assumed independent on the observation of \tilde{x} , which may or may not be a good assumption, the prior is:

$$p(\theta, \delta, \beta_\delta, \beta_\zeta | \tilde{x}) = p(\theta) p_\delta(\delta | \beta_\delta) p(\beta_\delta) p(\beta_\zeta)$$

^[4]The moment-generating function is defined as the expectation $E[\exp(tX)]$ for a continuous random variable X and real number t .

The posterior distribution then reduces to:

$$p(\theta, \delta, \beta_\delta, \beta_\zeta | \tilde{x}, y) \propto p_\zeta(y - f(\tilde{x} - \delta; \theta) | \tilde{x}, \theta, \delta, \beta_\zeta) p(\beta_\zeta) p_\delta(\delta | \beta_\delta) p(\beta_\delta) p(\theta) \quad (4.5)$$

The posterior in (4.5) can be evaluated directly with the distributions for the input and structural error as well $p(\theta)$ chosen appropriately. There is no free lunch though, since the posterior in (4.5) now includes the input error δ . For one thing this is a problem if the spread of $p_\delta(\delta | \beta_\delta)$ and/or $p_\zeta(\delta | \beta_\zeta)$ is too large since each one observation is affected by two sources of error from δ and ζ . If prior information about x is known, which is the case for wind speed as described in appendix C.1, then including x in the posterior instead of δ gives the parameter prior:

$$p(\theta, x, \beta_\delta, \beta_\zeta | \tilde{x}) \propto p(\theta) p_\delta(\tilde{x} - x | \beta_\delta, \tilde{x}) p(\beta_\delta) p(x) p(\beta_\zeta),$$

omitting a few details. The additional prior information about x might help parameter estimation. Another issue is the dimensionality of the parameters to be estimated. To see this, consider a set of N observed inputs $\tilde{\mathbf{x}} \in \mathbb{R}^N$ and outputs $\mathbf{y} \in \mathbb{R}^N$. The parameters θ , β_δ and β_ζ are assumed the same for each observation but δ and ζ changes so let $\boldsymbol{\delta} \in \mathbb{R}^N$ and $\boldsymbol{\zeta} \in \mathbb{R}^N$ be vectors of dimension N containing the individual input and structural errors. If the errors are mutually independent the posterior is:

$$p(\theta, \boldsymbol{\delta}, \beta_\delta, \beta_\zeta | \tilde{\mathbf{x}}, \mathbf{y}) \propto \prod_{n=1}^N \left(p_\zeta(y_n - f(\underbrace{\tilde{x}_n - \delta_n}_{x_n}; \theta) | \tilde{x}_n, \delta_n, \theta, \beta_\zeta) p_\delta(\delta_n | \beta_\delta) \right) p(\beta_\zeta) p(\beta_\delta) p(\theta), \quad (4.6)$$

The posterior in (4.6) is still not affected by the realisations of ζ , but due to the dependence on the realisations on δ , the dimension of the parameters in the posterior now scales linearly with the number of observations N . Any optimization algorithm would have to estimate every realisation of δ in order to estimate the remaining parameters. This issue is also discussed in [34], but it is found that if the input is measured “fairly accurately” [34, p. 3] then parameter estimation is possible using *Markov chain Monte Carlo* (MCMC) methods when the parameters are initialised carefully [34, p. 7]. These methods was attempted for this thesis, but the issue with dimensionality proved to great and convergence was not achieved using one MCMC method know as the Metropolis Algorithm. For the AW2P model, introduced in section 3.3, which is the modelling of power over a 10 minute period, the issue of dimension is aggravated even further since each observation is based on 600 individual realisations of the input error over 10 minutes in such a way that each observations adds 600 latent variables to be estimated. Thus using the method posed by [34] without alteration is deemed infeasible here.

4.1.1 Approximate Bayesian Total error analysis

To overcome the outlined issues with BETAE this section proposes A-BETAE as an alternative method/modelling philosophy. The basic idea is summarized in a few points:

- Design a model according to some physical process in an uncertainty framework featuring input, structural and output error (or some of these) without making compromises in regards to obtaining analytical solutions of the different distributions.

- Perform marginalisation by integrating out nuisance variables where necessary in order to get a low dimensional posterior distribution. If analytical solutions are not available, use numerical methods such as numerical integration or Monte Carlo simulation.
- Use approximate distributions to get a parametrisation for the likelihood if the exact distribution cannot be obtained.

For the uncertainty framework given in (4.2) and (4.3), A-BETAE suggests to perform marginalisation over β_δ in order to get a distribution for $x|\tilde{x}$ as shown in (4.11) either by solving the integral or using numerical methods. Then the distribution $f(x;\theta)|\tilde{x},\theta$ is found either by using the change of variable method or by approximating the distribution. An example of an approximation would be estimating the mean and variance then assuming a Gaussian distribution. Finally, the distribution of

$$f(x;\theta) + \zeta|\tilde{x},\theta,\beta_\zeta$$

is obtained either directly or by approximating the distribution. If both $f(x;\theta)$ and ζ are Gaussian distributed and independent their sum would also be Gaussian with their means and variances added.

The formulation of A-BETAE is somewhat vague since the implementation is specific to the modelled case. The remaining chapter will give one example where section 4.3 uses marginalisation to get a distribution for the free wind speed and section 4.4 uses an approximate distribution for produced power by estimating the mean and variance of a Gaussian distribution. Using numerical methods and approximate distributions might introduce some undesired effects like bias in the parameter estimation - see chapter 7 for further discussion.

4.2 Non-informative Bayesian inference priors

Bayes' theorem given in the previous section states that:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Here the prior is used to infer known information into the model also known as *á priori* information. The impact of the prior on the posterior tends to depend on the sample size N . When N is small the prior might have a great impact, while for larger N the likelihood, which is often a product of N terms, will dominate and the effects of the prior vanishes for $N \rightarrow \infty$. The sample sizes for this thesis are in the order of 10^5 or even 10^6 so the prior distribution might not have a great impact on the posterior but it should nonetheless be considered. A few different prior distributions was modelled in chapter 3 based on information about of the physical phenomena at play, but when little information is known, *non-informative* priors can be used.

When designing a non-informative prior, one might suggest the following conditions:

- The prior have little to no impact on the posterior distribution.
- The prior do not assume *á priori* information and lets the data speak for itself.

Based on these points an obvious suggestion would be a uniform prior distribution however if the bounds for the input is not known or if the input is unbounded, then it might be necessary make some assumptions. Another issue with the uniform distribution is the fact that the choice of parametrisation - say for a parameter θ - impacts inference. To see this, consider an example where θ follows a uniform distribution between $a, b \in \mathbb{R}$ such that $\theta \sim \mathcal{U}(a, b)$. Had one instead chosen the parametrisation $\psi = g(\theta)$ such that $\psi \sim \mathcal{U}(g(a), g(b))$ then:

$$P(a \leq \theta \leq b) = \int_a^b p_\theta(\theta) d\theta \neq \int_{g(a)}^{g(b)} p_\psi(\psi) d\psi = P(g(a) \leq \psi \leq g(b)), \quad (4.7)$$

except for some trivial cases. If equality is achieved in (4.7) for any choice of g it is then said that the prior is *invariant under reparametrisation* [46, p. 361]. A family of priors known as Jeffreys' prior achieves this property by defining the prior based on the likelihood.

Definition 4.2.1 (Jeffreys' prior) [46, p. 361] Let $p_Y(\mathbf{y}|\boldsymbol{\theta})$ be a likelihood pdf for the random variable \mathbf{Y} parametrised by $\boldsymbol{\theta}$. The prior distribution for $\boldsymbol{\theta}$ follows Jeffreys' prior if:

$$p_\theta(\boldsymbol{\theta}) \propto \sqrt{\det(\mathbf{i}(\boldsymbol{\theta}))},$$

where $\mathbf{i}(\boldsymbol{\theta})$ is the expected information corresponding to \mathbf{y} and $\boldsymbol{\theta}$ - see appendix B.2.1.

The expected information can be interpreted as how much information about $\boldsymbol{\theta}$ an observation of the random variable gives and having the prior distribution depend on this in Jeffreys' prior might be good property. Proposition B.2 in the appendix gives that Jeffrey's prior is invariant under reparametrisation (up to proportionality) if the Jeffrey's calculated calculated with respect to $\psi = g(\theta)$ is the same the same as reparametrising Jeffrey's prior for θ then using the change of variable proposition in appendix B.2.1 to get p_ψ . In other words the prior is invariant if it does not matter if p_ψ is obtained directly from Jeffrey's prior or if Jeffrey's prior for p_θ is first found then transformed using the change of variable proposition. Invariance is proven by the proposition below for a scalar variable.

Proposition 4.2.1 Let $\theta \in \mathbb{R}$ be a random variable with pdf $p_\theta(\theta) \propto \sqrt{\det(i_\theta(\theta))}$, $p_\psi(\psi) \propto \sqrt{\det(i_\psi(\psi))}$, and likelihood $p_\theta(y; \theta) = p_\psi(y; g(\theta))$ such that $\psi = g(\theta)$. Then

$$p_\theta(\theta) \propto \left| \frac{d}{d\theta} g(\theta) \right| p_\psi(g(\theta)). \quad (4.8)$$

Proof The proof is inspired by [56]. For scalar expected information $\det(i_\theta(\theta)) = i_\theta(\theta)$ and it follows that:

$$p_\theta(\theta) \propto \sqrt{i_\theta(\theta)} \propto \left| \frac{d}{d\theta} g(\theta) \right| p_\psi(g(\theta)) \Leftrightarrow i_\theta(\theta) \propto \left(\frac{d}{d\theta} g(\theta) \right)^2 p_\psi^2(g(\theta))$$

Using lemma B.2.1 and the chain rule for derivatives:

$$\begin{aligned}
 i_{\theta}(\theta) &= - \int \left(\frac{d^2}{d^2\theta} \ell_{\theta}(y; \theta) \right) p_{\theta}(y; \theta) dy \\
 &= \int \left(\frac{d}{d\theta} \ell_{\theta}(y; \theta) \right)^2 p_{\theta}(y; \theta) dy \quad (\text{Lemma B.2.1}) \\
 &= \int \left(\frac{d}{d\theta} \ell_{\psi}(y; g(\theta)) \right)^2 p_{\psi}(y; g(\theta)) dy \\
 &= \int \left(\frac{d}{d\psi} \ell_{\psi}(y; \psi) \frac{d}{d\theta} g(\theta) \right)^2 p_{\psi}(y; \psi) dy \quad (\text{Chain rule}) \\
 &= \left(\frac{d}{d\theta} g(\theta) \right)^2 \underbrace{\int \left(\frac{d}{d\psi} \ell_{\psi}(y; \psi) \right)^2 p_{\psi}(y; \psi) dy}_{i_{\psi}(\psi)} \\
 &\propto \left(\frac{d}{d\theta} g(\theta) \right)^2 p_{\psi}^2(\psi) = \left(\frac{d}{d\theta} g(\theta) \right)^2 p_{\psi}^2(g(\theta)) \quad \blacksquare
 \end{aligned}$$

Jeffrey's prior for Gaussian distributed random variables is given by the proposition below.

Proposition 4.2.2 Let $Y \sim \mathcal{N}(\mu, \sigma^2)$ with parameters $\theta = (\mu, \sigma)$. If θ follows Jeffreys' prior then:

$$p_{\theta}(\theta) \propto \frac{1}{\sigma^2}$$

Proof Let $Y \sim \mathcal{N}(\mu, \sigma^2)$ and $\theta = (\mu, \sigma)$, the log-likelihood is given by

$$\log p(\theta|y) = -\log \sigma - \frac{(y - \mu)^2}{2\sigma^2}$$

By applying the definition of the expected information the following is obtained

$$\begin{aligned}
 \mathbf{i}(\theta) &= -\mathbb{E} \left[\frac{\partial^2}{\partial \theta \partial \theta^T} \ell(\theta|y) \right] \\
 &= -\mathbb{E} \left[\begin{pmatrix} \frac{\partial^2}{\partial \mu^2} \ell(\theta|y) & \frac{\partial^2}{\partial \mu \partial \sigma} \ell(\theta|y) \\ \frac{\partial^2}{\partial \sigma \partial \mu} \ell(\theta|y) & \frac{\partial^2}{\partial \sigma^2} \ell(\theta|y) \end{pmatrix} \right] \\
 &= -\mathbb{E} \left[\begin{pmatrix} -\frac{1}{\sigma^2} & -\frac{2(y - \mu)}{\sigma^3} \\ -\frac{2(y - \mu)}{\sigma^3} & \frac{1}{\sigma^2} - \frac{3(y - \mu)^2}{\sigma^4} \end{pmatrix} \right]
 \end{aligned}$$

By applying the expected value it is observed that $\mathbb{E}[(y - \mu)] = 0$, $\mathbb{E}[(y - \mu)^2] = \sigma^2$ and the expression becomes

$$i(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix} \quad (4.9)$$

the Jefferys prior becomes

$$\sqrt{\det(\mathbf{i}(\boldsymbol{\theta}))} = \sqrt{\frac{1}{\sigma^2} \cdot \frac{2}{\sigma^2}} \propto \frac{1}{\sigma^2} \quad (4.10) \quad \blacksquare$$

Remark 4.1 Jeffrey's prior for Gaussian random variable is improper in the sense that:

$$\int_0^\infty \sigma^{-2} d\sigma = \lim_{a \rightarrow 0^+} \int_a^1 \sigma^{-2} d\sigma + \lim_{b \rightarrow \infty} \int_1^b \sigma^{-2} d\sigma = \lim_{a \rightarrow 0^+} \left(\frac{1}{a} - 1 \right) + \lim_{b \rightarrow \infty} \left(1 - \frac{1}{b} \right) = \infty$$

In general, this not necessarily a problem as long the posterior pdf with input y given by:

$$p(\boldsymbol{\theta}|y) = \frac{p(y|\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{p(y)} = \frac{p(y|\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\int p(y|\boldsymbol{\theta}')p_{\boldsymbol{\theta}}(\boldsymbol{\theta}')d\boldsymbol{\theta}'}$$

is a possible pdf which happens when the denominator is finite [56]. The proof is omitted.

In the proposition it is observed that $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is constant with respect to μ , thus when the variance σ^2 is fixed a flat prior is obtained.

4.3 Obtaining marginal wind speed distributions

In order to estimate the first and second order moment of the aggregated power \bar{P} , a few marginal distributions of the wind speed are needed. Similar to previous setups, let $\mathbf{v} = [v_1 \dots v_N]^T \in \mathbb{R}^N$ be the vector of wind speeds over a 10 minute period and let \bar{v} be the measured average over that period. $v_t \in \mathbf{v}$ is used as shorthand for the free wind speed $v_{f,t}$ omitting the subscript f unless the distinction between free and nacelle wind speed is needed. The conditional pdf $p_{v_t}(v_t|\bar{v})$ and the joint conditional pdf $p_{v_t, v_{t'}}(v_t, v_{t'}|\bar{v})$ for time instances $t \neq t'$ are desired. Two distributions are modelled: One where \bar{v} is considered the true mean, i.e $\bar{v} = E[v_t]$ and one where uncertainty with respect to \bar{v} is considered through the nacelle transfer function.

4.3.1 Assuming known mean free wind speed

Assuming that \bar{v} is the true mean of the free wind speed gives the distribution introduced in (3.4) which was:

$$\mathbf{v}|\bar{v}, \sigma_v \sim \mathcal{N}(\bar{v}\mathbf{1}_N, \sigma_v^2\mathbf{I}_{N \times N}) \quad \text{and} \quad \sigma_v|\bar{v} \sim \mathcal{N}_+(\mu_{\sigma_v}, \sigma_{\sigma_v}^2)$$

To obtain the marginal distribution of $v_t|\bar{v}$ requires integration over σ_v as well as all $v_{t'}$ for $t' \neq t$. First the joint distribution of $\mathbf{v}|\bar{v}, \sigma_v$ is factored as:

$$p_v(\mathbf{v}|\bar{v}, \sigma_v) = \prod_{t=1}^N p_{v_t}(v_t|\bar{v}, \sigma_v^2),$$

due to the i.i.d property where $p_{v_t}(v_t|\bar{v}, \sigma_v^2)$ is the pdf of a Gaussian distribution with mean \bar{v} and variance σ_v^2 . Now using the chain rule of conditional probability:

$$p_{v_t}(v_t, \sigma_v|\bar{v}) = p_{v_t}(v_t|\bar{v}, \sigma_v)p_{\sigma_v}(\sigma_v|\bar{v}). \quad (4.11)$$

The distribution of $v_t | \bar{v}$ is obtained by marginalisation:

$$p_{v_t, \sigma_v}(v_t | \bar{v}) = \int_0^\infty p_{v_t}(v_t | \bar{v}, \sigma_v) p_{\sigma_v}(\sigma_v | \bar{v}) d\sigma_v, \quad (4.12)$$

where p_{σ_v} is the pdf of a folded normal distribution with parameters according to (3.1) and (3.2). The integral in (4.12) can be computed numerically^[5]. For this an adaptive numerical integration method will be applied - see [49] for the used software.

The joint distribution $p_{v_t, v_{t'}}(v_t, v_{t'} | \bar{v})$ is obtained similarly:

$$p_{v_t, v_{t'}}(v_t, v_{t'} | \bar{v}) = \int_0^\infty p_{v_t, v_{t'}}(v_t, v_{t'} | \bar{v}, \sigma_v) p_{\sigma_v}(\sigma_v | \bar{v}) d\sigma_v, \quad (4.13)$$

where $p_{v_t, v_{t'}}(v_t, v_{t'} | \bar{v}, \sigma_v)$ is the pdf of a bivariate uncorrelated Gaussian distribution which has the feature that it only depends on v_t and $v_{t'}$ through their total Euclidean distance to the mean \bar{v} :

$$\begin{aligned} p_{v_t, v_{t'}}(v_t, v_{t'} | \bar{v}, \sigma_v) &= p_{v_t}(v_t | \bar{v}, \sigma_v) p_{v_{t'}}(v_{t'} | \bar{v}, \sigma_v) \\ &= \frac{1}{2\pi\sigma_v^2} \exp\left(-\frac{1}{2\sigma_v^2} \left((v_t - \bar{v})^2 + (v_{t'} - \bar{v})^2\right)\right) \\ &= \frac{1}{2\pi\sigma_v^2} \exp\left(-\frac{1}{2\sigma_v^2} \underbrace{\left\| \begin{bmatrix} v_t - \bar{v} \\ v_{t'} - \bar{v} \end{bmatrix} \right\|^2}_{d^2}\right), \end{aligned}$$

where $\|\cdot\|$ denotes the ℓ_2 norm - see appendix B.2.2. Thus (4.13) is fully characterised only with respect to the one dimensional variable d which is also equivalent to tabulating (4.12) with $(v_t - \bar{v})^2 = d^2$ and scaling the result by $(\sqrt{2\pi}\sigma_v)^{-1}$. See section 4.6 for implementation details.

