Using Machine Learning Methods for Wind Turbine Power Curve Modeling

by

Mehrdad Mehrjoo

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg

Abstract

Wind turbine power curve modeling plays an important role in wind energy management. Accurate estimation of power curves can help reducing power systems maintenance costs. In this thesis, we use machine learning techniques such as clustering, spline regression as well as statistical learning approaches such as multilevel modeling and isotonic regression to reduce bias and/or variance of fitted power curves to improve their performance. First, we focus on reducing the effect of outliers in the wind speed-power data. To this end, we propose to enforce the inherent property of manufacturer power curve on fitted power curves to reduce outliers' impact. Manufacturer power curve is a worthy source of information about turbines' performance which has been ignored in the literature. So, we propose two nonparametric techniques based on the tilting method and monotonic spline regression methodology to preserve monotonicity on fitted power curves according to manufacturer power curve. Another challenging issue in fitting empirical power curve which was investigated seldom in the literature is heteroscedasticity of the wind speed-power data set. Age of turbine, location, air density, wind direction, and measurement errors are some of the reasons which may cause heteroscedasticity or non-homogeneity among observed data. To overcome this problem, we propose a novel methodology to use a hybrid estimation approach based on weighted balanced loss functions that account for both estimation error and goodness of fit by shrinking estimates toward standardized target models. Our proposed approach is very

general and can be used with any desirable weighting scheme as an effective tool to improve the performance of existing power curve modeling approaches. Finally, we investigate improving wind farm aggregated power curve modeling instead of fitting power curves for individual turbines to reduce the complexity of wind farm management analyses. To solve this issue, we propose a novel clustering feature set based on turbines' overall performance and utilize K-Means clustering to classify turbines into homogeneous groups accordingly. We then apply multilevel modeling methods, including random intercept and random slope models on turbine clusters, to consider the hidden correlation among different clusters. We show that the proposed method is a solution for handling the complexity-accuracy trade-off issue since its accuracy is significantly higher than the single aggregated method alongside an equal complexity.

Acknowledgment Page

I would like to thank my academic advisors Dr. Miroslaw Pawlak and Dr. Mohammad Jafari Jozani who jointly supervised me during my PhD research.

I would like to express my appreciation to Dr. Miroslaw Pawlak for his advice and assistance during this research.

I am indebted to Dr. Mohammad Jafari Jozani for all his thoughtful academic guidance, positive encouragement, and friendship throughout the duration of my research. He has been supportive since the day I met him. He taught me a lot of lessons not only as a mentor but also as an honest friend. I owe him a lot not only because of his academic assistance but also because of how he made me a better person after collaborating with him.

I wish to extend my thanks to my PhD examining committee: Dr. Bagen Bagen, Dr. Pradeepa Yahampath, and Dr. Aerambamoorthy Thavaneswaran for their valuable comments and efforts in the evaluation of this work.

I would also like to acknowledge the Natural Sciences and Engineering Research Council (NSERC) of Canada, Mathematics of Information Technology and Complex Systems (MITACS), Manitoba Hydro Ltd, and the Faculty of Graduate Studies of the University of Manitoba for providing financial support for this work.

Finally, I wish to extend my deepest thankfulness to my family, especially to my dear mother "Khadijeh" and my dear brothers, "Saeed" and "Vahid", for their love and support, and my uncle "Ahmad" who has been a true mentor in my life. I also appreciate all my friends for their unconditional support.

To my wonderful wife, “Farnaz”, for all her love and endless support!

And in bright memory of my beloved father “Mohammad”.

Acronyms

IEC International Electrotechnical Commission.

LR Local Regression.

MAE Mean Absolute Error.

MPC Manufacturer's Power Curve.

NMAPE Normalized Mean Absolute Percentage Error.

NS Natural Spline.

PR Polynomial Regression.

RMSE Root Mean Squared Error.

sMAPE Symmetric Mean Absolute Percentage Error.

WFAPC Wind Farm Average Power Curve.

WPP Wind Power Plant.

WTPC Wind Turbine Power Curve.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

With the recent improvement of wind turbine technology and the presence of abundant wind resources in different countries, the number of installed wind farms for power generation has increased significantly. Although wind energy is one of the most promising renewable energy resources, there exist many challenges in wind farm managing. For example a major challenge is its high maintenance and operation cost, which accounts for 20-25% of the total lifetime costs of an offshore wind farm (Hassan, 2013). Another challenge is accurate power forecasting, which plays an important role in power system reliability and marketing. Also, monitoring the wind farm condition for detecting faults in the early stages facilitates wind farm management. Efficient estimations of wind turbine and wind farm power curves play an important role in resolving some of the aforementioned issues and perform better wind farm management. To this end, in this thesis, we develop methodologies based on machine learning techniques to improve the accuracy of fitted wind turbine power curves as well as reduce the impact of outliers. In addition, we study the heteroscedasticity property of wind farm data sets and propose hybrid methods to reduce the effect of heteroscedasticity on the wind turbine power curve fitting. Besides, we study wind farm aggregated models, which are useful in predicting annual power output. A novel methodology is proposed to reduce the complexity of the aggregated model without scarifying accuracy.

In this chapter, we provide a short introduction to preliminary concepts and outline the thesis objectives and describe our research problems. To this end, Section 1.1 explains the recent growth of the wind power industry. Section 1.2 presents the challenges regarding wind farm and wind power prediction. Section 1.3 provides an overview of different wind turbine power curves and various procedures to calculate them. Section 1.4 explains the need for an empirical power curve and its applications. Section 1.5 discusses different commonly used empirical power curve modeling techniques in the literature and defines statistical metrics to evaluate their accuracy. Section 1.6 explains the impact of outliers on the power curve fitting models and some outlier detection techniques available in the literature. Section 1.7 provides an overview of the heteroscedasticity property of the wind farm data sets and the uncertainty it causes in fitting empirical power curves. Section 1.8 describes wind farm aggregated models and the impact of the wake effect on wind farms. Section 1.9 describes the problems this thesis attempts to address and lists the thesis objectives. Finally, Section 1.10 provides an overview of the thesis, briefly describing the content of each chapter.

1.1 Wind Power

Due to the future shortage of fossil fuel and its pollution, clean and renewable energy sources are required to be used instead of fossil fuel. Clean energy may promote the environment and human lives by reducing the use of fossil fuels. Wind energy is one of the fastest-growing beneficial renewable sources of energy. In the last decade, the number of wind farms is increased significantly. According to Statistics Canada, more than 4% of electricity generation in Canada comes from wind energy (Banitalebi et al., 2020). Canada is an ideally suited country to capitalize on large amounts of wind energy due to its favorite environmental conditions. In 2018, Ontario and Quebec were two provinces with the most wind energy capacity, with 5,076 MW and 3,882 MW of power. Fig. (1.1) shows the wind energy capacity of each province in Canada. According to the Canadian Wind Energy Association (CWEA), among any form of renewable electricity generation, wind energy has created the most electricity in Canada between 2009 and 2019. In fact, 301 wind farms, operating from coast to coast, provide electricity for over three million Canadian homes. Despite this progress, there are critical obstacles in providing electricity from wind energy and still controlling power system in a wind farm is a complex and challenging task due to the fluctuation of wind and its randomness. Accurate wind information helps to better manage wind farms, control

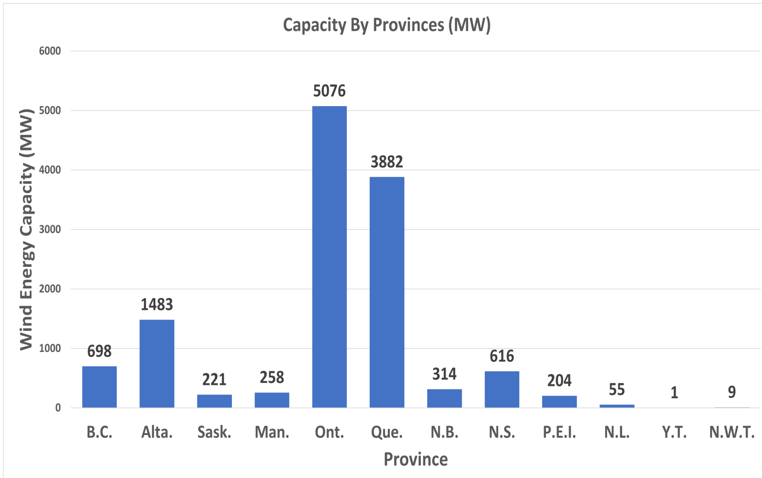


Figure 1.1: Wind energy capacity of each province in Canada by 2018. Source: https://www.nrcan.gc.ca/science-data/data-analysis/energy-data-analysis/energy-facts/renewable-energy-facts/20069

system operation, and perform more accurate fault detection. Using current and past wind information, one may utilize different wind power prediction methods as a practical approach for turbine fault detection, power system simulation, load analysis, and wind farms management and maintenance (Wang et al., 2020; Zhou et al., 2019; Wang et al., 2019; Nabat et al., 2020; Lydia et al., 2014).

Wind farms consist of numerous and complicated systems. Different factors affect wind farm performance. In recent years numerous studies pertinent to wind farm performance monitoring have been done using different computational and/or machine learning techniques. For an overview of such methods, see Wang et al. (2020).

1.2 Wind Farm Challenges

The traditional concept of wind power systems is that wind power systems consist of a few large wind power plants. Recently, this has been substituted with a new wind power system concept, which includes distributed small or medium wind power plants (Association et al., 2010). This new power system concept leads to more penetration of wind power into the power networks. However, one of the challenges in integrating wind energy into the power network is its randomness and wind speed fluctuation, which reduces the operating system's stability, reliability, and power quality (Tong, 2010).

Accurate wind speed and power prediction is required to increase the reliability of wind power systems and help power system management by giving clear and precise information about the future (e.g., available power, fluctuation of power and voltage, etc.) Wind power prediction may help in decisions such as connecting a load, utilizing battery storages, changing blade distance, and other control behaviors (Lange and Focken, 2006). Besides, for electric power companies and the marketing team, it is vital to extract the maximum available energy from wind speed and have a precise prediction of available wind energy provided to customers and clients. This can help companies to perform better management decisions. For example, their maintenance can be done on days with the least expected available wind energy (Li and Shi, 2010).

Below, we provide a short list of challenges and objectives pertinent to wind farm managing that highly depend on wind power forecasting accuracy:

- The quality of power supply must be sufficient with an acceptable level of reliability and stability (Marinelli, 2011).
- Energy distribution companies should be able to forecast available wind power energy accurately for future days to avoid the shortage of power for their clients (Billinton et al., 2006).
- Evaluating available power for the long term (e.g., next year) can help managers and marketing teams in their decision making and cost analysis (Jónsson et al., 2010).
- Accurate wind speed and power prediction are vital for energy companies to choose the location of new wind farms as well as select wind turbines' models (Oh et al., 2012).
- According to the high cost of battery storage, deciding on the sufficient storage capacity is another challenge which wind farm managers have to deal with (Shokrzadeh, 2014).
- Turbines maintenance should be done frequently, and managers prefer to maintain turbines when the available wind energy is low, so that all turbines can generate their maximum power when wind speed is high. Otherwise, the wind farm will not perform most efficiently (Besnard et al., 2009).

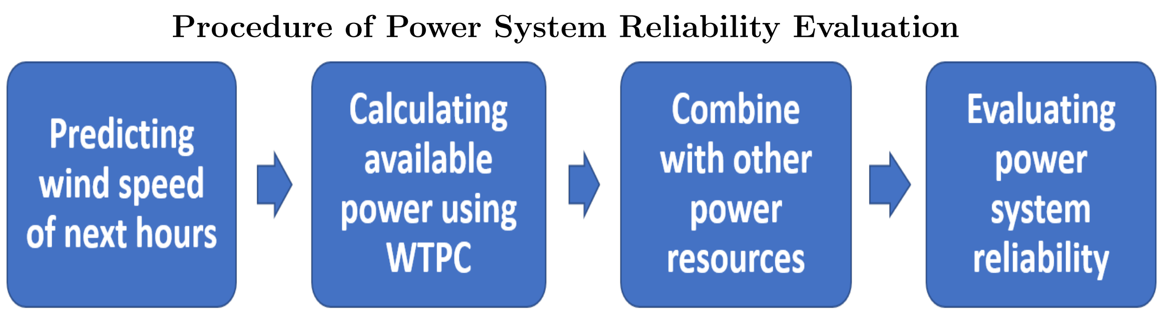


Figure 1.2: Four general phases of a power system reliability analysis. Phase 1: predicts the future wind speed. Phase 2: Estimate wind power using a power curve model. Phase 3: Combine with other power resources. Phase 4: Evaluate power system reliability.

- Wind farm condition monitoring and early fault detection can facilitate maintaining the wind farm. Identifying the turbines that are not working in normal conditions in the early stages can reduce the repair cost significantly (Pandit and Infield, 2018).

The solution to most of the aforementioned challenges depends on the accurate estimation of wind turbine power curves. These curves are often used for the operational management of wind farms, performance monitoring of turbines, and effective wind energy utilization into the power systems (Lydia et al., 2014).

1.3 Wind Turbine Power Curves

A power system reliability analysis can be considered as a procedure with four general phases as depicted in Fig (1.2). In the first phase, one needs to predict the future wind speed at the location of wind farms. Either physical or statistical approaches are utilized for this purpose. In the physical approach, all the details about the physical description of the wind farm site and its surrounding are required to model the on-site condition (Lange and Focken, 2006). In contrast to the physical approach, the statistical approach does work solely based on the historical data of wind speed and corresponding generated power from the wind farm. This approach, which is inexpensive compared to the physical approach, builds statistical models based on historical data. However, this approach needs an acceptable amount of historical data to train efficient statistical models. Some of the most commonly used classical wind speed forecasting methods include time series techniques such as Autoregressive (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA), and the Box-Jenkins approach (Akaike, 1969; Box et al., 2015). These models are easy to formulate and are capable of providing timely forecasts. Another approach is based on Artificial Neural Networks (ANN) and some of its variants such as Particle Swarm Optimization - Artificial Neural Networks (PSO-ANN), Modified Hybrid Neural Network, Complex-Valued Neural Network, and Adaptive Wavelet Neural Network (AWNN) (Amjady et al., 2011; Kitajima and Yasuno, 2010; Bhaskar and Singh, 2012). In addition to these methods, there exist many more machine learning algorithms for time series prediction e.g., K-nearest neighbor regression method (KNN) (Yesilbudak et al., 2013), CART (Classification & Regression Trees)(Lee et al., 2006), and support vector machine (SVM) (Marvuglia and Messineo, 2012; De Gooijer and Hyndman, 2006; Taieb et al., 2012; Jung and Broadwater, 2014).