4.3.2 Free wind speed measurements through nacelle transfer function

Characterising the marginal distributions as having input uncertainty with respect to the mean is more involved since the relation between nacelle measured wind speed v_{nc} and free wind speed v_f should be considered through the nacelle transfer function. Denoting $\mu_v = E[v_f]$ as the mean free wind speed over 10 minutes, the assumptions according to section 3.1 in regards to the free wind speed are then:

$$\mathbf{v}_f | \mu_v, \sigma_v \sim \mathcal{N}(\mu_v \mathbf{1}_N, \sigma_v^2 \mathbf{I}_{N \times N}), \quad \sigma_v | \bar{v} \sim \mathcal{N}_+(\mu_{\sigma_v}, \sigma_{\sigma_v}^2),$$

The measured average wind speed is:

$$\bar{v} = \frac{1}{N} \sum_{t=1}^N v_{nc,t},$$

^[5]To obtain an analytical solution requires evaluating an integral of the form $\int x^{-1} \exp(ax^{-2} + bx^2 + cx + d) dx$ for real numbers a, b, c, d which, to the extent of the knowledge of the authors, has no closed form solution.

where the free and nacelle wind speeds are linked through the NTF:

$$v_{f,t} = \text{NTF}(v_{nc,t}; \boldsymbol{\eta}) + \xi_t = \eta_1 v_{nc,t} + \eta_2 + \xi_t, \quad (4.14)$$

where $\boldsymbol{\eta}$ are NTF coefficients following a bivariate Gaussian prior distribution:

$$\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta),$$

and ξ_t is additive noise drawn independently from a Gaussian distribution:

$$\xi_t | \sigma_\xi \sim \mathcal{N}(0, \sigma_\xi^2) \quad \text{with} \quad \sigma_\xi \sim \mathcal{N}_+(\mu_{\sigma_\xi}, \sigma_{\sigma_\xi}^2).$$

For a given 10 minute period, the parameters σ_v , σ_ξ and $\boldsymbol{\eta}$ are assumed constant but each is drawn independently between periods^[6]. Note that values for the prior distribution parameters $\boldsymbol{\mu}_\eta$, $\boldsymbol{\Sigma}_\eta$, μ_{σ_ξ} and σ_{σ_ξ} were not available in the lookup tables so the values suggested in section 3.1.2 based on the literature are used in the implementation. To obtain $p_{v_t}(v_t | \bar{v})$ and $p_{v_t, v_{f,t}}(v_t, v_{f,t} | \bar{v})$, marginalisation with respect to σ_v , $\boldsymbol{\eta}$, σ_ξ is now required. Note that by marginalising out $\boldsymbol{\eta}$ this parameter cannot be estimated based on the marginal distributions so if $\boldsymbol{\eta}$ is desired another method is needed.

Obtaining the marginal distributions could be done similarly to the previous sections through numerical integration but the increased dimension offers a challenge. Computing e.g. $p_{v_t}(v_t | \bar{v})$ requires marginalisation over σ_v , $\boldsymbol{\eta}$ and σ_ξ but the parameters are coupled so the multidimensional integral in the marginalisation do not factor into separate integrals. In [41, p. 200], parameters like this are referred to as *crossed random effects*. The computational complexity of numerically integrating over crossed random effects is exponential with respect to the number of crossed random effects [41, p. 199], so instead the marginal distributions are obtained through Monte Carlo Simulations.

To estimate the distributions using Monte Carlo simulation first requires realisations of the form $\{\bar{v}_n, \mathbf{v}_{f,n}\}_{n=1}^{N_{\text{sim}}}$ according to the assumptions above where each \bar{v}_n is the average observed wind speed over a period with free wind speeds $\mathbf{v}_{f,n}$. Algorithm 1 describes how to obtain realisations given the mean wind speed μ_v and the algorithm can be repeated multiple times for different μ_v in a suitable range to get a set of representative realisations - see further details in section 4.6. Given the realisations, the marginal distributions can be estimated using a *Gaussian kernel density estimator*. In general if $\{\mathbf{x}_n\}_{n=1}^{N_{\text{sim}}}$ are realisations of a random process \mathbf{X} of dimension D , the isotropic Gaussian kernel density estimate is [5, p. 123]:

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{N_{\text{sim}}} \sum_{n=1}^{N_{\text{sim}}} \frac{1}{\sqrt{2\pi h^2}^D} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{2h^2}\right),$$

where h is a *smoothing parameter/bandwidth* chosen according to Scott's rule described in appendix B. To obtain $p_{v_t}(v_t | \bar{v}_0)$ for a particular \bar{v}_0 , choose the subset of realisations that have observed average wind speed \bar{v}_0 and denote them $\{\bar{v}_{n,0}, \mathbf{v}_{f,n,0}\}_{n=1}^{N_{\text{sim},0}}$. The distribution

^[6]The assumption that $\boldsymbol{\eta}$ is drawn independently each 10 minute period is made to ensure temporal independence of power distribution presented in section 4.4 although some dependence is perhaps more realistic from a modelling perspective.

is then estimated using the Gaussian kernel using each $\mathbf{v}_{f,n,0}$ as N_{sim} separate observations of dimension $D = 1$. To obtain the joint marginal distribution $p_{v_t, v_{t'}}(v_t, v_{t'} | \bar{v})$ divide each $\mathbf{v}_{f,n,0}$ into two parts of size $N/2$ and use those as separate observations in a kernel estimator with $D = 2$.

4.4 Power distribution

As described in section 4.1.1, A-BETAE seeks to obtain an approximate distribution of the modelled quantity using marginalised distributions describing the input uncertainty. Here the modelled quantity is the total power over a 10-minute period \bar{P} and the input uncertainty is characterised by the marginal distributions for $v_t | \bar{v}$ and $v_t, v_{t'} | \bar{v}$ with or without assuming known mean free wind speed according to section 4.3.1 and 4.3.2 respectively. In either case, the distribution of \bar{P} is given by the wind speed distributions and the W2P model from (3.7). \bar{P} was modelled by the sum of P_t generated over periods of Δt with Gaussian additive noise:

$$\bar{P} = \text{AW2P}(\mathbf{v}; \boldsymbol{\theta}) + \bar{\zeta} = \sum_{t=1}^N \Delta t \text{W2P}(v_t; \boldsymbol{\theta}) + \sum_{t=1}^N \zeta_t = \sum_{t=1}^N P_t + \sum_{t=1}^N \zeta_t \quad (4.15)$$

Let $\hat{\bar{P}} = \text{AW2P}(\mathbf{v}; \boldsymbol{\theta})$ and $\hat{P}_t = \Delta t \text{W2P}(v_t; \boldsymbol{\theta})$ denote the modelled power via the AW2P and W2P model respectively. The Central limit theorem states that the sum of any identically distributed independent random variables converges in distribution to a Gaussian distribution [46, p. 268]. In the case of $\hat{\bar{P}}$, the requirement of identically distributed variables is fulfilled since each \hat{P}_t has the same distribution, but the requirement of being independent is not. In the setup where \bar{v} is assumed to be the true mean of the free wind speed, $v_t | \bar{v}, \sigma_v$ is independent of $v_{t'} | \bar{v}, \sigma_v$ for $t \neq t'$ but the distribution of $v_t | \bar{v}$ is dependent on $v_{t'} | \bar{v}$ since they share the same standard deviation - thus factoring of the joint pdf $p_{v_t, v_{t'}}(v_t, v_{t'} | \bar{v})$ is not possible^[7]. Accounting for uncertainty with respect to \bar{v} further complicates the dependence since now each realisation also shares NTF coefficient $\boldsymbol{\eta}$ as well as the NTF structural error standard deviation $\sigma_{\bar{\zeta}}$.

To give some context, distributions like these are seen in *hierarchical models* and in some cases the dependence between the variables can be made explicit. As an example if $X_i = U + \varepsilon_i$ for $i = 1, \dots, N$ where U is zero mean Gaussian distributed with variance σ_u^2 independent of ε_i which is also zero mean Gaussian with variance σ_ε^2 and each ε_i are mutually independent, then the covariance is given by [41, p. 163]:

$$\text{Cov}[X_i, X_j] = \begin{cases} \sigma_u^2 + \sigma_\varepsilon^2 & \text{for } i = j \\ \sigma_u^2 & \text{for } i \neq j \end{cases} \quad (4.16)$$

This example is somewhat similar to the distribution of $\mathbf{v} | \bar{v}$ since each $v_t \in \mathbf{v}$ share the same realisation of standard deviation and possibly also $\boldsymbol{\eta}$ and $\sigma_{\bar{\zeta}}$, but are otherwise independent. In summary, $\hat{\bar{P}}$ is the sum of dependent identically distributed random variables. An analytical expression of the distribution - or even the first and second order moment - is not trivially obtained since the W2P model is highly non-linear and since the marginal distributions for the wind speeds are also not available as analytical expressions. To overcome

^[7]Technically this has not been shown.

this issue the principle of A-BETAE is used by using an approximate distribution - here a Gaussian distribution for \hat{P} is used. This is a somewhat crude approximation and should be analysed further, but it gives an efficient way to characterise the distribution using its moments and at least the central limit theorem gives an asymptotic Gaussian distribution if the wind speeds were independent. See further discussion of this in chapter 7. Modelled power is therefore approximated using:

$$\hat{P} | \bar{v}, \theta \sim \mathcal{N}(\mu_{\hat{P}}(\bar{v}; \theta), \sigma_{\hat{P}}^2(\bar{v}; \theta)), \quad (4.17)$$

such that the mean and variance are functions of observed average wind speed \bar{v} and parametrisation of the W2P model θ . Furthermore, in section 3.1 it was assumed that wind speeds for different 10-minute periods given statistics are independent - i.e $\mathbf{v}_n | \bar{v}_n, \sigma_{v,n}, \boldsymbol{\eta}_n, \sigma_{\xi,n}$ is independent of $\mathbf{v}_{n'} | \bar{v}_{n'}, \sigma_{v,n'}, \boldsymbol{\eta}_{n'}, \sigma_{\xi,n'}$ when modelling the effect of the NTF. In section 4.3.2 it was assumed that $\sigma_v, \boldsymbol{\eta}$ and σ_{ξ} are drawn independently for different 10 minute periods giving that $\mathbf{v}_n | \bar{v}_n$ is independent of $\mathbf{v}_{n'} | \bar{v}_{n'}$ after marginalisation over the statistics. Since the modelled power \hat{P} is just a function of \mathbf{v} it then follows that the power between two time periods \hat{P}_n and $\hat{P}_{n'}$ are conditionally independent for $n \neq n'$ such that^[8]:

$$\begin{aligned} p_{\hat{P}_n, \hat{P}_{n'}}(\hat{P}_n, \hat{P}_{n'} | \bar{v}_n, \bar{v}_{n'}, \theta) &= p_{\hat{P}_n}(\hat{P}_n | \bar{v}_n, \bar{v}_{n'}, \theta) p_{\hat{P}_{n'}}(\hat{P}_{n'} | \bar{v}_n, \bar{v}_{n'}, \theta) \\ &= p_{\hat{P}_n}(\hat{P}_n | \bar{v}_n, \theta) p_{\hat{P}_{n'}}(\hat{P}_{n'} | \bar{v}_{n'}, \theta), \end{aligned} \quad (4.18)$$

where \bar{v}_n and $\bar{v}_{n'}$ are the average wind speeds for the respective periods.

Since \hat{P} is as sum over \hat{P}_t , the statistics can be obtained by characterising \hat{P}_t . Specifically the mean and variance is:

$$\begin{aligned} \mathbb{E}[\hat{P} | \bar{v}, \theta] &= \sum_{t=1}^N \mathbb{E}[\hat{P}_t | \bar{v}, \theta] = N \mathbb{E}[\hat{P}_1 | \bar{v}, \theta] \\ \text{Var}[\hat{P} | \bar{v}, \theta] &= \sum_{t=1}^N \text{Var}[\hat{P}_t | \bar{v}, \theta] + \sum_{t=1}^N \sum_{\substack{t'=1 \\ t' \neq t}}^N \text{Cov}[\hat{P}_t, \hat{P}_{t'} | \bar{v}, \theta] \\ &= N \text{Var}[\hat{P}_1 | \bar{v}, \theta] + N(N-1) \text{Cov}[\hat{P}_1, \hat{P}_2 | \bar{v}, \theta], \end{aligned}$$

using that \hat{P}_t is identically distributed for each $t = 1, \dots, N$ - thus characterising mean and variance for \hat{P}_1 and covariance between \hat{P}_1 and \hat{P}_2 is sufficient. In other words the covariance structure is the same in the off diagonal elements, similar to the example in (4.16). Now denote $p_{v_t}(v_{t,m} | \bar{v})$ and $p_{v_t, v_{t'}}(v_{t,m}, v_{t,m'} | \bar{v})$ for $m = 1, \dots, M$ as the distributions sampled equidistantly with distance Δv at M and M^2 discrete points respectively in an appropriate range using the methods in the previous section. The moments of $\hat{P}_t = \Delta t \text{W2P}(v_t; \theta)$

^[8]Generally, if two random variables (X, Y) are independent then it follows directly from the multivariate change of variable theorem that $(U, V) = (f(X), f(Y))$ are also independent for some function f [46, p. 184].

can then be approximated using the midpoint-integration rule:

$$E[\hat{P}_t | \bar{v}, \theta] = \int_{-\infty}^{\infty} \Delta t \text{W2P}(v_t; \theta) p_{v_t}(v_t | \bar{v}) dv_t \approx \sum_{m=1}^M \Delta t \text{W2P}(v_{t,m}; \theta) p_{v_t}(v_{t,m} | \bar{v}) \Delta v \quad (4.19)$$

$$E[\hat{P}_t^2 | \bar{v}, \theta] = \int_{-\infty}^{\infty} (\Delta t)^2 \text{W2P}^2(v_t; \theta) p_{v_t}(v_t | \bar{v}) dv_t \approx \sum_{m=1}^M (\Delta t)^2 \text{W2P}^2(v_{t,m}; \theta) p_{v_t}(v_{t,m} | \bar{v}) \Delta v \quad (4.20)$$

$$\begin{aligned} E[\hat{P}_t \hat{P}_{t'} | \bar{v}, \theta] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\Delta t)^2 \text{W2P}(v_t; \theta) \text{W2P}(v_{t'}; \theta) p_{v_t, v_{t'}}(v_t, v_{t'} | \bar{v}) dv_t dv_{t'} \\ &\approx \sum_{m=1}^M \sum_{m'=1}^M (\Delta t)^2 \text{W2P}(v_{t,m}; \theta) \text{W2P}(v_{t',m'}; \theta) p_{v_{t,m}, v_{t',m'}}(v_{t,m}, v_{t',m'} | \bar{v}) (\Delta v)^2 \end{aligned} \quad (4.21)$$

From these the variance and covariance estimates are obtained through:

$$\begin{aligned} \text{Var}[\hat{P}_t | \bar{v}, \theta] &= E[\hat{P}_t^2 | \bar{v}, \theta] - E[\hat{P}_t | \bar{v}, \theta]^2 \\ \text{Cov}[\hat{P}_t, \hat{P}_{t'} | \bar{v}, \theta] &= E[\hat{P}_t \hat{P}_{t'} | \bar{v}, \theta] - E[\hat{P}_t | \bar{v}, \theta] E[\hat{P}_{t'} | \bar{v}, \theta] = E[\hat{P}_t \hat{P}_{t'} | \bar{v}, \theta] - E[\hat{P}_t | \bar{v}, \theta]^2 \end{aligned}$$

The estimates for mean and variance will be used for parameter inference, but given the parameters the mean estimate $\mu_{\hat{P}}(\bar{v}; \theta)$ is also used to model the power output given inputs \bar{v} and θ . This model will depend on the distributions for the wind speed so let the *probabilistic AW2P* model refer to the model using the wind speed distributions assuming known mean free wind speed given in section 4.3.1 and let the *probabilistic NTF-AW2P* model refer to the model using the distributions in section 4.3.2 where the free wind speed is modelled via the NTF. As shorthand these models will just be referred as the AW2P and NTF-AW2P model for the remainder of the thesis (ignoring the previous definition of the AW2P model). For a given \bar{v} and parameter θ the AW2P model is then:

$$\text{AW2P}(\bar{v}; \theta) = \mu_{\hat{P}}(\bar{v}; \theta) = E[\hat{P} | \bar{v}, \theta] \quad (4.22)$$

where the wind speed distributions are given as in section 4.3.1. The NTF-AW2P model is evaluated similarly but using the wind speeds distributions given in section 4.3.2.

A few notes in regards to computation. For parameter inference it is important that the likelihood distribution $p_{\hat{P}}(\hat{P} | \bar{v}, \theta)$ can be evaluated readily under different values of θ . Since the pdf's $p_{v_t}(v_{t,m} | \bar{v})$ and $p_{v_t, v_{t'}}(v_{t,m}, v_{t',m'} | \bar{v})$ are invariant under changes to θ , these can be tabulated prior to inference. The main computational effort is then estimating the moments in (4.19)-(4.21) in particular the double sum in (4.21).

4.5 Posteriors and logarithm posteriors

As mentioned in section 4.1 parameter inference for a Bayesian setup uses the posterior distribution for the parameters given observed data. The parameters to be estimated are $\theta = (C_{\max}, a_1, a_2)$ from the W2P model and the standard deviation for the structural error $\sigma_{\bar{\zeta}}$. The posterior is constructed based on a training data set including observed power

$\bar{\mathbf{P}} \in \mathbb{R}^{N_{\text{dat}}}$ and average wind speed $\bar{\mathbf{v}} \in \mathbb{R}^{N_{\text{dat}}}$ over a total of N_{dat} 10-minute periods from one or more WTG's. Similarly to section 4.1 the posterior distribution is obtained using Bayes formula:^[9]

$$p(\boldsymbol{\theta}, \sigma_{\bar{\zeta}} | \bar{\mathbf{P}}, \bar{\mathbf{v}}) = \frac{p(\bar{\mathbf{P}} | \boldsymbol{\theta}, \sigma_{\bar{\zeta}}, \bar{\mathbf{v}}) p(\boldsymbol{\theta}, \sigma_{\bar{\zeta}} | \bar{\mathbf{v}})}{p(\bar{\mathbf{P}} | \bar{\mathbf{v}})}$$

In chapter 5 the Metropolis Algorithm will be introduced as a method for inference. This algorithm only considers ratios of posterior distributions so it is only necessary to know the posterior up to proportionality giving:

$$p(\boldsymbol{\theta}, \sigma_{\bar{\zeta}} | \bar{\mathbf{P}}, \bar{\mathbf{v}}) \propto p(\bar{\mathbf{P}} | \boldsymbol{\theta}, \sigma_{\bar{\zeta}}, \bar{\mathbf{v}}) p(\boldsymbol{\theta}, \sigma_{\bar{\zeta}}),$$

assuming that the prior on the parameters are independent of $\bar{\mathbf{v}}$. According to the previous section each $\bar{P}_n \in \bar{\mathbf{P}}$ is modelled as the AW2P model with additive structural error both of which follow a conditional Gaussian distribution such that:

$$\begin{aligned} \bar{P}_n &= \text{AW2P}(\bar{v}_n; \boldsymbol{\theta}) + \bar{\zeta}_n = \hat{\bar{P}}_n + \bar{\zeta}_n \quad \text{with} \\ \hat{\bar{P}}_n | \bar{v}_n, \boldsymbol{\theta} &\sim \mathcal{N}(\mu_{\hat{\bar{P}}}(\bar{v}_n; \boldsymbol{\theta}), \sigma_{\hat{\bar{P}}}^2(\bar{v}_n; \boldsymbol{\theta})) \quad \text{and} \quad \bar{\zeta}_n | \sigma_{\bar{\zeta}}^2 \sim \mathcal{N}(0, \sigma_{\bar{\zeta}}^2) \Rightarrow \\ \bar{P}_n | \bar{v}_n, \boldsymbol{\theta}, \sigma_{\bar{\zeta}}^2 &\sim \mathcal{N}(\mu_{\hat{\bar{P}}}(\bar{v}_n; \boldsymbol{\theta}), \sigma_{\hat{\bar{P}}}^2(\bar{v}_n; \boldsymbol{\theta}) + \sigma_{\bar{\zeta}}^2) \end{aligned}$$