In the second phase of analyzing power system reliability, one estimates the generating power by turbines based on the predicted wind speed at the wind farms' location. To consider the turbines' performance, an accurate estimation of the empirical wind turbine power curve is required. Since many different factors, such as the turbine's model and age, wind farm location, and its surrounding obstacles impact the wind turbine's performance, efficient methods of fitting power curve models are necessary.

In the third phase, the predicted wind power would be combined with other power resources such solar power. Finally, in the last phase of power system reliability analysis, total amount of predicted power would be compared with the predicted customer load. Note that phase 1 and phase 2 are combined into one phase in some studies, and wind power is predicted directly using historical data.

The main focus of this thesis is on phase 2 of the power system reliability studies. To this end, we have developed new techniques to provide more efficient empirical power curves that can be used for wind farm management and other industrial applications. The wind turbine power curve shows the wind turbine's electrical power output for different wind speed values (Gasch and Twele, 2011). Cut-in (V_c), rated (V_r), and cut-out (V_s) are the three main wind speed values of a wind turbine power curve. The cut-in point shows the wind speed threshold value that wind turbines start generating power as wind speed reaches this threshold. Between the cut-in and rated speeds, typical wind turbines generate power increasingly by increasing wind speed. Wind turbines are supposed to generate their rated or maximum capacity in the range between V_r and V_s. Wind turbines are often shut down to avoid any

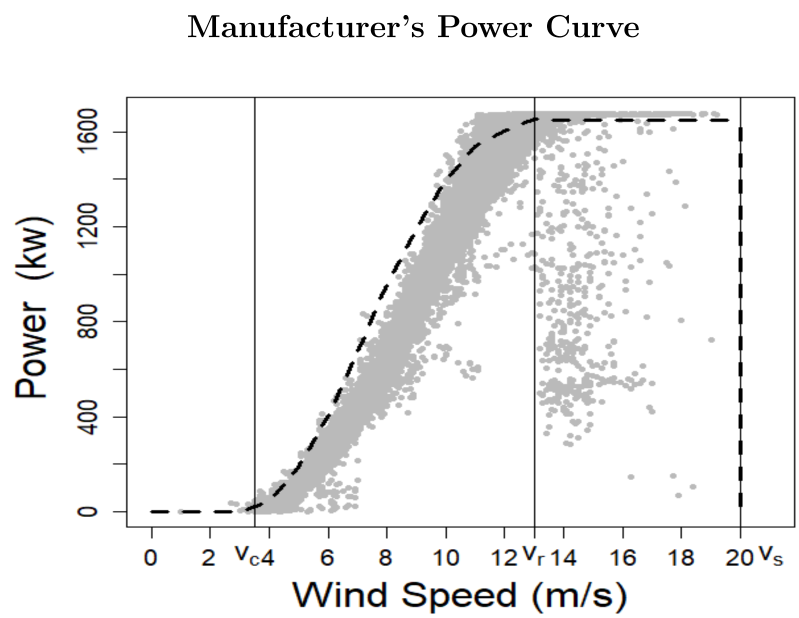


Figure 1.3: Manufacturer's power curve along with the cut-in (V_c), rated (V_r) and cut-out (V_s) points.

damage when wind speed is higher than V_s. Fig. 1.3 represents cut-in, rated, and cut-out points for a sample wind turbine alongside the manufacturer's power curve. We explain the theoretical, manufacturer, and empirical power curves in more details in the following sections.

1.3.1 Theoretical Power Curve

To better manage and maintain wind farms, it is necessary to have a clear idea about how a wind turbine produces energy and how the wind, as the primary resource of power in a wind farm, behaves (Marinelli, 2011). In fact, a wind turbine converts some portion of available energy in the moving air mass to rotational kinetic energy. Then, a generator will produce electrical energy from the rotational kinetic energy. The relation between available energy in the wind and produced electrical energy is not linear. Besides, some other constraints apply to the relationship such that there exist minimum and maximum threshold wind speed for the generator to operate (Walker and Jenkins, 1997).

Different factors, including wind speed, time, weather, location, and obstacle near the wind farm, impact the wind turbine's power output. The following formula shows the generated power from the mass flow rate of air

$$P_w = \frac{1}{2}\eta C_p \rho A v^3, \tag{1.1}$$

where P_w is the generated electrical power in $(Watt)$, η represents the overall efficiency of the turbine, C_p is the dimensionless power coefficient, ρ is the air density in kg/m^3, A is the turbine rotor area in m^2, and v is the wind speed in m/s (Walker and Jenkins, 1997; Ackermann and Söder, 2000). C_p coefficient, which shows the theoretical amount of mechanical power that can be extracted by the turbine rotor, is a function of the turbine blade pitch angle and blade tip speed ratio (Ackermann, 2005). Moreover, the so-called Betz limit explains that the maximum theoretical mechanical power of wind turbines is 0.5926 (Betz, 1920). This formula clearly explains that there exists a nonlinear relationship between wind power and wind speed. Nonetheless, it is an oversimplification to consider that wind turbine output would follow the theoretical power curve in reality. So, wind turbine manufacturers release a manufacturer's power curve alongside the wind turbine to represent a more realistic power curve.

1.3.2 Manufacturer Power Curve

The classic design of wind turbines in the 1990s was such that it would operate at constant speed even in different wind speeds (Marinelli, 2011). Some advantages of this design were mechanical simplicity, robustness, and low maintenance cost. On the other hand, this type of design has some disadvantages (e.g., it draws more reactive power by producing more active power). In new wind turbine designs, rotors may rotate at variable speed to achieve maximum efficiency. New wind turbines' mechanical designs are more complicated than constant speed turbines, but they help capture more power energy. New designs also improve power quality and reduce mechanical stress on the turbine.

Even in the new designs, wind turbines have a maximum limit on generating power in the wind speed range between rated to cut-out points, and they can not produce more than rated power in this range of wind speed. This limitation is handled by the reduction of conversion efficiency from wind energy to rotational kinetic energy. Pitch regulation and stall regulation are two ways that blades can be designed to reduce the conversion efficiency (Association et al., 2012).

Pitch regulated systems have an active control system that controls the pitch angle around its axis to reduce the rotor speed, keeping the rotor speed at a constant level when wind speed reaches the rated point.

On the other hand, in a stall regulated system, the rotor speed and the produced power will decrease by incrementing wind speed after reaching a rated threshold until it reaches a cut-out point. In stall regulated machines, blade designs are such that they perform worse when wind speed is high, to avoid any damage to the turbine, without requiring any active control system (Ackermann, 2005).

In both designs, wind turbine manufacturers provide specifications of their turbine model alongside a table or graph representing the turbine's power curve. Fig. 1.3 shows an example of a manufacturer's power curve for a sample pitch regulated wind turbine. Manufacturers' power curve defines essential factors about the turbine model that help customers decide which turbine to select for their wind farm. There is a standard procedure that manufacturers should follow to prepare power curve information, which is called IEC, where more details in this regard can be found in (Commission et al., 2005).

1.3.3 Empirical Power Curves

The IEC-based power curve prepares information about the behavior of new wind turbines under the condition of the test site location that are designed to assure the power curve's accuracy. According to different atmospheric conditions (e.g., air density, wind velocity distribution, obstacle near the wind farm, mechanical and control issues, and so on) in the current site as well as wind turbine's age, compared to the test site, wind turbines may perform not the same as the expectation from the

IEC-based power curve. Hence, wind turbine power curves are often replaced with empirical power curves that are obtained using real data sets retained in the wind farm. Another reason for requiring an empirical power curve is that rapid fluctuation is often ignored in an averaging procedure done in the IEC standard. Besides, the IEC-based curve is neither site-specific nor considering the wear and tear of turbines of a wind farm (Kusiak et al., 2009). In conclusion, it is recommended to utilize the empirical wind turbine power curve in practice, and estimate the generated power using statistical methods based on real historical data from that specific wind farm site (Trivellato et al., 2012; Shokrzadeh et al., 2015).