Let the total variance be denoted as $\sigma_T^2(\bar{v}_n; \boldsymbol{\theta}, \sigma_{\bar{\zeta}}) = \sigma_{\hat{\bar{P}}}^2(\bar{v}_n; \boldsymbol{\theta}) + \sigma_{\bar{\zeta}}^2$. The likelihood distribution is then:

$$\begin{aligned} p(\bar{\mathbf{P}} | \boldsymbol{\theta}, \sigma_{\bar{\zeta}}, \bar{\mathbf{v}}) &= \frac{1}{(\sqrt{2\pi})^{N_{\text{dat}}} \prod_{n=1}^{N_{\text{dat}}} \sigma_T(\bar{v}_n; \boldsymbol{\theta}, \sigma_{\bar{\zeta}})} \exp \left(-\frac{1}{2} \sum_{n=1}^{N_{\text{dat}}} \left(\frac{\bar{P}_n - \mu_{\hat{\bar{P}}}(\bar{v}_n; \boldsymbol{\theta})}{\sigma_T(\bar{v}_n; \boldsymbol{\theta}, \sigma_{\bar{\zeta}})} \right)^2 \right) \\ &\propto \prod_{n=1}^{N_{\text{dat}}} \sigma_T^{-1}(\bar{v}_n; \boldsymbol{\theta}, \sigma_{\bar{\zeta}}) \exp \left(-\frac{1}{2} \left(\frac{\bar{P}_n - \mu_{\hat{\bar{P}}}(\bar{v}_n; \boldsymbol{\theta})}{\sigma_T(\bar{v}_n; \boldsymbol{\theta}, \sigma_{\bar{\zeta}})} \right)^2 \right), \end{aligned} \quad (4.23)$$

For the prior $p(\boldsymbol{\theta}, \sigma_{\bar{\zeta}}) = p(C_{\text{max}}, a_1, a_2, \sigma_{\bar{\zeta}})$, the parameters are assumed mutually independent which allows factoring the prior:

$$p(C_{\text{max}}, a_1, a_2, \sigma_{\bar{\zeta}}) = p(C_{\text{max}}) p(a_1) p(a_2) p(\sigma_{\bar{\zeta}}) \quad (4.24)$$

Little information is known about the W2P model parameters in $\boldsymbol{\theta}$. Theoretically $C_{\text{max}} \leq 16/27$ as described in section 1.1.2, but as shown in figure 3.7, uncertainty in the input might cause C_{max} to appear larger than the limit so it will be assumed that C_{max} follows a uniform distribution in the range $[0, 1]$. From the theory about pitch control it is also known that $a_1 \geq 0$ and $a_2 \leq 0$ - some experimentation with the available data shows that they are in the range $[0, 1]$ and $[-0.5, 0]$ for a_1 and a_2 so a uniform distribution in those ranges is assumed. Little is known about the structural error standard deviation $\sigma_{\bar{\zeta}}$ so the non-informative Jeffrey's prior based on Gaussian likelihood will be assumed. Thus:

$$p(C_{\text{max}}, a_1, a_2, \sigma_{\bar{\zeta}}) \propto \mathbf{1}_{[0,1]}(C_{\text{max}}) \mathbf{1}_{[0,1]}(a_1) \mathbf{1}_{[-0.5,0]}(a_2) \sigma_{\bar{\zeta}}^{-2}, \quad (4.25)$$

^[9]Again, the subscripts of the pdf's are omitted to ease the notation.

where $\mathbf{1}$ is the indicator function mapping to either 0 or 1 - see appendix B. For large N_{dat} , the numerical values for the posterior typically approach 0 and to overcome this the *log-posterior* distribution is used giving large negative values which suffer less from numerical issues. Combining (4.23) and (4.25) gives the log-posterior distribution:

$$\begin{aligned} \ell(\underbrace{C_{\max}, a_1, a_2}_{\boldsymbol{\theta}}, \sigma_{\bar{\zeta}} | \bar{\mathbf{P}}, \bar{\mathbf{v}}) &= \log p(C_{\max}, a_1, a_2, \sigma_{\bar{\zeta}} | \bar{\mathbf{P}}, \bar{\mathbf{v}}) \\ &\propto - \sum_{n=1}^{N_{\text{dat}}} \left(\log(\sigma_T(\bar{v}_n; \boldsymbol{\theta}, \sigma_{\bar{\zeta}})) - \frac{1}{2} \left(\frac{\bar{P}_n - \mu_{\hat{P}}(\bar{v}_n; \boldsymbol{\theta})}{\sigma_T(\bar{v}_n; \boldsymbol{\theta}, \sigma_{\bar{\zeta}})} \right)^2 \right) \\ &\quad - 2 \log(\sigma_{\bar{\zeta}}) + \ell(C_{\max}) + \ell(a_1) + \ell(a_2), \end{aligned} \quad (4.26)$$

where the last three terms are the *log-priors* for the W2P parameters which maps to either $-\infty$ or 0.

To aid the reader in understanding the process, figure 4.2 summarises the different computations needed to evaluate to log posterior.

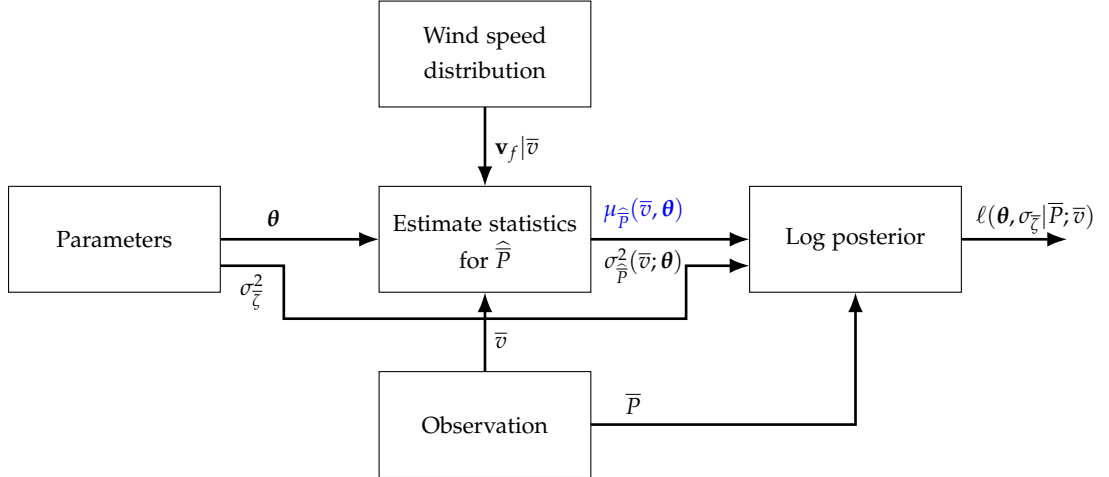


Figure 4.2: The process of computing the log posterior given parameters, observations and wind speed distribution. The example shown here is for a single observation, but the principle generalises to more observations. The wind speed distribution is derived in section 4.3 either assuming measured free wind speed (AW2P model) or modelling it via the NTF (NTF-AW2P model). The mean and variance for \hat{P} is computed based on the considerations in section 4.4. The mean $\mu_{\hat{P}}(\bar{v}; \boldsymbol{\theta})$ is particularly important since the AW2P and NTF-AW2P model estimate power \bar{P} using this. The log posterior is given by (4.26) using the shown inputs.

4.6 Implementation

The main computational effort of evaluating the log posterior, as shown in figure 4.2, lies in estimating the wind speed distributions $v_f | \bar{v}$ and evaluating the statistics for the modelled power \hat{P} . For any observation the statistics are determined by the measured wind speed,

wind direction and WTG number which determines which values to use from the lookup tables. Specifically, the wind speed distributions uses the mean and variance of the wind speed standard deviation μ_{σ_v} and σ_v^2 from the lookup tables. As described in section 2.2, the lookup tables provide different statistics for each WTG, each of 12 wind direction sectors and each wind speed between 3 and 30 m/s with 1 m/s interval. The statistics have been extrapolated and then interpolated to get values for each wind speed between 1.8 and 10 m/s with 0.05 m/s intervals giving statistics for 165 separate wind speeds^[10]. The South America and Asia wind farms both contain 32 WTG's so in total there are 2 wind farms \times 32 WTG's \times 12 wind sectors \times 165 wind speeds = 126720 different wind speed distributions. Computing the log posterior for a set of observations will only require a subset of the distributions depending on the experiment. Furthermore, the data shows that the majority of wind directions tends to fall within 3 or 4 sectors further limiting the needed distributions.

Wind speed distributions The free wind speed distributions have been estimated assuming no input uncertainty according to section 4.3.1 using numerical integration and assuming input uncertainty according to section 4.3.2 using Monte Carlo simulation. In each case the distributions are evaluated at 700 equidistant points for wind speeds between 0 and 15 m/s. As an example $p_{v_t}(v_t | \bar{v} = 4 \text{ m/s})$ is tabulated for v_t at 700 points between 0 and 15 m/s and $p_{v_t, v_{t'}}(v_t, v_{t'} | \bar{v} = 4 \text{ m/s})$ in an equivalent grid of 700^2 points^[11] For the Monte Carlo simulation each distribution kernel density estimate was based on about 10^5 realisations each. Figure 4.4 shows a few examples of the tabulated distributions.

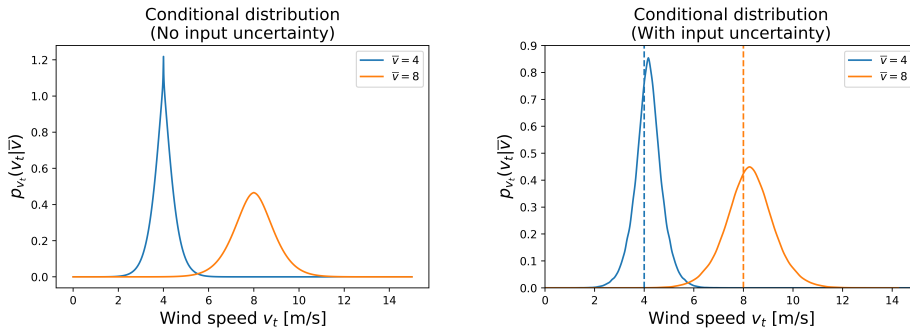


Figure 4.3: Examples of marginal wind speed distributions for a WTG at a particular section with different observed average wind speeds. Notice that, when accounting for input uncertainty, the most probable wind speed is not the observed average marked by the dashed line in the left plot.

^[10]Extrapolation is done with the least squares fit for a straight line for μ_{σ_v} and a second order polynomial for σ_v^2 to get the statistics for 1 and 2 m/s. Then linear interpolation is used to increase wind speed the resolution to 0.05 m/s. The range 1.8 to 10 is chosen such that it slightly exceeds the range of the cut-in (2.8 m/s) and rated (9.3 m/s) wind speed

^[11]Remember that $p_{v_t, v_{t'}}(v_t, v_{t'} | \bar{v})$ only depends on the Euclidean distance of v_t and $v_{t'}$ to \bar{v} so in practice is sufficient to evaluate the joint distribution on a line.

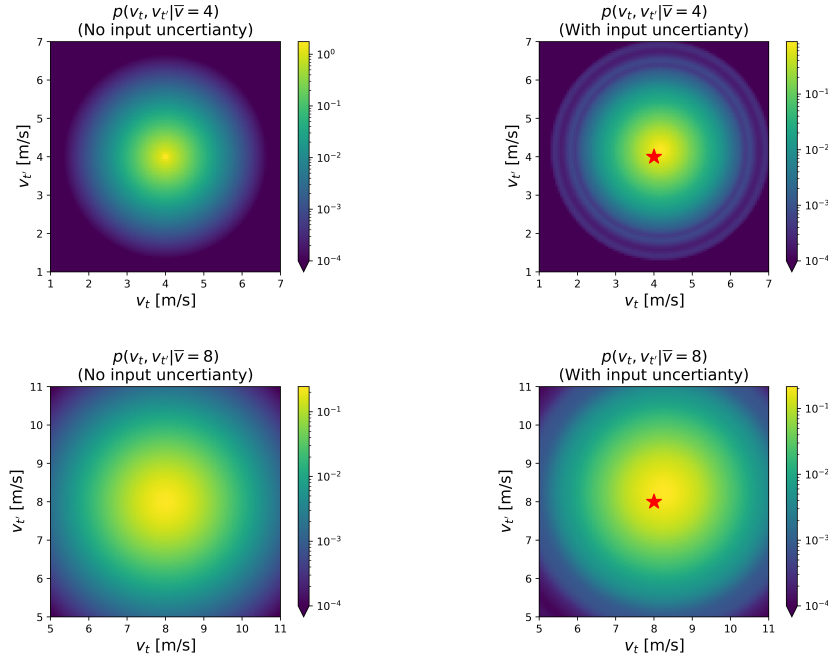


Figure 4.4: Examples of joint wind speed distributions for a WTG similar to figure 4.3. The red stars in the left plots mark the observed average. The upper left picture features some distinct ripples. These are not expected, but are explained by the use of Monte Carlo simulation to obtain the distribution and the fact that the logarithmic scale enhances small differences. Note that the ripples are necessarily circular since the distribution is only tabulated on a line and then repeated based on distance away from the center.

Power distributions Given the wind speed distributions, the mean and variance of the power distribution is estimated according to section 4.4. Figure 4.5 shows some examples at different wind speeds with no input uncertainty and figure 4.6 shows examples with input uncertainty. The script `likelihood_stats.py` computes these estimates - see appendix D. The figure clearly exemplifies some of the inaccuracies by using approximate distributions. At wind speeds between 3 and 9 m/s the assumption of a normal distribution seems reasonable but as the wind speeds approaches the cut-in, where power production is zero the distributions become left-screwed. Similarly, when the wind speed approaches the rated wind speed where power production becomes constant the distributions become right-screwed. The mean and variance seem to agree with the simulations in the upper plot but the lower plot where input uncertainty is assumed the variance estimates disagree with the sample variances. The authors were unable to locate the source of this disagreement. One explanation is that a slight bias, perhaps in the wind speed distributions, is amplified through the inference computations. In general the inference computations were very sensitive to the implementation method. As an example it was found that computing covariance as $\text{Cov}[v_t, v_{t'}] = E[(v_t - E[v_t])(v_{t'} - E[v_{t'}])]$ yielded much different results than $\text{Cov}[v_t, v_{t'}] = E[v_t v_{t'}] - E[v_t] E[v_{t'}]$ although they are theoretically equivalent. Despite this the variance estimates will be used in the remaining analysis despite their flaws. See chapter 7 for further discussion.

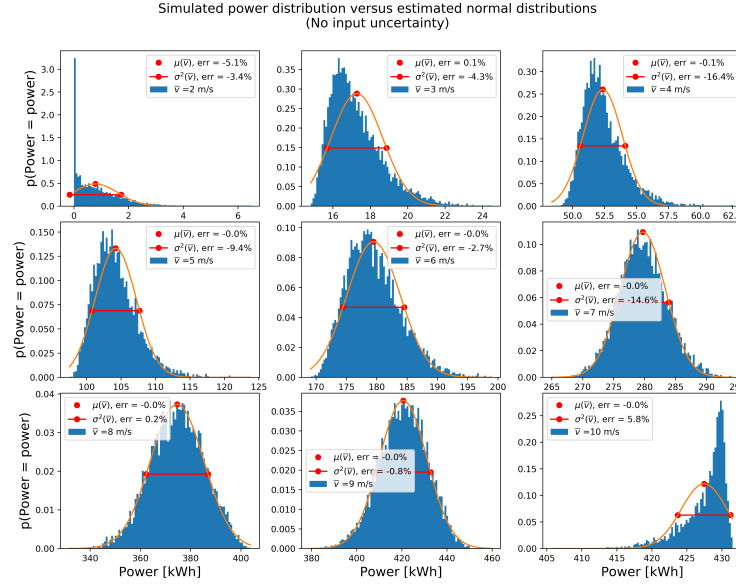


Figure 4.5: Distribution of power for one WTG using the AW2P model with no input uncertainty with respect to average measured wind speed. Average wind speeds between 2 and 10 m/s are simulated with known W2P model parameters θ . The single red dot mark the estimated mean and the line mark the 25th and 75th percentile of the fitted normal distribution. The blue histograms are simulated using the same θ . The sample mean and variance are compared and compared to the estimated values where the relative error to the sample statistics are shown in the legends.

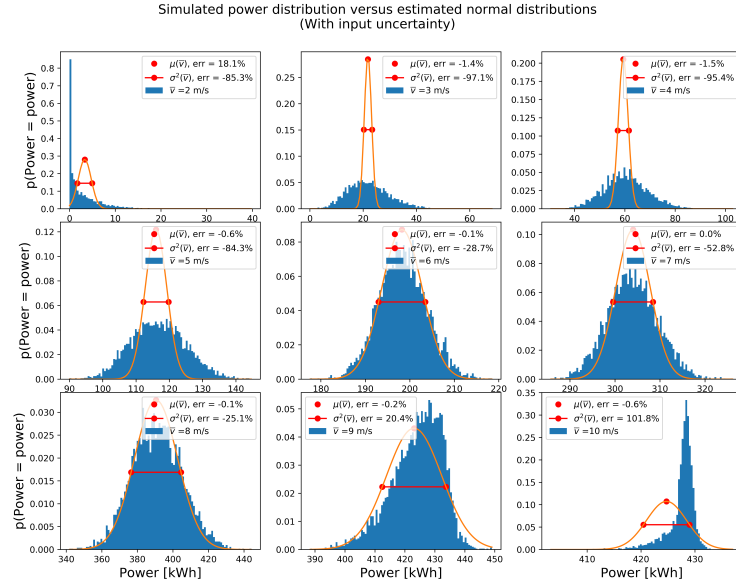
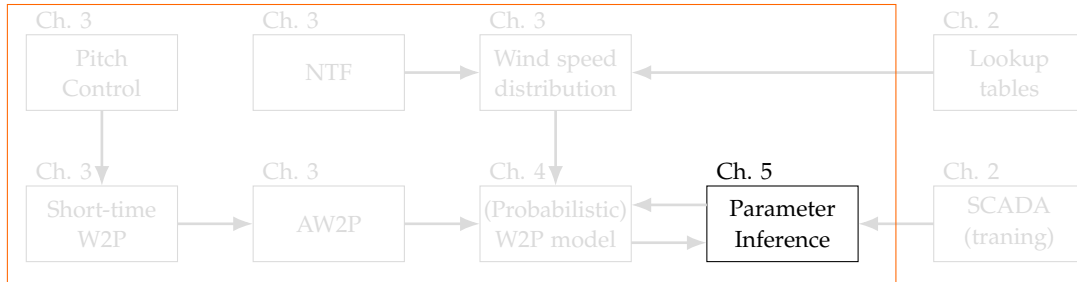


Figure 4.6: Distributions of power similar to figure 4.5 but using the NTF-AW2P model where input uncertainty is modelled through the NTF.

5 | Markov Chain Monte Carlo Methods

Chapter 4 presented an uncertainty framework to characterise the probability distribution of power generated from a W2P model using either the AW2P or NTF-AW2P model followed by the derivation of the posterior distribution for the parameters given a set of observations. To estimate parameters, [27] suggests using MCMC methods and in particular the Metropolis Algorithm. MCMC methods are used to sample from distributions - in particular, distributions that would be difficult to sample from using simulation methods such as the inverse transformation or the rejection method [46, ch. 5]. Some MCMC methods like the Metropolis Algorithm have the added flexibility that the pdf of distribution is only needed up to proportionality so any constants that are difficult to obtain analytically or are computationally complex can be left out. For parameter estimation, a sufficient number of simulations obtained using MCMC methods can be used to infer statistics about the parameters like the sample mean or even confidence intervals. This chapter will introduce the Metropolis Algorithm as a method for parameter estimation as highlighted below. A few theoretical facts are presented, but more detailed considerations are left out. Finally, the Metropolis Algorithm will be tested using synthetic data.