1.4 Modeling Applications

In the previous section, we explained that to overcome the drawbacks of the manufacturer's power curve, wind farm managers may build site-specific empirical wind turbine power curves using model fitting techniques. In this context, empirical power curve modeling is often done for one of the following reasons:

- Wind energy prediction and analysis
- Selecting appropriate wind turbines
- Condition monitoring
- Battery size estimation.

In the following sections, each of these objectives will be explained briefly.

1.4.1 Wind Energy Prediction and Analysis

One of the critical and challenging tasks for wind farm developers is to accurately estimate the wind farm's future energy production. Inaccurate estimates may cause considerable financial damage to investors. Finding an appropriate candidate site with high meteorological potential will provide sufficient available wind resources (Manwell et al., 2010). One can easily estimate available wind energy that can be produced over a period of time by having enough information about wind speed data in the candidate site and utilizing an accurate empirical power curve. In addition, an accurate wind turbine power curve can facilitate decisions about expanding wind farms (Norgaard and Holttinen, 2004). In Jin and Tian (2010), an analysis of wind energy production dynamics was investigated, and dynamic power curves were proposed. They also estimated the power uncertainty, especially when the wind turbine generator operates between the cut-in and the rated wind speed. In Olaofe and Folly (2013), accurate estimation and controlling the variability of a wind farm's power output were suggested to provide stable wind power to the grid. The authors concluded that their proposed method would improve the loss of load expectation and confirmed that utilizing a developed wind farm's empirical power curve is more accurate than the manufacturer's power curve. Accurate forecasting of wind power in hours or days ahead plays an essential role in electricity markets. In Botterud et al. (2010), the effect of an inaccurate wind power forecast was investigated and concluded that an inaccurate wind power forecast leads to irreparable damages.

1.4.2 Selecting Appropriate Wind Turbines

Wind farm developers can use wind distribution in wind farm locations and the estimation of generating power based on the power curve model to choose turbines with optimum efficiency and performance. In Simic and Mikulicic (2007), the wind turbine power curve's impact on the wind energy cost was studied. In addition, the optimal system configuration was investigated in a small wind off-grid power system. In Jangamshetti and Rau (2001), using the normalized power curves was proposed to help choose a wind turbine generator that yields higher energy at a higher capacity factor. The proposed generalized curves in Jangamshetti and Rau (2001) could help wind farm developers in the process of planning and development stages of wind power stations. In Bencherif et al. (2014), annual capacity factors were calculated based on the Weibull distribution function and power curve models. Also, the optimum selection of wind turbine generators from the site's viewpoint was investigated. Further researches studied the impact of power curve models on wind turbine selection and selected turbines' efficiency. For moew details see Cocina et al. (2015), Pallabazzer (2003) and reference therein.

1.4.3 Condition Monitoring

Empirical power curve models can be considered as effective condition monitoring tools (Pandit, 2018). They can be used as a valuable reference for monitoring the wind turbine's performance. Under normal conditions, the turbine's generating power should not be far from the expected power from the fitted power curve (Kusiak et al., 2009). In Marvuglia and Messineo (2012), a state model of a wind farm in normal condition was built using a data-driven approach. According to this state model, quality control charts were built for detecting anomalous functioning operation and faults detection in the wind farm. In Kusiak and Verma (2012), three different power curves were studied to monitor wind turbines' performance at a wind farm. In Kusiak and Li (2011), in addition to fault detection using the empirical power curve, the severity of faults were investigated. In Pandit and Infield (2018), an algorithm based on the Gaussian process using the Supervisory Control and Data Acquisition (SCADA) data was proposed as a cost-effective approach to wind turbine health monitoring. In Gill et al. (2011), early identification and detection of faults, including blade, yaw, and pitch errors, were studied using copula model and statistical signatures of faults or anomalies. In Kusiak and Li (2011), power curve models based on the data mining approaches were utilized to predict specific faults 60 min before occurrence.

1.4.4 Battery Size Estimation

Energy storage systems are required to integrate high penetrated wind and solar renewables on an energy network to overcome the issue of wind and solar energy fluctuations. To have an electric grid using 100% renewable-based intermittent renewables like wind and solar, an appropriate size of energy storage systems is necessary (Alotto et al., 2014; Castillo and Gayme, 2014; Denholm and Hand, 2011). Shokrzadeh (2014) concludes that one may convert intermittent renewables to baseload generation with sufficient size of energy storage systems by smoothing out the intermittency. Besides, it can allow penetrations of more renewables. The reduction in the cost of energy storage systems leads to the availability of larger-scale energy storage systems in power systems. In Shokrzadeh and Bibeau (2012), the authors discuss repurposing electric vehicle batteries by integrating energy storage systems and wind energy. They define a cost formula and discuss the impact of size and cost of storage system on the total cost of wind power production.

1.5 Empirical Power Curve Modeling and Its Accuracy

In this section, we explain the details of an empirical power curve modeling. The data set for fitting an empirical wind turbine power curve should contain wind speed and the corresponding generated power recorded at periodic intervals over a sufficiently long time. There exist different types of historical data sets. One of the commonly used data sets is collected from experimental wind farms which is used in this study. The other data set contains the Supervisory Control and Data Acquisition (SCADA) system data set. SCADA data set is not available for many wind farms, especially for the old ones. It causes a restriction for these wind farms to only use wind speed and power data for fitting the empirical power curve. In this thesis, we focus on fitting the power curve for turbines in wind farms without the SCADA data set. However, different wind farms may have different systems for collecting data.

There exist different methods to estimate empirical power curves using wind speed and corresponding power value. In general, one can classify the empirical power curve fitting methods into parametric and non-parametric techniques.

1.5.1 Parametric Techniques

Parametric techniques define the relationship between the input and output by a set of mathematical equations with a finite number of parameters. One can formulate the mathematical expressions of parametric models as follows:

$$P(v; \mathbf{\Theta}) = \begin{cases} f_1(v; \Theta_1) & 0 < v \leq v'_1, \\ f_2(v; \Theta_2) & v'_1 < v \leq v'_2, \\ f_3(v; \Theta_3) & v'_2 < v \leq v'_3, \\ \vdots & \\ f_b(v; \Theta_b) & v'_{b-1} < v \leq v'_b, \end{cases} \tag{1.2}$$

where v'_i is wind speed at the i^{th} wind speed segment. Each of $f_i, i = 1, \ldots, b$, may have several fixed number of parameters Θ_i, which are usually collected together to form a single parameter vector $\mathbf{\Theta} = (\Theta_1, \Theta_2, \cdots, \Theta_b)^\top$. For instance, in Khalfallah and Koliub (2007), piecewise linear approximations are fitted using the equation of a straight line as the empirical power curve. Different polynomial methods, including quadratic, binomial, and cubic polynomial methods, are proposed to fit the power curve in Diaf et al. (2008); Giorsetto and Utsurogi (1983); Deshmukh and Deshmukh (2008). Four- and five-parameter logistic approximations are another set of proposed methods for fitting power curves (Villanueva and Feijóo, 2016b; Sohoni et al., 2016a; Lydia et al., 2015). Thapar et al. (2011) proposed two different kinds of models. The first type of model was built based on the fundamental equation of power available in the wind, and the other model was formulated using the concept of the turbine's power curve. Models based on the power curve of the turbines performed more accurately than equation-based models. Also, they concluded that equation-based models were very cumbersome.

1.5.2 Non-Parametric Techniques

Nonparametric models make no assumption about the functional form of the relationship between wind speed and wind power. So, these models are capable of fitting a wide range of power curves' shapes, compared to parametric models (Eubank, 1988; Friedman et al., 2001; Hollander et al., 2013).

Shokrzadeh et al. (2014) introduced two nonparametric methods, cubic spline regression and penalized spline regression model, which outperformed parametric methods. Different modification of artificial neural network models are also proposed in Li et al. (2001); Pelletier et al. (2016); Manobel et al. (2018). Data mining algorithm methods are another class of commonly used methods for wind turbine power curve. These methods are powerful in extracting patterns in huge data sets.

In conclusion, there is no best empirical power curve modeling approach for all types of turbines in all wind farms. For each particular data set, a specific fitted curve might work best. So, it is critical to evaluate the performance of different power curve fitting methods based on statistical metrics on a data set. Then, one can decide which method performs better for a given data set.

1.5.3 Modeling Accuracy

Model accuracy is the most important criterion to use while comparing different power curve modeling approaches. Researchers compare their proposed methods based on different statistical metrics (Marvuglia and Messineo, 2012). The most commonly used metrics in the literate include mean absolute error (MAE), symmetric mean absolute percentage error (sMAPE), normalized mean absolute percentage error (NMAPE), root mean squared error (RMSE). We explain each of these metrics by presenting the corresponding equation of each metric. First, we define absolute error (AE) and relative error (RE) associated with i^{th} observation in a data set as follow:

$$AE(i) = |\hat{P}_i - P_i|, \tag{1.3}$$

$$RE(i) = |\frac{\hat{P}_i - P_i}{P_i}| \times 100, \tag{1.4}$$

where $\hat{P}_i$ is the predicted power value and P_i is the actual power. One can formulate different accuracy metrics using Eq. (1.3) and Eq. (1.4). For example,

$$MAE = \frac{1}{N}\Sigma_{i=1}^{N}|\hat{P}_i - P_i| = \frac{1}{N}\Sigma_{i=1}^{N}AE(i), \tag{1.5}$$

$$RMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(\hat{P}_i - P_i)^2} = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}AE^2(i)}, \tag{1.6}$$

$$sMAPE = \frac{1}{N}\Sigma_{i=1}^{N}\frac{|\hat{P}_i - P_i|}{(|\hat{P}_i + P_i|)/2} \times 100 = \frac{1}{N}\Sigma_{i=1}^{N}\frac{2P_i}{\hat{P}_i + P_i}RE(i), \tag{1.7}$$

$$NMAPE = \frac{1}{N}\Sigma_{i=1}^{N}\frac{|\hat{P}_i - P_i|}{\max_{i=1}^{N}(P_i)} \times 100 = \frac{1}{N}\Sigma_{i=1}^{N}\frac{P_i}{\max\limits_{1\leq j\leq N}(P_j)}RE(i). \tag{1.8}$$

One can use any of these metrics for evaluating the goodness-of-fit for any fitted curve. A better fit corresponds with the one with a smaller value of the underlying metric. However, to better evaluate the performance of a method, these criteria should be calculated on a test data or be averaged over different folds in a cross-validation procedure.

1.6 Outlier Issues

In this section, we discuss outliers data that can highly impact the accuracy of fitted curves. Wind farm data usually contain abnormal data that are far beyond the expected output of the power curve. An application of condition monitoring using an empirical power curve is to detect anomalies or outliers. Therefore, accurate wind power data facilitate the secure operation of wind turbines and help in optimizing the control strategy. Accurate wind power data is dependent on the number of outliers in the historical data set. Different reasons such as turbine shutdown, wind speed sensor failure, load shedding, dirty or damaged turbine blades, communication noise, and equipment failure may cause the existence of outliers in the wind farm data set (Shen et al., 2018). Statistical analysis of wind power data will be distorted if data contain a high portion of anomalies and outliers. Wind power characteristics may change according to these outliers. So, it is required to overcome this issue by removing outliers or reducing their impact. As a pre-processing stage, data cleaning is done to improve the quality of wind farm data by removing outliers (Swapna et al., 2016).

1.6.1 Outlier Detection and Pre-processing Data

Outlier detection is a well established topic in the literature (Ben-Gal, 2005; Rousseeuw and Leroy, 2005; Domingues et al., 2018; Daki et al., 2017). Identification and cleaning of wind turbine outliers became an exciting area of wind energy research, and outstanding achievements are established in this area. In Zhao et al. (2017), a quartile algorithm was utilized for identifying the outliers in wind turbines data. The quartile algorithm is a commonly used method that is the most effective in a dataset with a small portion of outliers. In Zhao et al. (2014), in addition to utilizing the quartile algorithm, the K-means method was applied to the data set to find clusters of outliers. This method suffers from two points. One is the complexity of finding the correct number of clusters, and the other is that this method may eliminate lots of normal data by considering them as a cluster of outliers. In Lou et al. (2016), an effective solution to detect stacked outliers below the power curve based on the intra-group optimal variance algorithm was proposed; however, this method was not accurate to detect outliers above the power curve. In Kusiak et al. (2009), a novel method based on a nonlinear power curve modeling method was proposed for filtering outliers. This method performs well for a data set with a high portion of normal data, but it is a time-consuming method.

One may classify outlier detection methods into three types (Shen et al., 2018). In the first class, outliers are detected based on the data density or distance from other data points (Zhao et al., 2014; Zheng et al., 2014). This class of methods performs poorly in the detection of densely distributed stacked outliers. The second class of outlier detection methods is based on mathematical modeling of wind turbine power curves. Such methods filter out outliers according to the residual of data from the fitted power curve and are not effective when the wind farm performance is abnormal, which cause the existence of a high number of outliers in the data set or high variance of generated power in the data set (Kusiak et al., 2009). In the third class (e.g., Zhao et al., 2017), outlier detection is based on the location of data in the distribution of total data. The main idea of this class is that outliers are beyond the acceptable range of normal data distribution (Lou et al., 2016; Zheng et al., 2014). This class of method is universally applicable to different types of outliers. All of these methods try to remove outliers, and none of them are capable of removing all outliers.

1.7 Heteroscedasticity

In addition to outliers, heteroscedasticity of wind speed-power data plays a critical role in improving fitted power curves' accuracy. Heteroscedasticity is a usual issue in the wind farm data set, which means the variance of recorded wind power is dependent on wind speed. In classic literature, this issue was considered in the same way as outlier detection, and researchers have tried to overcome this issue by removing lots of data; however, these data are not outliers. Sometimes turbine's aging may cause them not to generate their rated power even when the wind speed is in the range between rated and cut-out points. Heteroscedasticity causes the estimated errors of wind power to become large and sometimes have a long tail distribution (Wang et al., 2018). One can find worthy studies in statistics regarding heteroscedasticity as in Carroll and Ruppert (1982); Park (1966); Rao (1970). However, considering heteroscedasticity in wind turbine power curve fitting is a new subject, and the number of articles related to this subject is meager. In Wang et al. (2018), the heteroscedastic spline regression model (HSRM) and robust spline regression model (RSRM) were utilized to detect inconsistent samples, followed by an ANN-based forecasting model to obtain power forecasts. In Rogers et al.