Contract W2P model



5.1 Metropolis Algorithm

Let $\mathbf{Z} \in \mathbb{R}^d$ be a random variable with pdf $p_{\mathbf{Z}}(\mathbf{z})$. The principle behind the Metropolis Algorithm is to simulate \mathbf{Z} with target pdf $p_{\mathbf{Z}}$ by simulating a sequence of realisations

$$\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(\tau)}$$

such that the distribution of $\mathbf{Z}^{(\tau)}$ converges to the distribution of \mathbf{Z} for $\tau \rightarrow \infty$ by some definition of convergence. This is done by imposing a *first order Markov property* on the sequence which gives that [46, p. 445]:

$$P\left(\mathbf{Z}^{(\tau+1)} = \mathbf{z}^{(\tau+1)} \mid \mathbf{Z}^{(\tau)} = \mathbf{z}^{(\tau)}, \dots, \mathbf{Z}^{(1)} = \mathbf{z}^{(1)}\right) = P\left(\mathbf{Z}^{(\tau+1)} = \mathbf{z}^{(\tau+1)} \mid \mathbf{Z}^{(\tau)} = \mathbf{z}^{(\tau)}\right),$$

which reduces the problem of simulating the sequence to simulating $\mathbf{z}^{(\tau+1)}$ given $\mathbf{z}^{(\tau)}$. The Metropolis Algorithm does this in a rejection sample manner where new states \mathbf{z}^* are simulated from a proposal distribution with pdf $q(\mathbf{z}^*|\mathbf{z}^{(\tau)})$, which depends only on the previous state $\mathbf{z}^{(\tau)}$. The proposal distribution chosen should be sufficiently simple such that direct sampling can be done in an efficient manner. In each iteration of the algorithm a candidate realization \mathbf{z}^* is drawn from the proposal distribution and accepted with the following probability:

$$A(\mathbf{z}^*, \mathbf{z}^{(\tau)}) = \min \left\{ 1, \frac{p_Z(\mathbf{z}^*)}{p_Z(\mathbf{z}^{(\tau)})} \right\}. \quad (5.1)$$

This is realised by simulating a random number u with distribution $U \sim \mathcal{U}(0,1)$ and accepting the candidate if $A(\mathbf{z}^*, \mathbf{z}^{(\tau)}) > u$. By doing so, \mathbf{z}^* is accepted with probability 1 if it increases the probability $p_Z(\mathbf{z}^*) \geq p_Z(\mathbf{z}^{(\tau)})$ and otherwise according to the ratio in (5.1). If \mathbf{z}^* is accepted a new element is added to the sequence such that $\mathbf{z}^{(\tau+1)} = \mathbf{z}^*$, otherwise \mathbf{z}^* is rejected and $\mathbf{z}^{(\tau+1)} = \mathbf{z}^{(\tau)}$. The ratio in (5.1) is only applicable for symmetric proposal distribution where $q(\mathbf{z}^*|\mathbf{z}^{(\tau)}) = q(\mathbf{z}^{(\tau)}|\mathbf{z}^*)$. For asymmetrical q the proposal density is included in the ratio which is known as the *Metropolis-Hastings Algorithm* [5, p. 541]. As mentioned, p is only needed up to proportionality - (5.1) shows this since any constants simply vanish in the ratio so if $p_Z(\mathbf{z}) = c\tilde{p}_Z(\mathbf{z})$ it is sufficient to know $\tilde{p}_Z(\mathbf{z})$. The Metropolis Algorithm is summarized in algorithm 2 and implemented in the script `metropolis.py` - see appendix D.

Algorithm 2 Metropolis Algorithm

Input: Target distribution pdf p , proposal density q , number of iterations T and initial state $\mathbf{z}^{(0)}$

for $\tau = 1, \dots, T$ **do**

Simulate $(\mathbf{Z}^* = \mathbf{z}^*) \sim q(\cdot|\mathbf{z}^{(\tau)})$ Sample proposal

Simulate $(U = u) \sim \mathcal{U}(0,1)$

if $A(\mathbf{z}^*, \mathbf{z}^{(\tau)}) > u$ **then**

$\mathbf{z}^{(\tau+1)} \leftarrow \mathbf{z}^*$ Accept sample

else

$\mathbf{z}^{(\tau+1)} \leftarrow \mathbf{z}^{(\tau)}$ Reject sample

return $\{\mathbf{z}^{(\tau)}\}_{\tau=0}^T$

From an optimization point of view the Metropolis Algorithm seems reasonable and the fact that candidate states are sometimes accepted despite lowering the pdf can help \mathbf{z} escape from a local minimum of p . It also turns out that $\mathbf{Z}^{(\tau)}$ converges to the true distribution and in the Metropolis Algorithm but the details and exact conditions are omitted here - see [5]. Implementing the Metropolis Algorithm in practise requires some considerations.

Proposal distribution The choice of proposal distribution will highly affect the *acceptance ratio* which is defined as the ratio of accepted samples to the total number generated. A suggestion is to use an uncorrelated Gaussian proposal distribution with mean in the current state and the variance σ_q^2 such that $\mathbf{Z}^*|\mathbf{z}^{(\tau)} \sim \mathcal{N}_d(\mathbf{z}^{(\tau)}, \sigma_q^2 \mathbf{I}_{d \times d})$. The value of σ_q^2 can be used to obtain the desired acceptance ratio - a rule of thumb is to choose it such

that 20 – 50% of samples are accepted on average to sufficiently explore the target distribution [1, p. 20]. If the range of likely values for the different entries of \mathbf{z} differs by one or more orders of magnitude, the target proposal distribution can be changed to give an individual variance for each parameter such that the covariance structure is a matrix with $(\sigma_{q,1}^2, \dots, \sigma_{q,d}^2)$ on the main diagonal and zeroes elsewhere.

Log distributions Metropolis Algorithm is easily converted to handle densities in the logarithmic domain like the log-posterior distribution derived in section 4.5. Given $\ell(\mathbf{z}) = \log(p(\mathbf{z}))$, samples will be accepted with the following pseudo-probability:

$$A_\ell(\mathbf{z}^*, \mathbf{z}^{(\tau)}) = \min \left\{ 0, \ell(\mathbf{z}^*) - \ell(\mathbf{z}^{(\tau)}) \right\}.$$

Accepting \mathbf{z}^* when $A_\ell(\mathbf{z}^*, \mathbf{z}^{(\tau)}) > \log(u)$ gives an equivalent algorithm for u simulated uniformly between 0 and 1.

Burn in When using the Metropolis Algorithm to infer statistics about \mathbf{Z} - e.g mean or variance - it is often required to discard some of the initial sequence depending on how $\mathbf{z}^{(0)}$ is initialised. This is referred to as *burn-in* and from the generated sequence $\mathbf{z}^{(0)}, \dots, \mathbf{z}^{(\tau)}$ the first N_{burn} states are discarded.

5.2 Experiment on synthetic data

In this section the Metropolis Algorithm will be used to simulate the parameters for the AW2P and NTF-AW2P model given their respective log-posterior distributions using synthetic data generated according to algorithm 1. The two models share the same set of parameters $(a_1, a_2, C_{\max}, \sigma_{\bar{\zeta}})$ as well as log posterior in (4.26), but the mean and variance estimates of modelled power $\mu_{\bar{P}}(\bar{v}; \theta)$ and $\sigma_{\bar{P}}^2(\bar{v}; \theta)$ and settings for simulating data are different.

In the following test the state consists of $\mathbf{z}^{(\tau)} = [a_1^{(\tau)} \ a_2^{(\tau)} \ C_{\max}^{(\tau)} \ \sigma_{\bar{\zeta}}^{(\tau)}]^T \in \mathbb{R}^4$. When choosing the initial state $\mathbf{z}^{(0)}$ to ease the parameter estimation $a_1^{(0)}, a_2^{(0)}$ and $C_{\max}^{(0)}$ are set to the least squares estimate of the deterministic W2P model which the AW2P and NTF-AW2P models are based on. The model is given by (3.6) and (3.7) and is piecewise linear with respect to the parameters, so given 10-minute power \bar{P} and 10-minute average wind speed \bar{v} the parameters are found by constructing an appropriate design matrix and using the normal equations - see appendix B. Choosing the initial state for $\sigma_{\bar{\zeta}}^{(0)}$ is done somewhat arbitrarily. When using real data in the next chapter it will be based on a qualified guess and in the experiment here it is chosen as the true value with added random noise.

For proposal distribution an uncorrelated multivariate Gaussian was chosen. The parameters (a_1, a_2, C_{\max}) are numerically bounded between -0.5 and 1 and have been grouped to have the same variance $\sigma_{q, \text{W2P}}^2$ in the proposal distribution. The structural error $\sigma_{\bar{\zeta}}$ is 2 in the synthetic data and is given a separate variance $\sigma_{q, \bar{\zeta}}^2$ in the proposal distribution. The

exact values of $\sigma_{q,W2P}^2$ and $\sigma_{q,\zeta}^2$ are chosen based on empirical experimentation to get an acceptance rate of 20 – 50%. An experiment simulating the parameters with log-posterior given by the AW2P model is shown below. The experiment uses 10,000 simulated data-points (\bar{P}_n, \bar{v}_n) for $n = 1, \dots, 10000$ with wind speed statistics from the same WTG in the South American wind farm and average wind speed for each between 1.8 and 10 m/s. Simulation is done without the influence of the NTF on the average measured wind speed by setting NTF = FALSE in algorithm 1. See figure 5.1 for results.

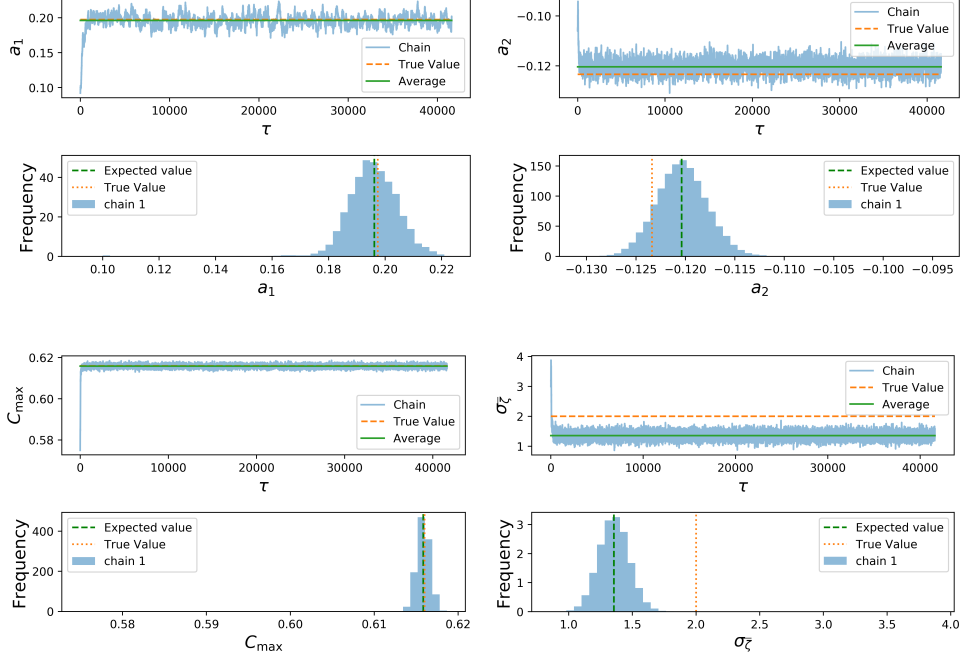


Figure 5.1: Traceplots and histograms for the different states using the Metropolis Algorithm given the log-posterior for the AW2P model. The algorithm used 10^5 iterations and an acceptance rate of 40% was achieved. The average of each entry in the sequence after a burn-in of 5000 states is shown in the traceplots.

A similar experiment is done for the NTF-AW2P model using the log-posterior for this. Synthetic Data is simulated with the influence of the NTF on the average measured wind speed by setting NTF = TRUE in algorithm 1. See figure 5.2 for results.

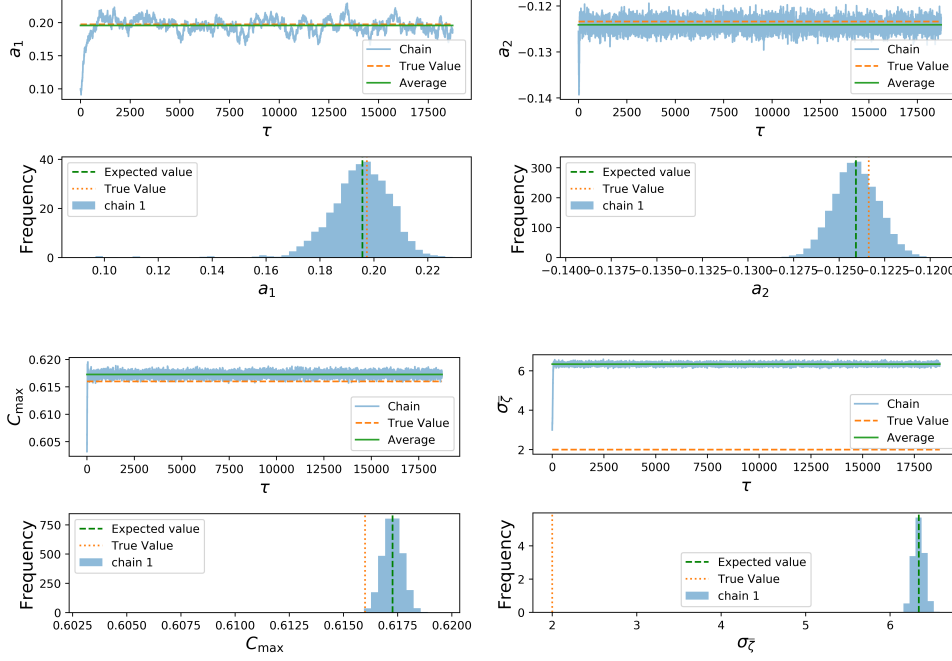


Figure 5.2: Traceplots and histograms for the different states using the Metropolis Algorithm given the log-posterior for the NTF-AW2P model. The algorithm used 10^5 iterations and an acceptance rate of 19% was achieved. The average of each entry in the sequence after a burn-in of 5000 states is shown in the traceplots.

Results from the two experiments show that the parameters converge after a few thousand steps with relatively small variance after the burn-in period. Some of the parameters seem to exhibit a bias when comparing the average values after burn-in to the true parameter values. This is quantified in table 5.1.

	$(\hat{a}_1 - a_1)$	$(\hat{a}_2 - a_2)$	$(\hat{C}_{\max} - C_{\max})$	$(\hat{\sigma}_{\zeta} - \sigma_{\zeta})$
AW2P	$-1.60 \cdot 10^{-3}$	$3.00 \cdot 10^{-3}$	$-2.00 \cdot 10^{-4}$	$-6.44 \cdot 10^{-1}$
NTF AW2P	$-3.30 \cdot 10^{-3}$	$-7.00 \cdot 10^{-4}$	$1.30 \cdot 10^{-4}$	4.34

Table 5.1: Difference between the true parameter values $(a_1, a_2, C_{\max}, \sigma_{\zeta})$ and the estimates ones for synthetic data $(\hat{a}_1, \hat{a}_2, \hat{C}_{\max}, \hat{\sigma}_{\zeta})$ for the two experiments.

The table shows that the bias in parameter estimation is in the order of 10^{-4} to 10^{-3} for a_1 , a_2 and C_{\max} . The error could be due to random fluctuations although some of the histograms of the simulated sequences suggest a structural error. In either case the estimation error is relatively small. The error for estimated σ_{ζ} , however is larger in particular for the NTF-AW2P model. It is hypothesised that this is due to the variance estimate problems presented in section 4.6 - see further discussion in chapter 7.

In order to test if the performance of the Metropolis Algorithm is consistent, multiple sequences have been simulated on the same synthetic data but with different initial states. A good measure of consistency is convergence to the same parameters independent of initial state and random seed for the proposal distribution. Three different sequences are simulated for both the AW2P and NTF-AW2P model and the results can be seen in figure 5.3.

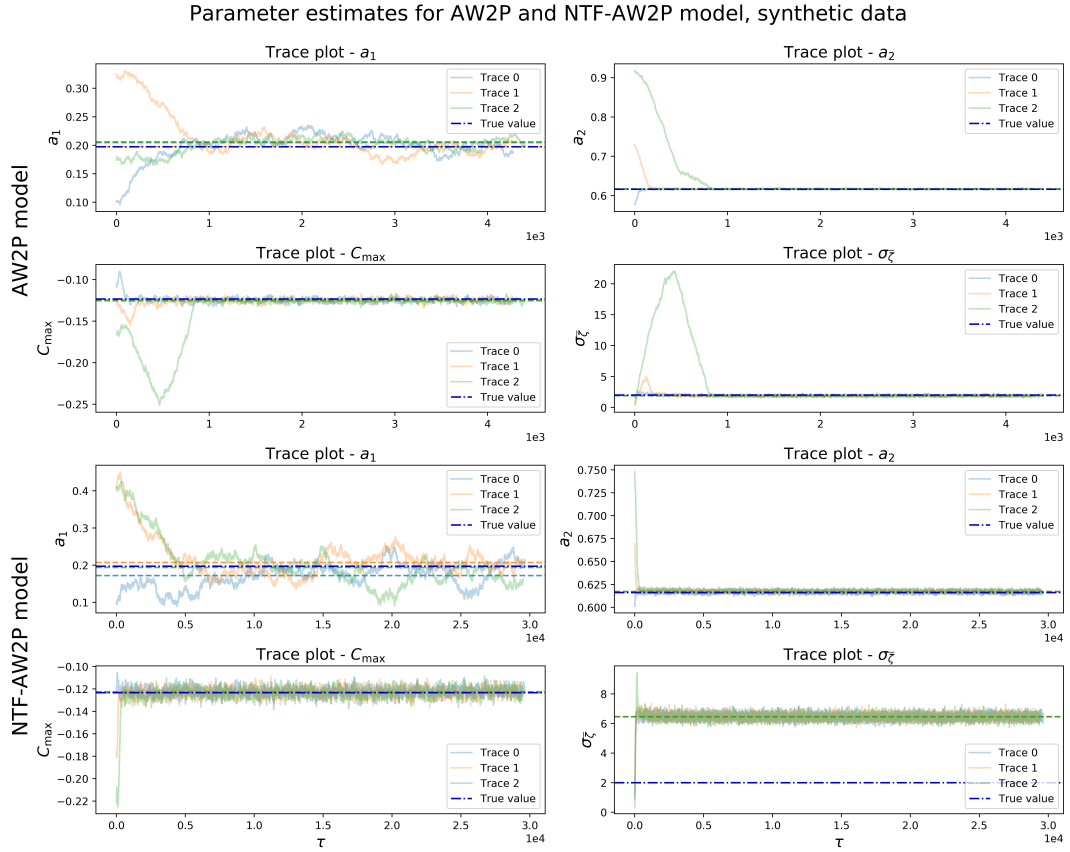


Figure 5.3: Traceplots for three runs of the of the Metropolis Algorithm on the same synthetic using either the AW2P or NTF-AW2P model. The dashed orange, light blue and green lines are the sample averages after the burn-in period although some these are difficult to distinguish here.