(2020), probabilistic methods were proposed as a natural framework to quantify risk for facilitating decisions. The probabilistic methods are fitted based on the heteroscedastic Gaussian process method. The authors utilized the heteroscedastic property of data to calculate the estimation's uncertainty without any attempt to overcome this issue. Overcoming the heteroscedasticity issue plays an important role in wind farm management and fitting power curve. In Chapter 3 of this thesis, we propose a novel approach using two different weighting schemes to take into account target power curve models based on locally balanced loss function to reduce the negative impact of the heteroscedasticity issue.

1.8 Wind Farm Aggregated Models

So far, we have explained power curve modeling for individual wind turbines, but it is often necessary to study the wind farm aggregated power curve models. One of the wind farms aggregated power curve model applications is in estimating the annual wind power. Since the wind availability is dependent on the weather condition, estimating the annual wind power is very challenging.

In Commission et al. (2005), according to wind speed frequency distributions, the total energy production of a wind turbine in a test site is estimated under specific conditions during a one-year period, assuming 100% availability. However, energy production may vary from one year to another as well as from one wind farm to another (Marinelli, 2011). Forecasting a wind farm's energy production through predicting the energy production of each wind turbine in the wind farm increases the computational cost. In massive wind farms with tens or hundreds of wind turbines containing generators, turbines, transformers, cables, and control systems, detailed modeling of the power system is very complex, and any simulation is time-consuming (Ali et al., 2012; Fernandez et al., 2009). So, any approach to simplify the wind farm model, which reduces the simulation time and complexity of the wind farm model while keeping the model's accuracy at an acceptable level, can improve the estimation of the total energy production. To this end, one approach is to utilize the wind farm aggregated models while maintaining their accuracy (Chowdhury et al., 2013). Several different aggregated models are currently introduced, which can be classified into two general classes: single aggregated model and multiple aggregated model (e.g., Fernández et al., 2008; Li et al., 2012). Some of the commonly used methods for the single aggregated model include artificial neural networks combined with the wavelet transform (Catalão et al., 2011) as well as radial basis function neural network and back-propagation neural network (Yongqian et al., 2011).

Nevertheless, a single aggregated model can be useful for a small wind farm with one type of turbine model; otherwise, one single aggregated model will not accurately represent the complex behavior of the entire wind farm. To this end, often multiple aggregated models are utilized to reduce the risk associated with a single aggregated model and the computational cost associated with separating models for each turbine in a wind farm. In most of the multiple aggregated models, one first clusters turbines with similar characteristics into several groups, consequently, fits a single aggregated power curve model for each cluster. So, in multiple aggregated models, two components are fundamental and need to be selected carefully. One is the clustering features and the clustering method to classify similar turbines, and the other one is the appropriate method of fitting the wind power curve models. Several different methods are proposed in the literature. In Zou et al. (2015), utilizing the Fuzzy clustering method based on some indicators such as wind speed, slip ratio, and stator voltage is proposed. In Ali et al. (2012), the Support Vector Clustering (SVC) technique is proposed to cluster wind turbines based on features such as the location of turbines in the wind farm and incoming wind. Consequently, a probabilistic aggregated model is proposed by taking into account the wake effect. We briefly explain wake effect in the following section. In Zhang and Liu (2019), a matrix of features is built on the wake effect of turbines. To cluster turbines, the authors proposed a singular value decomposition (SVD) clustering algorithm.

1.8.1 Wake Effect

One of the factors which impacts wind farm efficiency is wind shadow or wake effect (Al-Shammari et al., 2016; Shamshirband et al., 2014). The wake effect represents the negative impact of an upper turbine on a downwind turbine, which reduces the wind power energy. In this case, the downwind turbine energy production will be reduced in that wind direction compared to the upper turbine (Changshui et al., 2011). Fig. 1.4, represents an example of the wake effect in a wind farm. The wake effect is dependent on the distance between turbines and the layout

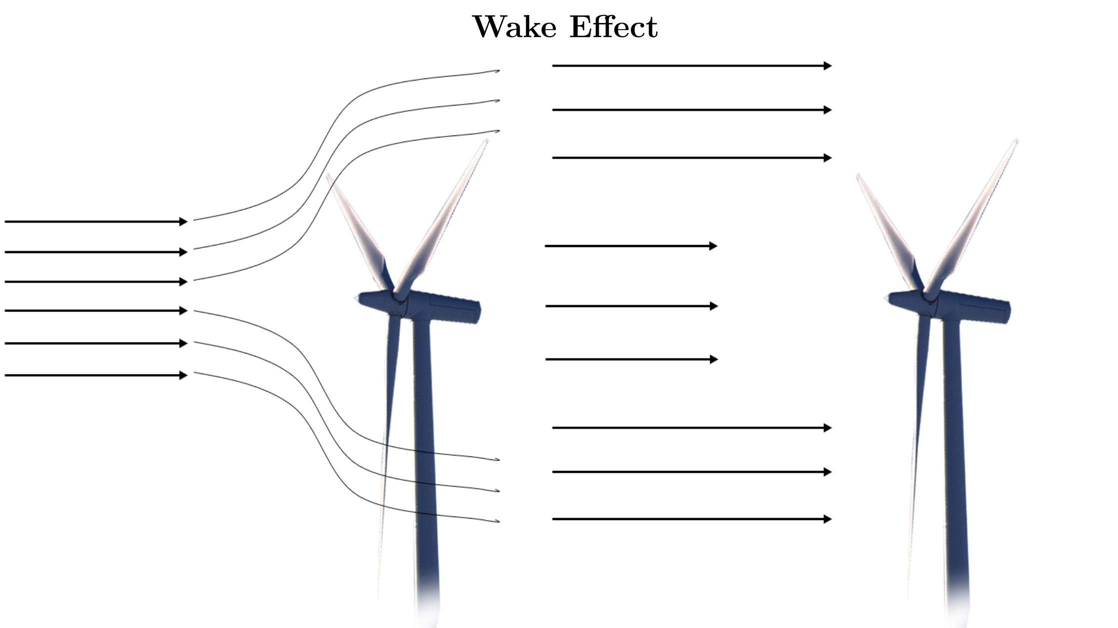


Figure 1.4: Wake effect of an upper wind turbine on a downwind turbine which causes the downwind turbine generates less power than expected.

of the wind farm. Consequently, the wake effect is a critical concept that should be considered in designing the wind farm layout and estimating the annual wind energy production. It also plays a vital role in wind turbine power curve fitting as well as clustering wind turbines in a wind farm (Ekonomou et al., 2012; Saavedra-Moreno et al., 2011; Eroğlu and Seçkiner, 2012; Yin and Wang, 2012).

Several different factors including the wind farm location, number of turbines in a wind farm and their size, the terrain morphology, the distribution of wind direction, the wind speed, and the design of blades may have impacts on the wake effect in a wind farm (Chen et al., 2013; Petković et al., 2014; Rašuo and Bengin, 2010). The wake effect is one of the most common reasons which reduces the efficiency of wind power production (Lignarolo et al., 2014; Dörenkämper et al., 2015). Wake effect not only reduces the efficiency of downwind turbines but also may lead to possible mechanical failure in the cases with lots of interference and should be considered as a vital factor in fitting aggregated power curves (Subramanian et al., 2016; Chowdhury et al., 2015).