The results show that the simulated sequences converge to the same parameters with both models although the a_1 parameter has a larger variance than the others and the estimated means are slightly different. The different sequences for a_1 crosses paths multiple times so based on the Markov property, it is expected that the estimated means will also converge to the same values for longer chains.

6 | Performance Analysis

This Chapter presents the results from a few tests where the AW2P and NTF-AW2P models have been evaluated on the available SCADA data. For comparison, results using the method of bins model and 5-parameter logistic model presented in section 3.4 will also be shown. Section 6.1 describes the performed tests in detail and section 6.2 covers parameter estimation. Section 6.3 then compares the different models using various measures of error and section 6.4 analyses the residuals for the AW2P and NTF-AW2P models.

6.1 Case setup: Training and test

As described in chapter 2, the available data includes a South American and an Asian wind farm. To analyse the performance, the following case is considered.

Case 10 WTG's in the South America wind farm are installed and have been operational over a period of time. A customer now wants to invest in two new projects installing:

- 22 additional WTG's in the South America wind farm.
- 32 WTG's at a new location in Asia.

All the new WTG's are of the same type, but the site in Asia is mainland whereas the South America site is a coastal wind farm. The customer wants to know how much power production they can expect from the new projects under various inputs such as wind speed, temperature, and air density. To do this, lookup tables are produced using WRA software for the proposed new WTG's location. A power curve based on the lookup tables is estimated and delivered to the customer, who decides to invest in the projects based on the power curves^[1]. The WTG's are installed and after a while, the customer wants to know how accurate the power curves they were delivered based on measurements from the installed WTG's.

To simulate the proposed case, WTG 1-10 in South America will be treated as the existing WTG's which can be used as training data to estimate parameters for the models acting as the power curves. Given the parameters, the models are evaluated on the test data which are the remaining WTG's in the South America wind farm and the entire Asia wind farm. With the data setup, it is possible to analyse how well the models generalises from training to test data. In particular, it is interesting to see how well the models generalise to the Asia wind farm where the mainland climate is different from the climate of the coastal South America wind farm. Note that the time of installation described in the case does not reflect the actual time of installation - the time horizons is as described in chapter 2 and the training/test data is simply separated in groups of WTG's.

^[1]In a real case, the AEP (annual energy production) would be computed from the power curve using long-time weather predictions, but this is not considered here.

A few notes in regard to the SCADA data. In raw form, the SCADA data contains quite a bit of outliers when comparing produced power at different wind speeds. Some of these outliers are explained by the SCADA through status variables as described in section 2.1. After filtering for status variables, the data still contains outliers - in particular below the average power curve. The outliers can be filtered out by sorting the data into small wind speed bins and removing power outside some interval for each bin determined by certain quantiles. Figure 6.1 shown an example where data have been removed based on status variables and outliers have been removed based on chosen percentiles^[2]. Both training and test data have been filtered using the status variables, but outliers based on percentiles are only removed in the training data in order to get realistic error measurements in the test^[3]. Additionally, data points for wind speeds below 1.8 m/s and above 10 m/s are also removed from the data since the power should be 0 below 1.8 m/s and nominal power above 10 m/s and any deviations from this are assumed to be an error^[4]. After filtering the training data for WTG 1-10 in South America contains about 170,000 measurements, the test data for WTG 11-33 in South America contains about 410,000 measurements and the test data for the Asian wind farm contains about 2,500,000 measurements.

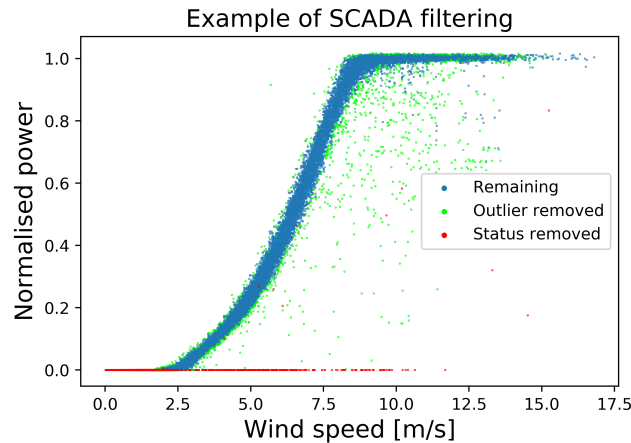


Figure 6.1: Removed data points for one WTG.

^[2]Larger outliers are seen at high wind speeds, so the values outside the 2nd and 99.5th percentile in the interval 0 to 7.5 m/s are removed while values outside the 7th and 99.5th percentile above 7.5 m/s are removed. The width of each wind speed bin is 0.05 m/s

^[3]Another reason why outliers have not been removed in the training data is the fact that finding the percentile values that removes all the “true outliers” and nothing else is quite challenging in particular when applying the percentiles across different WTG’s and the two wind farms.

^[4]In principle the power should be 0 below the cut-in wind speed at 2.8 m/s and nominal power above rated wind speed at 9.3 m/s, but in practise, these bounds are extended due to fluctuations in the wind speed within each 10-minute period.

6.2 Parameter estimation

The parameter estimates using the training data with the Metropolis Algorithm for the AW2P and NTF-AW2P models will be presented in this section. Similar to the test with synthetic data, the parameters to be estimated are the W2P parameters $\theta = (a_1, a_2, C_{\max})$ and the structural error standard deviation σ_{ζ} . The variances for the transition kernel was changed slightly from those used in the synthetic data to get an acceptance rate between 20% and 50%. The traceplots and histograms for the parameters can be seen in figure 6.2 and 6.3 for the AW2P and NTF-AW2P models respectively.

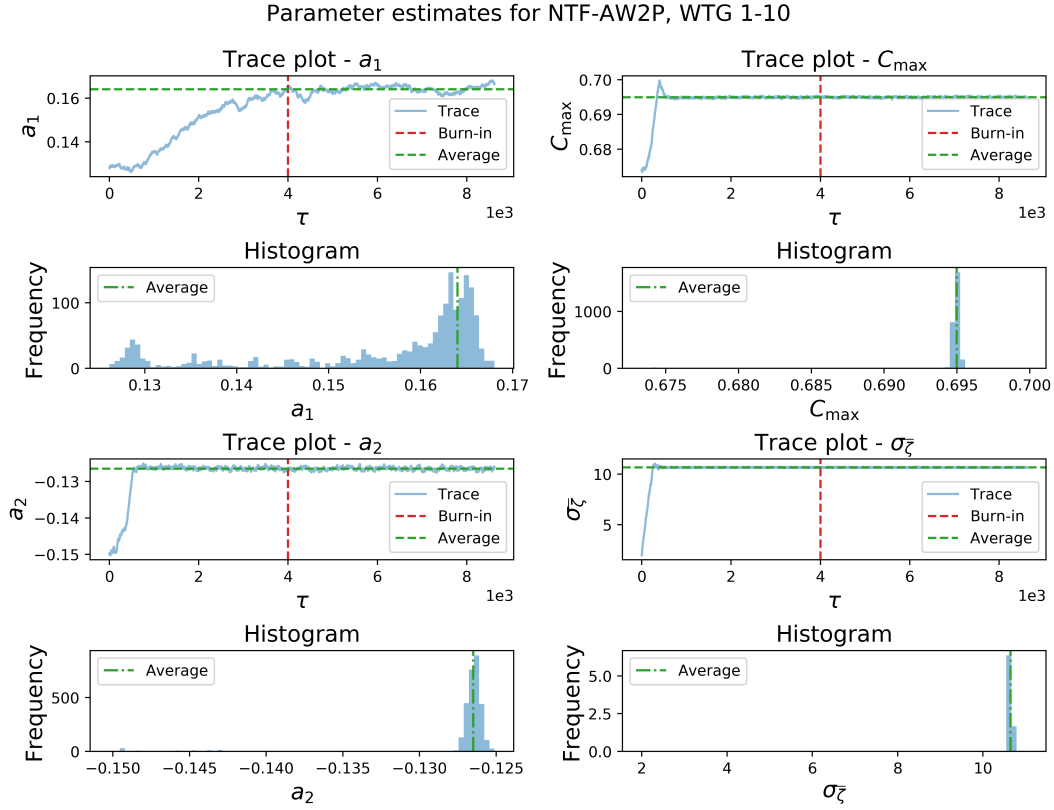


Figure 6.2: Parameter estimates for the AW2P model using training data with a burn-in of 4000. The average parameters and their sample standard deviation after the burn - i.e $\mu \pm \sigma$ - are $a_1 = 0.16 \pm 1.8 \cdot 10^{-3}$, $a_2 = -0.13 \pm 3.9 \cdot 10^{-4}$, $C_{\max} = 0.69 \pm 1.6 \cdot 10^{-4}$ and $\sigma_{\zeta} = 11 \pm 2.4 \cdot 10^{-2}$

In figure 6.2, the parameters for the AW2P model show convergence although a_1 is by far the slowest to converge requiring about 4 times as long before stabilising. Similar convergence behaviour was seen in parameter estimation using synthetic data suggesting that a_1 in the AW2P model generally converges slowly using the Metropolis Algorithm. The acceptance rate was 34% for the AW2P model.

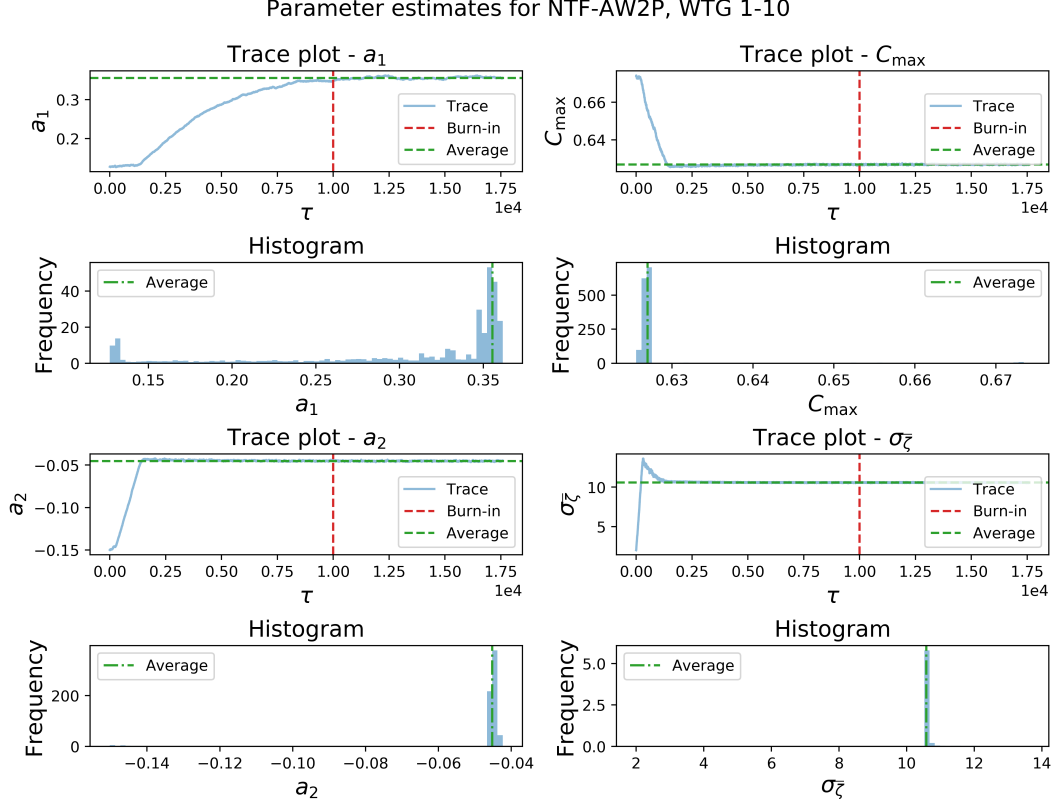


Figure 6.3: Parameter estimates for the NTF-AW2P model using training data with a burn-in of 10,000. The average parameters and their sample standard deviation - i.e $\mu \pm \sigma$ - after the burn in are $a_1 = 0.36 \pm 2.8 \cdot 10^{-3}$, $a_2 = -4.5 \cdot 10^{-2} \pm 4.2 \cdot 10^{-4}$, $C_{\max} = 0.63 \pm 1.7 \cdot 10^{-4}$ and $\sigma_{\zeta} = 11 \pm 2.4 \cdot 10^{-2}$

Similar convergence behaviour is seen for parameter estimation with the NTF-AW2P model in figure 6.3 although a_1 now required 10,000 accepted samples before convergence. The acceptance rate was similarly 34% for the NTF-AW2P model.

Comparing the parameters from the two models it is seen that σ_{ζ} are the same within the given precision while the other parameters differ more. C_{\max} for the NTF-AW2P model is lower than for the AW2P model but both values are above the theoretical limit at $16/27 \approx 59\%$ perhaps due to measuring uncertainties.

6.3 Performance

The performance of the AW2P and NTF-AW2P model will be assessed in this section by calculating the errors between modelled and observed power and the errors are summarised in a few different statistics. As a baseline, the method of bins (bin) and the 5-parameter logistic (logistic) model presented in section 3.4 are used as reference models. Before presenting the error statistics, the modelled power for the test and training data is presented in figure 6.4.

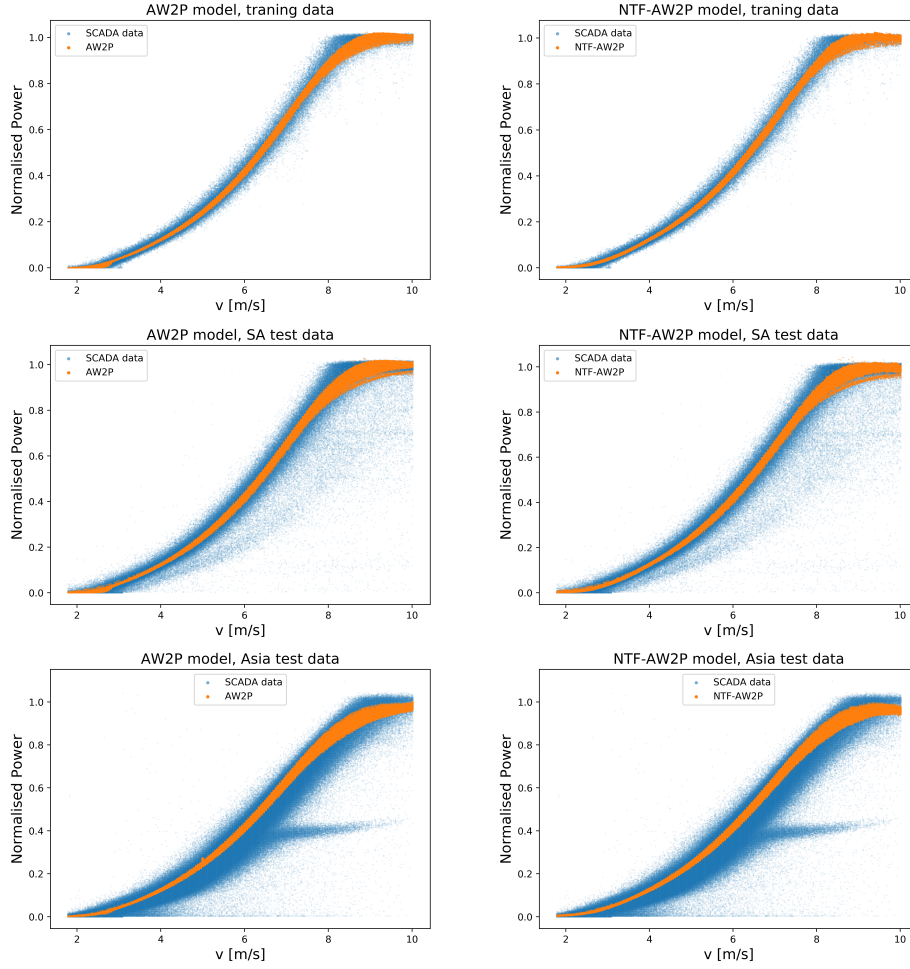


Figure 6.4: Modelled versus observed power for wind speeds between 1.8 and 10 m/s. The Asia test data in the lower plots shows some observations that seem systematically below the remaining power curve - this could be explained by some control errors e.g. in pitch control.

The error between modelled and observed power for the different models is defined as the difference measured in kWh. Given the error, the mean squared error (MSE), mean absolute error (MAE), median absolute error (50th percentile) (Med AE), and the mean error (ME) is computed. The training set and test sets have data from several WTG's so the minimum and maximum MAE (min MAE, max MAE) between WTG's for each data set is also computed. Finally, the MAE is computed in three regions depending on the wind speed. The regions are chosen based on the limits for the analysed wind speeds and different regions for the power coefficient C_P described in section 3.2.1. The regions are 1.8 – 3.87 m/s (reg 1), 3.87 – 7.93 m/s (reg 2) and 7.93 – 10 m/s (reg 3). In table 6.1 the error measurements are shown for the training and test data sets. To visualise the error figure 6.5 shows boxplots for the absolute error (AE) across the three data sets.

SA Training	MSE	MAE	Med AE	ME	min MAE	max MAE	reg 1 MAE	reg 2 MAE	reg 3 MAE
AW2P	153.0	8.1	4.9	0.4	6.7	9.6	4.0	10.1	12.6
NTF-AW2P	150.0	8.1	4.9	0.1	6.7	9.6	4.0	10.1	12.0
logistic	209.3	10.7	7.9	0.9	9.5	12.1	5.7	11.9	16.3
bin	144.3	7.9	4.9	0.1	6.4	10.0	3.9	9.9	11.7
SA Test	MSE	MAE	Med AE	ME	min MAE	max MAE	reg 1 MAE	reg 2 MAE	reg 3 MAE
AW2P	681.8	12.4	5.9	-3.8	9.1	41.8	5.0	14.2	21.2
NTF-AW2P	680.8	12.3	5.9	-4.0	9.1	41.9	5.1	14.1	21.0
logistic	716.8	14.4	8.5	-3.7	11.4	42.6	6.6	15.9	22.4
bin	679.0	12.1	5.8	-4.3	8.8	41.8	5.0	14.0	20.2
Asia Test	MSE	MAE	Med AE	ME	min MAE	max MAE	reg 1 MAE	reg 2 MAE	reg 3 MAE
AW2P	433.5	12.8	8.2	-9.6	8.9	19.4	5.8	16.5	24.4
NTF-AW2P	434.8	12.8	8.3	-9.8	8.8	19.6	5.8	16.5	24.8
logistic	359.5	10.9	6.2	-3.2	8.2	14.7	4.7	13.5	24.5
bin	344.3	10.8	6.5	-4.3	8.4	15.3	4.7	13.3	25.5

Table 6.1: Error statistics for the training and two test data sets for the four models AW2P, NTF-AW2P, 5-parameter logistic, and method of bins model. The smallest value for each error statistic in the training and test sets is marked with bold font.

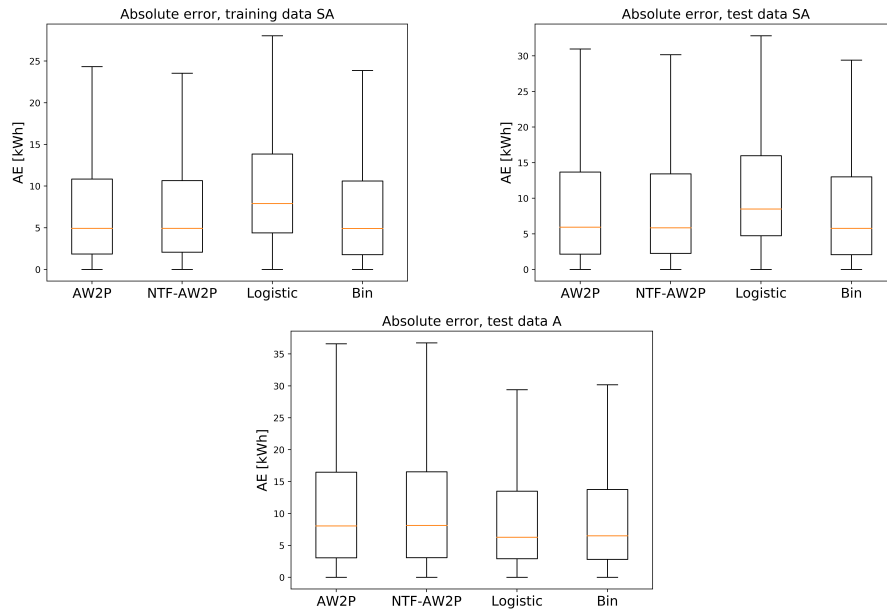


Figure 6.5: Boxplot for the absolute over the four models. The boxes mark Q_1 and Q_3 (25th and 75th percentile) and the whiskers are located at $Q_1 - 1.5(Q_3 - Q_1)$ and $Q_3 + 1.5(Q_3 - Q_1)$. Outliers outside the whiskers are not shown here.

In the training data, it is observed that the method of bins performs the best but comparable to that of the AW2P and NTF-AW2P across all error measurements, except for max MAE. The same trend is seen in the test data for the South-American wind farm, with the bin model having a slightly lower error in most of the error measurements. In the

Asian test set the bin and logistic model is seen to perform better compared to the AW2P and NTF-AW2P in all but one statistic, with the AW2P and NTF-AW2P having similar performance. The test data sets have negative mean error (bias) for all models, but this is explained by the negative outliers which were removed in the training set but not the test sets.

Lastly, the models are compared visually in a power curve. The models are not unique for particular wind speed, so to be able to compare the models in a single plot the average modelled power in wind bins of size 0.05 m/s is taken similarly to the method of bins model. See figure 6.6.

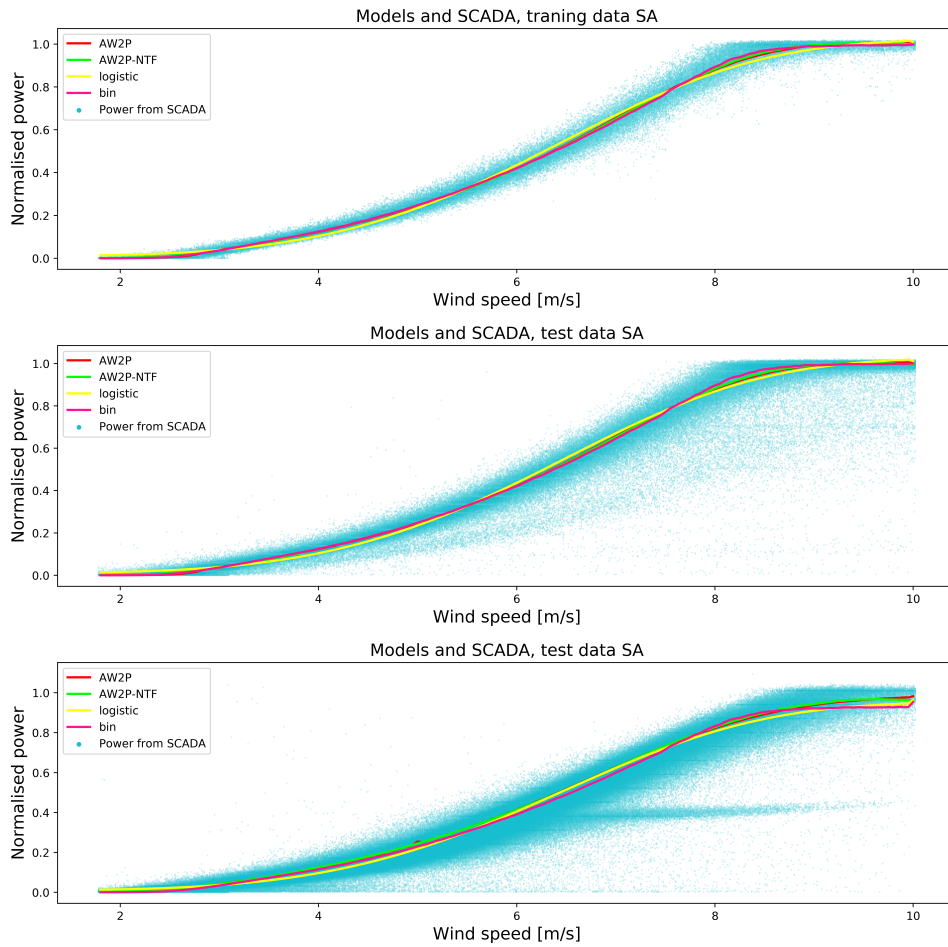


Figure 6.6: Average produced power for the four models at different wind speeds on the training and test data sets.

6.4 Residual analysis

To get further insight into the performance of the AW2P and NTF-AW2P model, the residual between the observed power and estimated power of the test data sets are evaluated in this section.

Since the distribution of produced power is available, approximated as a Gaussian distribution in section 4.4, the confidence intervals can be readily obtained. Figure 6.7 shows the estimated power along with the estimated 95% confidence intervals for the training data. The residual plots show that at little less than 95% of the observed data points are inside the 95% confidence for the test data while the training data show better alignment. This is partly expected since outliers are not removed in the test data, but other effects might cause the data points to exceed the confidence intervals more than they should according to the models.

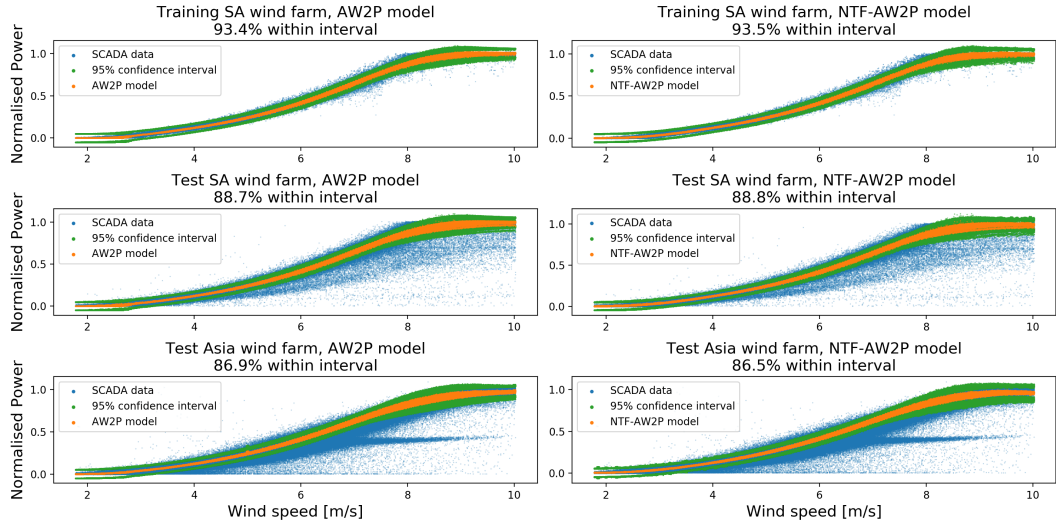


Figure 6.7: Power at different wind speeds observed in SCADA data and estimated via the AW2P and NTF-AW2P model with accompanying 95% confidence intervals. The figure titles reveal how many percent of the observed data points fall within the confidence intervals.

AW2P and NTF-AW2P models both yield an estimate for the mean and variance for the approximated Gaussian distribution. To check how accurate the approximation is, the standardised residuals can be analysed using that they should follow a standard Gaussian distribution - i.e:

$$\frac{\bar{P} - \mu_{\bar{P}}(\bar{v}, \theta)}{\sqrt{\sigma_{\bar{P}}^2(\bar{v}, \theta)}} \sim \mathcal{N}(0, 1)$$

This is analysed in a few ways. Firstly, figure 6.8 shows the standardised residuals plotted along different inputs to check for any remaining dependence on the input variables. Secondly the distribution is more directly analysed in figure 6.9 in a histogram and a Quantile-Quantile (Q-Q) plot.

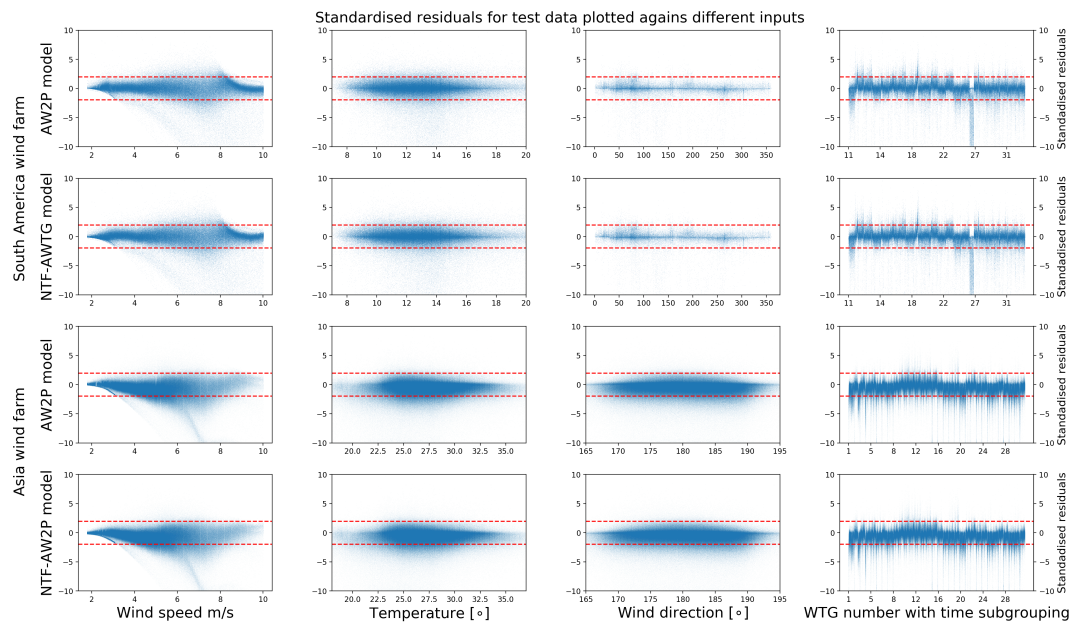


Figure 6.8: Standardised residuals (along the ordinate) for the different models and wind farms in the test data with. It should be noted that WTG 26 in the South American wind farm has significant outliers. The dashed red lines are the 95% confidence intervals of a standard Gaussian.

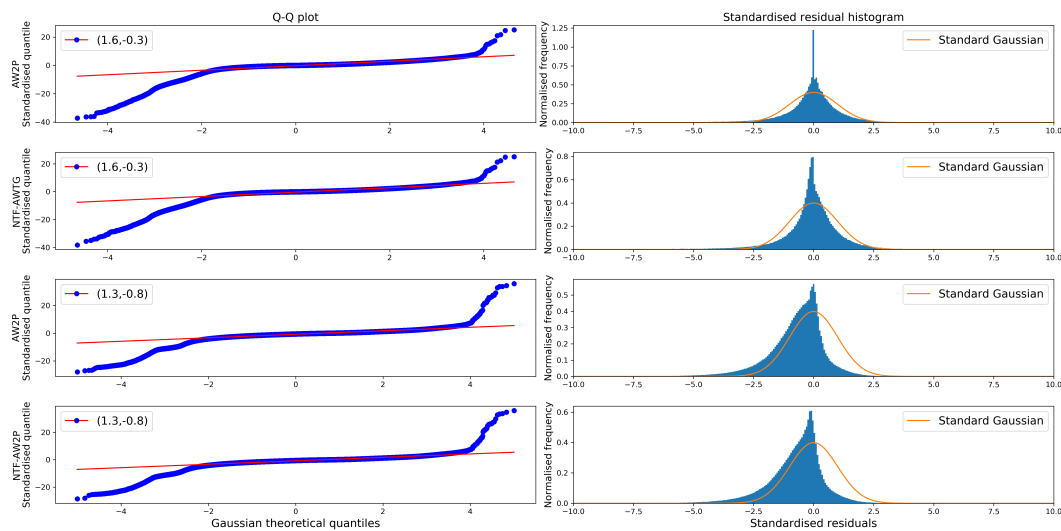


Figure 6.9: Q-Q plots and histograms for the standardised residuals for the different models and wind farms in the test data. The labels in the Q-Q plots show the slope and intercept of the best straight line fit which should be (1,0) if the normalised residuals were truly standard Gaussian distributed.

The residual plots could suggest that the wind speed influences the normalised residuals while dependence on the temperature and wind direction is not clear based on visual inspection. The residual plot concerning the WTG number shows that different WTG's have slightly different residual distributions although it is not clear how and the time subgrouping might also affect the residuals due to e.g. seasonal fluctuations.

The previous observations showed that data mainly have negative outliers and the Q-Q plots confirm this by showing larger values for the large quantiles - in particular for the negative quantiles. The histograms also show that the residual distribution is not quite Gaussian with heavier tails. It is interesting to observe that the distribution for the residuals in the Asia wind farm are negatively skewed while it is more or less symmetric for the Southern American wind farm.

Based on the different visualisations it is safe to conclude that the normalised residuals are not exactly standard Gaussian distributed although some similarities are seen. Reasons why the models do not align with the data could be due to operational reasons such as errors in the pitch control of the WTG blades, unreported regulation of the produced power due to energy grid considerations or errors in the measurements. Of course, the AW2P and NTF-AW2P models only consider a few of the physical phenomena at play, so the distribution of the residuals is surely also affected by processes that are not modelled - a few of these is mentioned in chapter 9.

7 | Discussion

The research question asked:

How is the power generation from a wind turbine generator modelled using parametric models accounting for uncertainties using Bayesian methods?

This chapter will discuss the methods and results presented so far in the context of the research question, as well as the accompanying study questions presented in section 1.3. The issue with variance estimation using the NTF-AW2P model is discussed initially.

Variance estimation error The AW2P and NTF-AW2P model approximates the distribution for the modelled power with a Gaussian distribution by estimating its mean and variance. In figure 4.6 it was shown that the variance estimates for wind speeds below 6 m/s and above 9 m/s for the NTF-AW2P model were noticeably lower than the sample variance obtained through Monte Carlo simulation. Possible explanations for this disagreement include:

1. Inaccurate wind speed distribution. Since the NTF-AW2P model is based on wind speed distributions obtained through Monte Carlo simulation, having too few realisations in the simulations could cause inaccuracies in the distributions that propagate through to the variance estimation. The method of simulation as described in algorithm 1 could also be mistaken.
2. Numerical issues when estimating the variance by numerical integration.
3. Procedural error. Some modelling or theoretical aspect could be mistaken in the variance estimation.
4. Implementation error.

Given that the AW2P and NTF-AW2P only differ in the modelling assumptions and methods for simulating wind speed distributions, it is likely that error in the variance estimate is caused by option 1 or 3 in the above list. This is only speculation, but a relevant question is where and how the error affects the model. Importantly, the variance estimate is used to evaluate the log-posterior which is used for parameter inference. It is also used to give confidence intervals for produced power at different wind speeds. Figure 5.1 and 5.2 showed that the parameter estimates using synthetic data were relatively accurate for the W2P model parameters $\theta = (a_1, a_2, C_{\max})$ but the standard deviation for the structural error σ_{ξ} was overestimated for the NTF-AW2P model which aligns well with the observation that the variance for the modelled power is underestimated. In regards to the performance of the NTF-AW2P, the problem with the variance should not be too severe since the mean estimate, which is the output of the model, is fairly accurate given the correct θ . Conclusions made for the NTF-AW2P model will take these considerations into account, but it will not be possible to draw a definite conclusion for the model due to the variance estimation issue.

Performance This thesis sought to establish a parametric model for WTG power production based on a physical understanding of the underlying processes using Bayesian methods to model different sources of uncertainty. The intended application was a contract W2P model which is calibrated before installation of a wind farm. The current design regime for W2P models is data-driven where models are fitted based on observations mostly without considering the underlying physical processes. So is it possible to design models outside this regime achieving better performance? The case used in the performance analysis is a setup where it is potentially beneficial to use models that are based on the physical process since these can directly incorporate the site-specific information available through the lookup tables before the wind farm is installed. The AW2P and NTF-AW2P model does this and the performance analysis compares these to the data-driven models. Based on the different error measurements presented in table 6.1 it is safe to conclude that the proposed models do not show any performance gain compared to the baseline models (logistic and bin model). The baseline models generalises better to the test data in the Asia wind farm than the AW2P and NTF-AW2P model across most performance measures which is somewhat surprising. This is not conclusive proof that models based on the physical processes of a WTG using Bayesian methods cannot be used to increased performance over data-driven methods, at least this thesis was not able to show any.

The AW2P and NTF-AW2P models are based on different assumptions about the input uncertainty of the wind speed, and comparing these gives some insight into how to model this. The AW2P model assumes that the measured average wind speed is the mean of the free wind speed in front of the WTG and the NTF-AW2P model omits this assumption and models the relation between free and nacelle wind speed using an NTF. One would expect that the NTF-AW2P model performs better assuming it is based on more realistic assumptions, but comparing the performance of the two models shows almost the same performance across all performance measures, so this might not be the case. One problematic assumption for the NTF model is for the prior information on the NTF-coefficients η which was inferred by comparing results from [43] and [28]. The used here WTG's are of different types and basing the prior information for η on WTG's of the same type and in similar terrain as the modelled ones would be a better practise. Another explanation for similar performance is the AW2P model being similar to the NTF-AW2P when fitted using the same data since they only differ by modelling wind speed through the NTF, which is linear function, thus AW2P is almost equivalent to NTF-AW2P under reparameterisation which is seen in (3.6) and (3.7) when ignoring random effects effects. The issue with variance estimation might also affect the performance, but in any case, it is assessed that no performance gain is found by modelling the NTF.

The developed models are based on several simplifying assumptions and a few approximations - improving any of these might yield better models. A hypothesis based on the standardised residuals in figure 6.8 is that the residuals have some dependence left on the wind speed and modelling this could improve the models. Finding a more appropriate distribution for the power than the Gaussian could also yield an improvement. The histograms in figure 6.9 show that the residuals have heavier tails than a Gaussian distribution with a distinct peak around zero. A distribution like the Laplace distribution which decreases less rapidly from its mean would characterise the residuals better [35,

p. 298]. Modelling the underlying physical processes responsible for power production, with more advanced models accounting for additional effect, is certainly also relevant when discussing improvements - some examples are given in chapter 9.

Power curve uncertainty The different models for this thesis output the total power over 10 minutes, since this is a common setting in the WTG industry for W2P models. As described in section 3.1.3, 10 minutes is specifically used due to the assumption of a spectral gap but modelling the 10-minute power poses some challenges though. For one, it complicates characterising the power distribution as in section 4.4 where the 10 minute-total power is the sum over several random variables, and obtaining an analytical expression for the distribution is not feasible in many cases. Secondly, by modelling the power aggregated over 10 minutes based on a single observation through the average wind speed \bar{v} , the spread of the power distribution becomes quite large. This is seen in the simulated power distributions in figure 4.5 where the distance between the 25th and 75th percentile spans more than 10 kWh at wind speeds above 4 m/s. This observation is based on the assumptions behind the model but the power distribution necessarily has to have a non-zero spread unless it is assumed that the wind speed over 10 minutes is deterministic given \bar{v} . This means that the absolute precision of any model is limited by the variation in the wind - i.e turbulence. Modelling the power over shorter periods would lessen the spread but it may not be desired to model shorter periods depending on the application. As described in section C.1, the models designed in this thesis can be used for AEP calculation where the model has to output 10-minute power since the calculation is based on statistics for 10-minute average wind speeds. In applications where e.g. the wind speed is measured more frequently, it will be possible to increase the accuracy of the model, but these have not been studied here.