1.9 Problem Description and Thesis Objectives

In recent years with a high increment in the number of wind farms and the number of turbines in wind farms, solving the challenges pertinent to wind turbines and wind farms became a priority in renewable power systems. In the previous sections, we explained the wind turbine power curve fitting process and its importance on wind farm management and maintenance cost. Although there have been significant improvements in some aspects of power curve fitting models, several challenging issues are required to be addressed to increase the efficiency of wind turbines and wind farms' performance. Some important challenges are studied in this thesis that are outlined in Chapters 2, 3 and 4 of this thesis. These chapters are based on two published papers in Renewable Energy (2020) and Energy (2021), one conference paper presented in the 6th International Conference on Green Energy (2018) as well as an accepted paper for publication in the IEEE Transactions on Sustainable Energy (2021). More details regarding the content of each chapter that addressed specific goals toward the objective of this PhD thesis are given below.

1.9.1 Chapter 2

The first challenging issue is to reduce the effect of outliers on wind turbine power curve modeling. Outliers may substantially change the shape of the empirical power curves and subsequently reduce the accuracy of all analyses. A clean wind speed-power data set will have the most similar empirical power curve to the manufacturers' power curve. However, after a while, the turbine's aging issue decreases the turbine's rated power and efficiency. Consequently, the shape of the empirical power curve may differ from the manufacturer in a minor amount. But, it is clear that abnormal data, which are far from other normal data points, may pull the fitted empirical power curve toward themselves and significantly change the power curve's shape. This affected power curve will perform poorly in power forecasting and management decision applications compared to the empirical power curve, which is impacted only by turbines' aging. There are many different methods for detecting abnormal data and outliers in the literature; however, researchers did not pay sufficient attention to methods that may reduce the effect of outliers on fitting empirical power curves. As none of the outlier detection methods can identify all the outliers in a data set, it is desirable to have fitting methods that reduce the outliers' effect. To this end, considering some inherent features of the manufacturer's power curve, e.g., monotonicity of power curve, and applying them in fitting empirical power curves on wind farm data set can reduce outliers' impact. By enforcing the fitted curve to have a similar shape to the manufacturer's power curve, one can restrict the fitted curve from changing majorly toward outliers, which are far from the normal data.

By this approach, one can reduce the effect of outliers. The overall contributions associated with this Chapter are:

- investigating the impact of enforcing inherent properties of the manufacturer's power curve on reducing outliers effect on fitting power curves,
- enforcing the empirical power curve to be monotone, which is one of the manufacturer's inherent properties,
- proposing an approach to enforce monotonicity for both parametric and nonparametric power curve fitting methods,
- reducing the variance by keeping the bias of fitted curve unaffected,
- comparing the accuracy of monotonic power curves and other commonly used power curves in estimating power prediction.

1.9.2 Chapter 3

In addition, we investigate the heteroscedasticity property of wind speed-power data. The heteroscedastic property of data increases the uncertainty of estimation according to the high variance of generated power, especially in the range of high wind speed. So, commonly used empirical power curve models are not able to accurately forecast the wind power. To this end, we propose an approach to take into account other valuable information resources such as similar turbine's power curve for a similar turbine, manufacturer's power curve, theoretical power curve, or wind farm aggregated power curve for the regions where there exists an uncertainty in estimation. In other words, we take advantage of other resources other than just the empirical power curve when the uncertainty of the fitted empirical power curve is high. So, we suggest to partially use some valid and well-defined target power curves, e.g., manufacturer's power curve or wind farm average power curve (in this study), in the region with high variance and uncertainty. To consider the level of uncertainty, we propose two different weighting schemes, which are based on the variance of data in different bins of wind speed. Using the weighting scheme, we propose an approach to produce a hybrid power curve using both the empirical power curve and target power curves. To have a systematic approach for combining two power curves (empirical and target), we define a locally balanced loss function, which is explained in this Chapter in more detail. The overall contributions associated with this Chapter are:

- exploring the heteroscedastic property of data in wind turbine power curve modeling,

- defining weighting schemes based on heteroscedasticity of wind speed-power data,

- proposing a locally weighted balanced loss function which considers goodness of fit and closeness to a target model,

- suggesting two different target models which are worthy resources for fitting power curves,

- reducing the bias of fitted power curve alongside keeping the variance unaffected,
- introducing a general methodology to produce hybrid power curves using any empirical power curve fitting model and target model,
- comparing hybrid methods with other commonly used method based on statistical metrics for measuring accuracy.

1.9.3 Chapter 4

In addition to the challenges pertinent to fitting power curve for individual turbines, fitting wind farm aggregated models has different challenges. Fitting power curves for each turbine individually is required for scheduled maintenance and fault detection. On the other hand, fitting the power curve for a wind farm is required for the overall analysis, such as estimating annual power production. As explained earlier, regarding the high complexity of modeling each turbine separately for annual power prediction or other applications in wind farm management, simple models are preferred for aggregated wind farm power curves. There exists a trade-off between complexity and accuracy. In other words, complex models suffer from their high computation cost, although they perform accurately. On the other hand, simple models take advantage of low complexity, making it quick in analysis and simulation, but their accuracy may not be sufficient. Consequently, it is required to find an appropriate approach which profits from the pros of both types of models. It means that the wind farm aggregated power curve model should perform accurately while being simple. Multiple aggregated power curves can be used to balance this trade-off. Accordingly, we propose a novel methodology for wind farm multiple aggregated power curves that reveals the high accuracy and reduces the complexity of the fitted model. Two important steps in fitting multiple aggregated power curves for wind farms include selecting an appropriate clustering approach as well as an appropriate fitting method. We study a new clustering approach to classify wind turbines in a farm by considering turbines' wake effect and overall performance in a wind farm. The clustering method is based on the K-means clustering technique using the Euclidean distance between turbine feature sets. We define a novel feature set for turbines to take into account their overall performance in generating power. Hence, the clustering approach's goal is to assign turbines into different groups such that the pattern of generating powers of turbines inside each group is homogeneous, and the pattern of generating power for turbines from different groups is non-homogeneous. Generally, researchers cluster wind turbines into multiple groups to reduce the model complexity, the number of parameters and simulation time. By clustering, one can reduce the order of fitted power curves from the number of turbines in a wind farm to the number of clusters. Another important factor ignored in the literature is that there is a hidden correlation between the fitted power curves for different clusters. By considering the correlation of fitted power curves for each cluster, one can improve the accuracy of the wind farm aggregated model and reduce the complexity of the model as well as the number of parameters. Thus, we propose to utilize multilevel modeling approaches to take into account the correlation of power curves fitted on each cluster when modeling the wind farm aggregated model. The details of our novel clustering feature selection approach based on generated power pattern that considers the wake effect regardless of the turbines' location, as well as the multilevel modeling approach for wind farm power curve modeling, are presented in Chapter 4.

The overall contributions associated with this Chapter are:

- defining a feature set for clustering wind turbines into several homogeneous groups which takes into account the overall performance of the turbine as well as their wake effect,
- considering the hidden correlation of different clusters in fitting wind farm multiple aggregated power curve by applying advance statistical models such as mixed effect models
- utilizing statistical tests to confirm the effectiveness of proposed wind farm multiple aggregated power curve.