BETAE in WTG modelling The AW2P and NTF-AW2P models are based on a deterministic W2P model in (3.7) mapping wind speed to power. The models become stochastic when modelling the input wind speed as a stochastic variable. Chapter 4 showed that obtaining the posterior distribution for the parameters for these models is non-trivial when the input for the models suffers from uncertainty. By using a few approximations for the various distributions it was possible to obtain a low dimensional expression for the posterior making parameter estimation feasible which was the idea of A-BETAE. The use of approximations also has its downsides, so a few points are discussed here.

The wind speed is modelled as a stochastic process described in section 3.1 and the assumptions for the various distributions are made with no regard to the feasibility of integrating out nuisance parameters which necessitates the use of numerical methods to perform the integration. This is a somewhat extremist approach and perhaps a compromise would be better especially when considering that the assumptions of prior distributions for the nuisance parameters are based on limited evidence. A good example of a compromise would be in the conditional distribution for the wind speed:

$$v | \bar{v}, \sigma_v \sim \mathcal{N}(v, \sigma_v^2) \quad \text{and} \quad \sigma_v | \bar{v} \sim \mathcal{N}_+(\mu_{\sigma_v}, \sigma_{\sigma_v}^2),$$

where a folded Gaussian distribution with known parameters $(\mu_{\sigma_v}, \sigma_{\sigma_v}^2)$ is assumed for the standard deviation σ_v . Had it instead been assumed that $1/\sigma_v^2$ followed a gamma distribution with parameters given by a reparametrisation $(\alpha(\mu_{\sigma_v}, \sigma_{\sigma_v}^2), \beta(\mu_{\sigma_v}, \sigma_{\sigma_v}^2))$, the marginal distribution $v | \bar{v}$ has an analytical expression though the t-distribution [41, p. 242].

Another point of critique is the use of a Gaussian distribution to approximate the distribution of power production. As previously mentioned the normalised residuals are characterised better by the Laplace, see appendix B.1, than Gaussian distribution, but from the simulations of synthetic data in figure 4.5 it was also seen that the power distribution becomes skewed for wind speeds close to the cut-in and rated wind speed. Using a distribution with a non-zero skew could yield a more accurate description, although if the skew is to be numerically estimated similar to the mean and variance it would require evaluating a triple sum, which would greatly increase the computational complexity using a numerical approach^[1].

Parameter estimation using Metropolis To round off the discussion, the parameter estimation for the AW2P and NTF-AW2P model using the Metropolis Algorithm is reviewed. Both models had relatively quick convergence using either synthetic or SCADA data. The only exception was for the NTF-AW2P model using SCADA data as shown in figure 6.3 where a burn-in of 10,000 was required, but this was mostly due to poor initialization of the a_1 parameter. Figure 5.3 also showed that the parameters converge to the same value with different Markov chains on the synthetic data although this has not been shown using SCADA data. One issue is that the standard deviation for the structural error $\sigma_{\bar{\zeta}}$ converged to the wrong value using synthetic data as shown in table 5.1. As previously mentioned in the discussion, this issue is possibly linked with the variance estimation error for the modelled power. This has been investigated further and it was found that the posterior approximately has its maximum value in the true parameter values when using synthetic data except for $\sigma_{\bar{\zeta}}$ which is noticeably different - see figure 7.1. The figure also shows that the posterior has the wrong maxima not only for the NTF-AW2P model which is known to have some errors but also for the AW2P model. The bias for the AW2P model with respect to $\sigma_{\bar{\zeta}}$ is certainly lower but it does show that the distribution for power under this model also has some inaccuracies perhaps due to numerical inaccuracies. From this, it is concluded that the issue is not with the Metropolis Algorithm but rather with inaccuracies in the posterior for the two models.

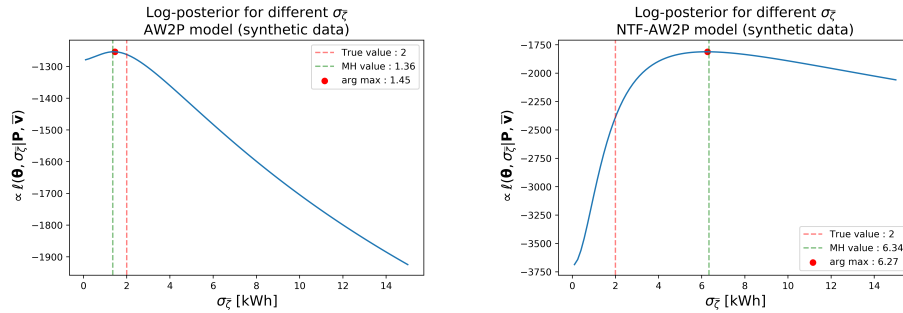


Figure 7.1: Log-posterior for the AW2P model (left) and NTF-AW2P model (right) using synthetic data. The parameter θ is the true value from the synthetic data and is not changed. The log posterior is computed up to proportionality according to (4.26) so the values shown on the ordinate are proportional to the log posterior.

^[1]The skew or skewness of a random variable is defined as the third-order standardised moment - i.e $E[((X - \mu)/\sigma)^3]$ for a random variable with mean μ and standard deviation σ [46, p. 153].

8 | Conclusion

The current design regime for contract W2P models is data-driven where models are fitted based on observations, mostly disregarding the physical processes driving the production of power from WTG's. This thesis has taken another approach by directly modelling the physical processes and using Bayesian methods to obtain a probability distribution of the produced power.

A deterministic W2P model has been designed to take wind speed and air density as input and modelling the power coefficient C_P as a function of wind speed based on some simplifying assumptions of how a WTG controls power generation through pitch control of its blades. Wind speed is modelled as a random variable ignoring temporal dependencies either assuming that the average observed wind speed is the true mean of the free wind speed in front of the WTG or modelling the relation between the free wind speed and the nacelle measured wind speed behind the blades using a nacelle transfer function. Applying the deterministic W2P model on the wind speed distributions and aggregating power production over 10 minutes yields the probabilistic AW2P and NTF-AW2P model. Methods to obtain a posterior distribution for model parameters have been presented, but it is found that this is intractable without using approximate distributions or by including latent parameters in the posterior which greatly increases its dimension. The former option is chosen and the power is approximated using a Gaussian distribution with estimated mean and variance. Here it is concluded that the variance estimate for the NTF-AW2P model suffers from inaccuracies due to an unknown error although that error does not affect the model output which is the estimated mean.

Parameter estimation using the Metropolis Algorithm shows convergence for both models using synthetic data as well as SCADA data. Experiments with synthetic data showed that both models approximately converge to the true parameters in the algorithm except for the structural error standard deviation and it is concluded that this is due to inaccuracies in the estimation of the posterior rather than issues with the Metropolis Algorithm.

In the performance analysis, the AW2P and NTF-AW2P models are evaluated on data from a training set in a South American wind farm and two test sets - one in the same South American wind farm but for different WTG's and one in an Asian wind farm. As a baseline, two existing models, namely a 5-parameter logistic model and a method of bins model, were used as a comparison. The thesis found no gain in performance using the AW2P and NTF-AW2P model compared to the existing models with similar performance for the test set in South America and worse performance for the test set in Asia. The AW2P and NTF-AW2P models showed similar performance compared to each other so modelling the free wind speed through the nacelle transfer function gives no noticeable gain in performance for the models here. It is concluded that the AW2P and NTF-AW2P model are inferior to the existing models with the current implementation as contract W2P models, although the used methods could be applied in more advanced models under different assumptions to try and achieve better performance.

9 | Further Studies

The developed models for this thesis could be improved on numerous fronts. Apart from improving the approximations in the statistical framework, the W2P models were based on several simplifying assumptions about the physical process behind power generations and improving these could potentially yield a gain in performance. A few of these are discussed here.

Terrain How wind flows at a WTG site is affected by the type of terrain like ground inclination, presence of surrounding objects like trees or cliffs, roughness of ground and onshore vs. offshore location etc [33, ch. 8]. In wind resource assessment (WRA) these effects are accounted for using software like the mentioned windPRO. An important factor which depends on terrain is wind speed statistics, as described in appendix C.1. Other effects are wind flow inclination and wind shear.

Wind flow inclination The anemometer placed on a WTG measures horizontal wind speed, but the free wind acting on the blades may also have a vertical component. To correctly model the energy in the wind, the anemometer can be calibrated according to:

$$v_f = v_{nc} \cos(\theta),$$

where θ is the inclination angle, v_f is the resultant wind speed and v_{nc} is the nacelle measured horizontal wind speed [59]. WRA software can estimate the average inclination angle, but the accuracy of the calibrated wind speed measurements is impacted by higher angles [33, p. 86]. Accounting for the impact of average inclination on wind speed statistics could result in a more accurate description of uncertainties in a W2P model.

Wind shear Wind shear is a measure of how wind speed changes for different altitudes. If v_1 and v_2 are wind speeds at heights above ground h_1 and h_2 , their ratio is modelled by the wind shear exponent γ [33, p. 38]:

$$\frac{v_2}{v_1} = \left(\frac{h_2}{h_1} \right)^\gamma$$

Wind shear is affected by the roughness of the terrain and WRA software can compute the expected wind shear exponent. As an example, $\gamma \approx 0.08$ at open sea and $\gamma \approx 0.39$ in a large city with tall buildings [33, p. 40]. Modelling how shear affects the force applied on WTG blades at different heights could improve a W2P model.

Wake effects Wake is the reduction of wind and increased turbulence caused by e.g. the presence of a WTG. In a wind farm, increased turbulence causes greater stress at the WTG structures and lower wind speeds causes decreased energy production in the range of 2 – 20% depending on the relative WTG positions [33, p. 161]. Modelling wake effects in a W2P model based on wind direction could be an improvement.

Bibliography

- [1] Alston L., Clair, Mengersen L., Kerrie, and Pettitt N., Anthony. *Case Studies in Bayesian Statistical Modelling and Analysis*. 1st ed. Wiley, 2012.
- [2] American Meteorological Society. *Convective Turbulence*. URL: http://glossary.ametsoc.org/wiki/Convective_turbulence.
- [3] AskCI Consulting. *Share of energy production in China in 2017, by source*. Visited on 16-12-2019. Dec, 2018. URL: <https://www.statista.com/statistics/950088>.
- [4] *Aviation Weather Student Guide*. URL: <http://navyflightmanuals.tpub.com/P-303/Mechanical-Turbulence-106.htm>.
- [5] Bishop, Christopher M. *Pattern Recognition and Machine Learning*. Springer, 2006. ISBN: 9780387310732.
- [6] CEC. *Electric power generation in China in 2018, by source*. Visited on 11-9-2019. Jan, 2019. URL: <https://www.statista.com/statistics/302233>.
- [7] CEC. *Wind power generation in China in 2016, by region*. Visited on 11-9-2019. Dec, 2017. URL: <https://www.statista.com/statistics/302355>.
- [8] Chen, Hung. *Transformations and Expectations*. Visited on 31-5-2020. URL: http://www.math.ntu.edu.tw/~hchen/teaching/StatInference/notes/lecture7.pdf?fbclid=IwAR2vupVJFVHlKCWpr89RLvQEnpnKsK5hB-ZwqwC77e3_dqJYERZz9oXZR0.
- [9] Chinese Wind Energy Association and Sohu. *Cumulative installed wind power capacity in China in 2018, by major company*. Visited on 11-9-2019. 2019. URL: <https://www.statista.com/statistics/1029752>.
- [10] Chinese Wind Energy Association and Sohu. *Export volume of Goldwind wind power generators from China in 2018, by country*. Visited on 11-9-2019. 2019. URL: <https://www.statista.com/statistics/1029993>.
- [11] Demurtas, Giorgio, Friss, Troels Pedersen, and Rozenn, Wagner. "Nacelle power curve measurement with spinner anemometer and uncertainty evaluation". In: *Wind Energy Science* (2016). DOI: 10.5194/wes-2016-29.
- [12] DSW. *Ten countries with the largest population in mid 2018*. Visited on 11-9-2019. Mar, 2019. URL: <https://www.statista.com/statistics/262879>.
- [13] energy, The big portal for wind. *Goldwind GW 121/2500*. Visited on 21-02-2020. 2018. URL: <https://en.wind-turbine-models.com/turbines/1192-goldwind-gw-121-2500#datasheet>.
- [14] Escalante Soberanis, M.A and Mérida, W. "Regarding the influence of the Van der Hoven spectrum on wind energy applications in the meteorological mesoscale and microscale". eng. In: *Renewable Energy* 81 (2015), pp. 286–292. ISSN: 0960-1481.
- [15] Everitt, B. S. *The Cambridge Dictionary of Statistics*. 2nd ed. Cambridge University Press, 2002.

- [16] Fairley, Peter. "China's Potent Wind Potential". In: *MIT Technology Review* (2009). Visited on 12-19-2019. URL: <https://www.technologyreview.com/s/415311/chinas-potent-wind-potential/>.
- [17] Frandsen, Sten Tronæs. "Turbulence and turbulence-generated structural loading in wind turbine clusters". In: *Risoe-R 1188(EN)* (2007).
- [18] Gao, Fuchang and Han, Lixing. "Implementing the Nelder-Mead simplex algorithm with adaptive parameters". In: *Springer Science* (2010).
- [19] Germanwatch. *Largest producers of territorial fossil fuel CO2 emissions worldwide in 2017, based on their share of global CO2 emissions*. Visited on 16-12-2019. Dec, 2019. URL: <https://www.statista.com/statistics/271748>.
- [20] Global Commission on the Geopolitics of Energy Transformation. *A New World, The Geopolitics of the Energy Transformation*. Tech. rep. 2019. URL: http://geopoliticsofrenewables.org/assets/geopolitics/Reports/wp-content/uploads/2019/01/Global_commission_renewable_energy_2019.pdf.
- [21] Goodfellow, Ian, Bengio, Yoshua, and Courville, Aaron. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [22] Greengard, Samuel. "Digital Twins Grow Up". In: (2019). Visited on 12-19-2019. URL: <https://cacm.acm.org/news/238642-digital-twins-grow-up/fulltext>.
- [23] Grieves, Michael W. "Virtually Intelligent Product Systems: Digital and Physical Twins". In: *Complex Systems Engineering: Theory and Practice* (2019), pp. 175–200.
- [24] Heier, Siegfried. *Grid Integration of Wind Energy*. 5th ed. Wiley, 2014.
- [25] Horrace, C. William. "Moments of the truncated normal distribution". eng. In: (2014). DOI: 10.1007/s11123-013-0381-8.
- [26] Hoven, Isaac Van der. "Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour". In: *Journal of Meteorology* 14 (1956).
- [27] Huard, David and Mailhot, Alain. "A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model "abc"". In: *Water Resources Research* 42 (2006). DOI: 10.1029/2005WR004661.
- [28] Hyeon-Wu, Kim, Kyung-Nam, Ko, and Jong-Chul, Huh. "Atmospheric turbulence affects wind turbine nacelle transfer functions". In: *Journal of the Korean Solar Energy Society* 33.4 (2013).
- [29] I-Scoop. "Digital twins – rise of the digital twin in Industrial IoT and Industry 4.0". In: (). Visited on 12-19-2019. URL: <https://www.i-scoop.eu/internet-of-things-guide/industrial-internet-things-iiot-saving-costs-innovation/digital-twins/>.
- [30] International Electrotechnical Commission. *International Standard, IEC 61400-12-2*. Tech. rep. Valid till 2020. 2013. URL: <https://webstore.iec.ch/publication/5430>.
- [31] IRA. *Distribution of primary energy supply worldwide in 2016, by source*. Visited on 10-9-2019. Sep, 2019. URL: <https://www.statista.com/statistics/270528>.
- [32] IRENA. *Energy production from wind energy in China from 2008 to 2017*. Visited on 11-9-2019. July, 2019. URL: <https://www.statista.com/statistics/224763>.

- [33] Jain, Pramod. *Wind Energy Engineering*. McGraw-Hill Professional, 2011. ISBN: 9780071714778.
- [34] Kavetski, Dmitri, Kuczera, George, and Franks, Stewart W. "Bayesian analysis of input uncertainty in hydrological modelling: 1. Theory". In: *Water Resources Research* 42 (2006). DOI: 10.1029/2005WR004368.
- [35] Kay, Steven. *Intuitive Probability and random processes using Matlab*. 4th ed. Springer, 2006.
- [36] Kitain, Lior. *The New Age of Manufacturing: Digital Twin Technology and IoT*. 2011. URL: <https://www.seebo.com/digital-twin-technology/>.
- [37] Krishna, B. Vinay, Ormel, Frank, and Hansen, Kurt S. "Alternative approach for establishing the Nacelle Transfer Function". In: *Wind Engineering* 40.4 (2016).
- [38] Le Gourières, Désiré. *Wind Power Plants: Theory and Design*. Pergamon Pr, 1982.
- [39] Lydia, M. et al. "A comprehensive review on wind turbine power curve modeling techniques". In: (2013). URL: <https://www.sciencedirect.com/science/article/abs/pii/S1364032113007296>.
- [40] Lydia, M. et al. "Advanced Algorithms for Wind Turbine Power Curve Modeling". In: *IEEE Transactions on Sustainable Energy* 4.3 (2013).
- [41] Madsen, Henrik and Thyregø, Poul. *Introduction to General and Generalized Linear Models*. CRC Press, 2011.
- [42] Marr, Bernard. "What Is Digital Twin Technology - And Why Is It So Important?" In: (). Visited on 12-19-2019. URL: <https://www.forbes.com/sites/bernardmarr/2017/03/06/what-is-digital-twin-technology-and-why-is-it-so-important/#3ed32f3e2e2a>.
- [43] Martin, Clara M. St et al. "Wind Turbine Power Performance Testing using Nacelle Transfer Function". In: *Wind Energy Science* (2016). DOI: 10.5194/wes-2016-45.
- [44] Mihet-Popa, L. and Groza, V. "Dynamic modeling, simulation and control strategies for 2 MW wind generating systems". In: *International Review of Modelling and Simulations* 5.4 (2010).
- [45] NOAA. *Annual anomalies in global land and ocean surface temperature from 1885 to 2018, based on temperature departure (in degrees Celsius)*. Visited on 10-9-2019. July, 2019. URL: <https://www.statista.com/statistics/224893>.
- [46] Olofsson, Peter and Andersson, Mikael. *Probability, Statistics, and Stochastic Processes*. eng. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011. ISBN: 9780470889749.
- [47] *Renewable Electricity Capacity and Generation Statistics*. Visited on 11-9-2019. 2019. URL: <http://resourceirena.irena.org/gateway/dashboard/?topic=4&subTopic=54>.
- [48] Scientific & Technical Terms, McGraw-Hill Dictionary of. *eddy kinetic energy*. Visited on 1-6-2020. URL: <https://encyclopedia2.thefreedictionary.com/eddy+kinetic+energy>.
- [49] SCIPY. *Integration and ODEs*. URL: <https://docs.scipy.org/doc/scipy/reference/integrate.html>.
- [50] SCIPY. *Optimization and Root Finding*. URL: <https://docs.scipy.org/doc/scipy/reference/optimize.html>.

- [51] Scott, David W. *Multivariate Density Estimation : Theory, Practice, and Visualization*. 2nd ed. John Wiley and Sons, Incorporated, 2015.
- [52] Shin, Dongheon, Hyeonwu, Kim, and Kyungnam, Ko. "Analysis of wind turbine degradation via the nacelle transfer function". In: *Journal of Mechanical Science and Technology* 29.9 (2015). DOI: 10.1007/s12206-015-0846-y.
- [53] Shumway, H. Robert and Stoffer, S. David. *Time Series Analysis and Its Application with R Examples*. 4th ed. Springer, 2017.
- [54] Stanway, David. "China to end subsidies for onshore wind power by 2021". In: *Sustainable Business* (2019). Visited on 12-19-2019. URL: <https://www.reuters.com/article/us-china-windpower/china-to-end-subsidies-for-onshore-wind-power-by-2021-idUSKCN1SU0M1>.
- [55] Statnett. *Slik fungerer kraftsystemet*. Visited on 12-19-2019. 2018. URL: <https://www.statnett.no/om-statnett/bli-bedre-kjent-med-statnett/slik-fungerer-kraftsystemet/>.
- [56] Surya, Tokdar T. *STA 114: Statistics - Notes 12. The Jeffreys Prior*. Visited on 30-05-2020. URL: <https://www2.stat.duke.edu/courses/Fall11/sta114/jeffreys.pdf>.
- [57] The welding institute. *Digital Twin Technology to monitor offshore assets*. Visited on 12-19-2019. 2017. URL: <https://www.twi-global.com/media-and-events/press-releases/2017-03-twi-embarks-on-lifecycle-engineering-asset-management-through-digital-twin-technology>.
- [58] Tuegel, Eric J., Ingrassia Anthony R. Eason, Thomas G., and Spottswood, S. Michael. "Reengineering Aircraft Structural Life Prediction Using a Digital Twin". In: *International Journal of Aerospace Engineering* (2011). DOI: 10.1155/2011/154798.
- [59] Vahidzadeh, Mohsen and Markfort, Corey. "Modified Power Curves for Prediction of Power Output of Wind Farms". In: *Energies* 12 (May 2019), p. 1805. DOI: 10.3390/en12091805.
- [60] *windPRO*. EMD International A/S. URL: <https://www.emd.dk/windpro/>.
- [61] Zhou, Feihang and Liu, Jun. "Pitch Controller Design of Wind Turbine Based on Nonlinear PI/PD Control". In: *Shock and Vibration* (2018). DOI: 10.1155/2018/7859510.

A | List of abbreviations

WTG	Wind turbine generator
W	Watt
Wh	Watt hours
kWh	Kilo Watt hours
WF	Wind farm
W2P	Wind to power (model)
WTPC	Wind turbine power curve
AEP	Annual energy production
SCADA	Supervisory Control and Data Acquisition
WRA	Wind resource assessment
pdf	Probability density function
C_P	Power coefficient
NTF	Nacelle transfer function
TSR	Tip speed ratio
PID	Proportional Integral Derivative
AW2P	Aggregated wind to power
std	Standard deviation
BETAE	Bayesian total error analysis
A-BETAE	Approximate Bayesian total error analysis
MCMC	Markov Chain Monte Carlo
Q-Q	Quantile-Quantile

B | Mathematical definitions and results

B.1 List of distributions

A few distributions are used through the thesis. Their probability density functions are listed here.

Let X be a real valued continuous random variable with range R of dimension N and pdf p_X . The thesis uses the following distributions:

- **Gaussian distribution:**

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

with $R = \mathbb{R}$ parametrised by mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}_+$, written $X \sim \mathcal{N}(\mu, \sigma^2)$.

- **Folded Gaussian distribution:**

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \left(\exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) + \exp\left(-\frac{1}{2\sigma^2}(x + \mu)^2\right) \right)$$

with $R = [0, \infty)$ parametrised by $\mu \in \mathbb{R}_+$ and $\sigma^2 \in \mathbb{R}_+$, written $X \sim \mathcal{N}_+(\mu, \sigma^2)$.

- **Multivariate Gaussian distribution:**

$$p_X(x) = \frac{1}{\sqrt{(2\pi)^N |\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

with $R = \mathbb{R}^N$ parametrised by mean $\boldsymbol{\mu} \in \mathbb{R}^N$ and $\Sigma \in \mathbb{R}^{N \times N}$ is the symmetric, positive semi-definite variance-covariance matrix, written $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

- **Uniform distribution**

$$p_X(x) = \frac{1}{b - a}$$

with $R = [a, b]$ parametrised by the $a \in \mathbb{R}$ and $b \in \mathbb{R}$ for $a < b$, written $X \sim \mathcal{U}(a, b)$.

- **Laplace distribution**

$$p_X(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}}{\sigma}|x - \mu|\right)$$

with $R = \mathbb{R}$ parametrised by the mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}_+$, written $X \sim \mathcal{L}(\mu, \sigma)$.

B.2 Results and definitions

Definition B.2.1 (Observed and expected information) [41, p. 18] Let $p_Y(\mathbf{y}|\theta)$ the likelihood for the stochastic variable \mathbf{Y} parametrised by real valued $\theta \in \mathbb{R}^m$ and $\ell_Y(\mathbf{y}|\theta) = \log p_Y(\mathbf{y}|\theta)$ the log likelihood. The observed information corresponding to \mathbf{y} and θ is:

$$j(\theta; \mathbf{y}) = -\frac{\partial^2}{\partial \theta \partial \theta^T} \ell(\theta|\mathbf{y}),$$

and the expected information corresponding to θ and \mathbf{Y} is:

$$i(\theta) = \mathbb{E}(j(\theta|\mathbf{Y}))$$

Lemma B.2.1 [41, p. 19] Let Y be a stochastic variable with likelihood $\ell_\theta(\theta; Y)$, for reasonable well behave likelihood the following holds:

$$\mathbb{E} \left[-\frac{d^2}{d^2\theta} \ell(\theta; Y) \right] = \mathbb{E} \left[\left(\frac{d}{d\theta} \ell(\theta; Y) \right)^2 \right]$$

Definition B.2.2 (ℓ_2 -norm) The ℓ_2 -norm is the function $\|\cdot\| : \mathbb{R}^N \rightarrow \mathbb{R}_+$ is defined as:

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^N x_i^2}$$

where $\mathbf{x} \in \mathbb{R}^N$ is some real valued vector.

Definition B.2.3 (Indicator function) Let $x \in X$ and $A \subseteq X$. The indicator function $\mathbf{1}_A : X \rightarrow \{0, 1\}$ is defined as:

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Proposition B.2.1 (Change of variable) [46, p. 94] Let X be a continuous random variable with density p_X and let g be a strictly increasing or decreasing differentiable function, further let $Y = g(X)$. Then Y has pdf

$$p_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| p_X(g^{-1}(y)) = \left| \frac{dx}{dy} \right| p_X(x)$$

Corollary B.2.1 Let X be a continuous random variable with pdf p_X with and let $Y = aX + b$ with $a \neq 0$. Then Y has pdf:

$$p_Y(y) = \frac{1}{|a|} p_X \left(\frac{y-b}{a} \right)$$

Proof Let $g(x) = ax + b$ which is differentiable and either strictly increasing or decreasing depending on a when $a \neq 0$. Using proposition B.2.1 and $g^{-1}(y) = (y-b)/a$ it follows immediately that:

$$p_Y(y) = \left| \frac{d}{dy} \frac{y-b}{a} \right| p_X \left(\frac{y-b}{a} \right) = \frac{1}{|a|} p_X \left(\frac{y-b}{a} \right) \quad \blacksquare$$

Scott's rule Choosing bandwidth h for a kernel density estimator, the heuristic known as *Scott's rule* can be used which uses [51, p. 164]:

$$h \approx \hat{\sigma}_X \cdot (N_{\text{sim}})^{-(D+4)},$$

where $\hat{\sigma}_X$ is the estimation standard deviation for an observed data set \mathbf{X} . Choosing bandwidth according to Scott's rule minimizes the according to certain measures when the underlying process is uncorrelated multivariate Gaussian [51, p. 163].

Normal Equations [41, p. 48]. Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{y} \in \mathbb{R}^n$. Then $\mathbf{x} \in \mathbb{R}^m$ is the least squares solution to $\mathbf{A}\mathbf{x} = \mathbf{y}$ if and only if \mathbf{x} is a solution to the normal equations:

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{y}$$

If $\text{rank}(\mathbf{A}) \geq m$ then the solution is given uniquely by:

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$$

Power spectral density Let $X[n]$ be a time discrete stochastic process with $n \in \mathbb{Z}$. If $X[n]$ is weak sense stationary defined as having constant mean

$$\mu_X[n] = \mathbb{E}[X[n]] = \mu \quad \forall n \in \mathbb{Z},$$

and autocovariance function (ACF) only dependent on lag:

$$r_X[n, n+k] = \mathbb{E}[X[n]X[n+k]] = r_X[k] \quad \forall n, k \in \mathbb{Z},$$

then the power spectral density (PSD) is defined as the discrete time Fourier transform of the ACF [35, p. 571]:

$$P_X(f) = \sum_{k \in \mathbb{Z}} r_X[k] \exp(-j2\pi f k),$$

if it exists.

Transformation of Random variable For some random variable $X \sim p_X$ it is given by definition that:

$$P(a \leq X \leq b) = \int_a^b p_X(x) dx$$

if a monotone (increasing) transformation $Y = g(X)$ is applied, then the following holds [8]:

$$\int_a^b p_X(x) dx = \int_{g(a)}^{g(b)} p_Y(y) dy, \quad (\text{B.1})$$

where $p_Y(y)$ is given by the change of variable theorem B.2.1.

C | WTG engineering: Additional theory and results

C.1 Annual Energy Production and wind speed statistics

AEP is the estimated the amount of energy produced over an entire year for a WTG or WF and is an important measure in wind resource assessment. Consider the case of estimating AEP for a single WTG with known power curve given by a W2P model with parameters θ . Let $v \in \mathbb{R}^N$ be a long-term forecast of wind speed over a year with N being the number of samples in a year for a given resolution of v . Estimated is AEP then simply:

$$\text{AEP} = \sum_{n=1}^N \text{W2P}(v_n; \theta). \quad (\text{C.1})$$

Forecasts of wind speed including uncertainty measures can be obtained by modelling the average wind speed short intervals as a random variable. In the wind turbine industry the convention is to model 10-minute average wind speeds as Weibull distributed [33, p. 32].

Definition C.1.1 (Weibull distribution) [33, p. 30] *If a random variable T has pdf on*

$$f(t) = \frac{\alpha}{\lambda} \left(\frac{t}{\lambda} \right)^{\alpha-1} \exp \left[- \left(\frac{t}{\lambda} \right)^{\alpha} \right], \quad \text{for } t \in [0, \infty)$$

then T is said to follow a two-parameter^[1] Weibull distribution with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$, written $T \sim \mathcal{W}(\alpha, \lambda)$.

Remark C.1 The Weibull distribution has mean

$$\mu = \lambda \Gamma(1 + 1/\alpha)$$

and variance

$$\sigma^2 = \lambda^2 \left(\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha) \right)$$

where Γ is the Gamma function [33, p. 33].

Remark C.2 For $\alpha = 1$ (C.1.1) reduces to the pdf of a exponential distribution and for $\alpha = 2$ it is equivalent to the pdf of a Rayleigh distribution [35, p. 302].

Modelling wind speed as Weibull distributed, $v \sim \mathcal{W}(\alpha, \lambda)$, the shape parameter is unitless and the scale parameter λ is measured in m/s. According to [33] modelling wind speed as a Weibull distribution tends to be a good assumption for many locations and as a

^[1]A three-parameter Weibull distribution exists with an additional parameter taking location into account, however only the 2 parameter version is used when dealing with wind profiles.

rule of thumb $\alpha \approx 2$. Furthermore, when modelling minute 10 average wind speed as a stochastic process v_n , it can be assumed that subsequent samples are uncorrelated - see further description in section 3.1.3. Illustrations of the Weibull distributions for different α and λ parameters can be seen in figure C.1.

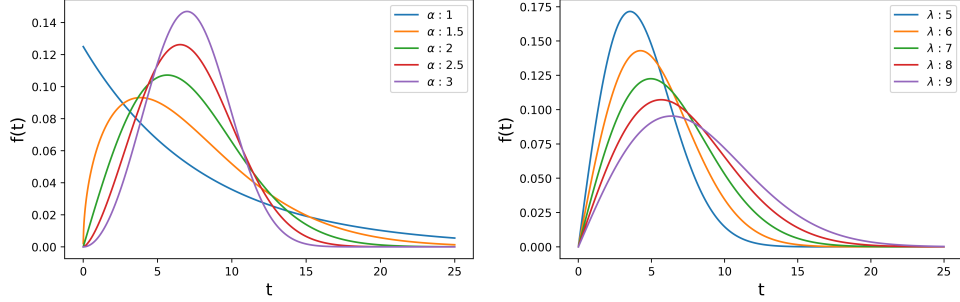


Figure C.1: (Left) figure shows a Weibull pdf with $\lambda = 8$ and varying scale parameter α . Similarly, the (Right) figure shows the same for $\alpha = 2$ and varying λ .

In regards to obtaining wind speed forecasts for AEP estimation, WRA software like the mentioned windPRO, is used to obtain (α, λ) in different wind direction sectors along with a frequency parameter f specifying what percentage of time the wind direction is expected to fall in each sector [33, ch. 7]. Given $(\alpha, \lambda, f)_i$ for each wind direction sector i , mean AEP and associated uncertainty can be obtained using (C.1).

The Weibull distribution is fitted on the available data to see how well the assumption applies for the two wind farms. Closed-form solutions for the maximum likelihood estimate for (α, λ) is not found in the literature, however the following approximations can be used [33, p. 34]:

$$\hat{\alpha} = \left(\frac{\sigma}{\mu} \right)^{-1.086} \quad \text{and} \quad (C.2)$$

$$\hat{\lambda} = \frac{\mu}{\Gamma \left(1 + \frac{1}{\hat{\alpha}} \right)}, \quad (C.3)$$

where μ and σ is the mean and standard deviation which is estimated for a given data set. An illustration of a Weibull distribution fitted using (C.2) and (C.3) can be seen in figure C.2.

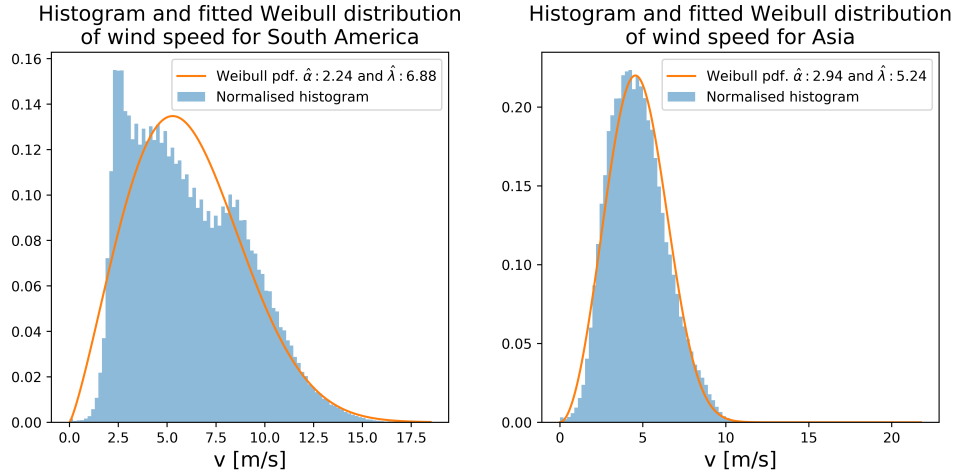


Figure C.2: Histogram for wind speed data and a fitted Weibull pdf. The (left) figure shows data for the South American wind farm and the (right) for the Asian wind farm.

C.2 Estimation of v_1 and v_2

v_1 and v_2 are estimated to be the values that minimises the sum of squared error (SSE) between estimated and observed power. Let $\mathbf{P} \in \mathbb{R}^{N_{\text{dat}}}$ be observed 10 minute total power and $\bar{\mathbf{v}} \in \mathbb{R}^{N_{\text{dat}}}$ 10 minute average wind speed for N_{dat} data points. The function for the power coefficient in (3.6) states that

$$v_c = 2.8 \text{ m/s} \leq v_1 \leq v_2 \leq v_r = 9.3 \text{ m/s}$$

so v_1 and v_2 are simply found by finding the point with lowest SSE in an appropriate grid. The SSE is obtained in the following steps:

1. Chose v_1 and v_2 at a point in the grid.
2. Calculate $\theta = (a_1, a_2, C_{\text{max}})$ as the least square estimate by the same method used in section 5.2 to initialise θ for the Metropolis Algorithm.
3. Estimate $\hat{P}_n = \mu_{\hat{P}}(\bar{v}_n, \theta)$ for $n = 1, \dots, N_{\text{dat}}$ using the AW2P or NTF-AW2P model. Note that $\mu_{\hat{P}}(\bar{v}_n, \theta)$ depends on v_1 and v_2 despite being omitted from notation.
4. Set $\text{SSE} = \|\hat{\mathbf{P}} - \mathbf{P}\|^2$.

This have been performed using a subset of the training data described in section 6.1 modelling power is modelled according to the AW2P model. Figure C.3 and C.4 shows the results.

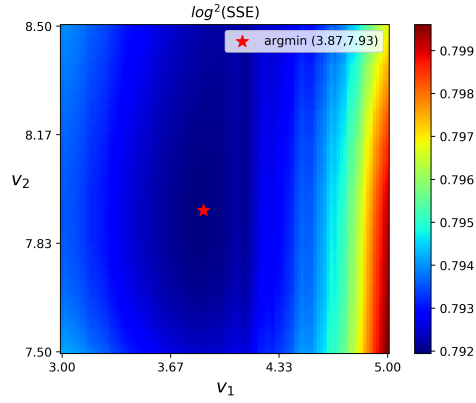


Figure C.3: SSE for different values of v_1 and v_2 showed as the 2nd nested log $\log^2(\text{SSE}) = \log(\log(\text{SSE}))$ to emphasise small differences. $v_1 = 3.87$ and $v_2 = 7.93$ have the lowest SSE.

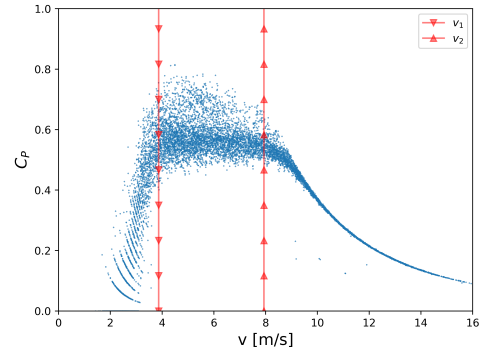


Figure C.4: Plot of power coefficient $C_p = P/P_{\text{wind}}$ as defined in (3.5) for the training data. The estimated v_1 and v_2 are shown in the plot.

D | Software

The relevant scripts developed during the project period has been attached with the hand-in of the project. The scripts are all made in Python 3.7.

The available scripts include:

- `sim_data.py` is used to generate synthetic data for the AW2P and NTF-AW2P model.
- `likelihood_stats.py` is used for calculating the moments see (4.19),(4.20) and (4.21).
- `metropolis.py` is used to perform parameter inference through the Metropolis Algorithm see algorithm 2.

References to data have been removed. Note that the scripts cannot be run as they rely on confidential data which is not included.