1.10 Thesis Overview

This thesis represents proposed methods, the comparison of proposed methods, and results obtained from the experiments investigated throughout the Ph.D. study. There are 5 chapters in this thesis. In Chapter 1, we introduced the research context and motivation, the background and challenges associated with wind turbine power curve modeling, and the aim and objectives of this research. Chapter 2 introduces a monotonic regression method for wind turbine power curve modeling, with a focus on the impact of enforcing the manufacturer's power curve properties in reducing the outlier effect on fitting an empirical power curve. In Chapter 3, we discuss the heteroscedastic issue in the wind speed-power data set and define a novel weighted locally balanced loss function that accounts for both estimation error (using observed data) and goodness of fit that reflects the proximity to a target model. We propose two different weighting schemes to deal with non-homogeneity and in particular high variability of generated power of each turbine in different wind speed regions. These weights are then used in the loss function to balance the importance of estimation error and proximity to target values. It leads to a hybrid method for the power curve fitting on a heteroscedastic data set. In Chapter 4, we discuss a clustering approach using a novel feature set that is based on wind turbine performance in generating power. We study multilevel modeling and discuss the utilization of the mixed effect model to reduce the order of wind farm aggregated power curve. We evaluate the performance of our methodology based on actual data sets of a wind farm in Canada. In summary, this thesis is structured based on the grouped manuscript style (i.e., sandwich thesis). Mr. Mehrjoo has been the main contributor and first author of all the manuscripts presented in this thesis. His contribution to this work includes conception, developing the research questions, designing the studies, developing all the codes,pre/processing data, conducting the analyses, writing up all manuscripts, submitting all the manuscripts, and responding to reviewers' comments. Dr. Jafari Jozani contributed to the conception and design of the study, mathematical theory, editing manuscripts and reviewing process. Dr. Pawlak contributed to mathematical theory and editing manuscripts.

Chapters 2 contains a peer-reviewed journal paper published in *Renewable Energy* Journal. Chapter 3 contains a peer-reviewed journal paper published in *Energy* Journal. Chapter 4 contains an under-review journal paper. These papers, in addition to a conference paper presented at *Green Energy and Expo*, collectively contribute toward the goal of this Ph.D. thesis. Following is the list of these papers:

- Chapter 2: Mehrdad Mehrjoo, Mohammad Jafari Jozani, and Miroslaw Pawlak. "Wind turbine power curve modeling for reliable power prediction using monotonic regression." Renewable Energy 147 (2020): 214–222.

- Chapter 3: Mehrdad Mehrjoo, Mohammad Jafari Jozani, and Miroslaw Pawlak. "Toward hybrid approaches for wind turbine power curve modeling with balanced loss functions and local weighting schemes." Energy 218 (2020).

- Chapter 3: Mehrdad Mehrjoo and Mohammad Jafari Jozani. "A weighted hybrid wind turbine power curve modeling approach using spline regression and theoretical power curve." 6^{th} International Conference on Green Energy and Expo (2018).

- Chapter 4: Mehrdad Mehrjoo, Mohammad Jafari Jozani, Miroslaw Pawlak and Bagen Bagen. "A multilevel modeling approach towards wind farm power curve." Accepted for publication in IEEE Transactions on Sustainable Energy, 2021.

Finally, in Chapter 5, we conclude the research work presented in this thesis and explain some suggestions for future study in this area.

Chapter 2

Monotonic Power Curves

This chapter encloses a peer-reviewed journal paper published in Renewable Energy Journal that consists of the author's work during his Ph.D. study[1].

Abstract: Wind turbine power curve modeling plays an important role in wind energy management and power forecasting and it is often done based on parametric or non-parametric methods. As wind-power data are often noisy, even after polishing data using proper methods, fitted wind turbine power curves could be very different from the theoretical ones that are provided by manufacturers. For example, it might be the case that the theoretical wind turbine power curve is a non-decreasing function of speed but the fitted statistical model does not necessarily meet this

[1]Mehrdad Mehrjoo, Mohammad Jafari Jozani, and Miroslaw Pawlak. "Wind turbine power curve modeling for reliable power prediction using monotonic regression." Renewable Energy 147 (2020): 214–222. https://doi.org/10.1016/j.renene.2019.08.060

desirable property. In this paper, we present two nonparametric techniques based on tilting method and monotonic spline regression methodology to construct wind turbine power curves that preserve monotonicity. To measure the performance of our proposed methods, we evaluate and compare our estimates with some commonly used power curve fitting methods based on historical data from a wind farm in Manitoba, Canada. Results show that monotone spline regression performs the best while the tilting approach performs similar to the methods we studied in this paper with the benefit of finding a curve that is more similar to the theoretical power curve.

2.1 Introduction

U.S. effort to achieve 20% electricity of whole country supplied from wind power by 2030 is one example to show how important are renewable energy sources (Lindenberg et al., 2008). A great deal of effort has been devoted to wind power forecasting over the past years and a considerable improvement in wind power prediction is achieved by developing new methods. Despite enormous research efforts and acquired knowledge, power reliability of wind power systems is still one of the most demanding areas in research pertaining to sustainable energy. Evaluating the power system reliability contains three general phases (Bagen and Billinton, 2008). First, one needs to predict the wind speed in the location of turbines using either physical or statistical approaches. The physical approach, uses the detailed physical description of the wind farm to model the on-site conditions. The statistical approach, which is rather easy and inexpensive, generally uses historical data to train necessary statistical models for reliable wind speed prediction. This is often done using conventional statistical approaches based on statistical time series models or machine learning algorithms (Jung and Broadwater, 2014).

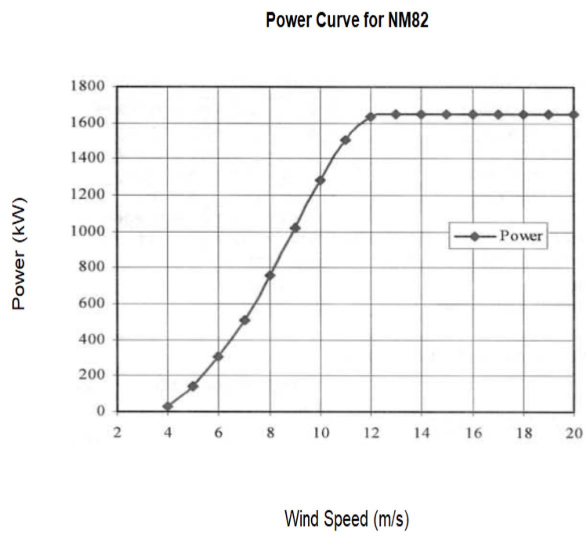
The second phase of power system reliability analysis, which is the focus of this paper, is to accurately estimate the generated power based on the predicted speed. Accurate estimation of the turbine power curve is required for effective integration of wind power into the power systems (Shokrzadeh et al., 2014; Zhang et al., 2014). To achieve an accurate estimation of generated power value, efficient methods of fitting power curve models are needed. Statistical methods to fit the empirical power curve of a wind turbine are abundant and can be classified into parametric and nonparametric methods (Jung and Broadwater, 2014). Parametric methods include segmented linear models, polynomial regression, exponential power curve, cubic power curve and models based on probabilistic distributions such as four or five parameters logistic distributions (Lydia et al., 2013; Carrillo et al., 2013; Villanueva and Feijoo, 2018; Marčiukaitis et al., 2017). On the other hand, nonparametric methods do not assume any specification for the underlying model and estimate the power curve to be as close as possible to the observed data, hence they can accurately model a wide range of possible shapes of power curves. Examples of these methods include neural networks, random forest, and various forms of k-nearest neighbor, kernel algorithms and natural cubic spline regression (Pelletier et al., 2016; Ouyang et al., 2017; Marvuglia and Messineo, 2012; Ai et al., 2003; Diaf et al., 2007;

Pandit et al., 2019).

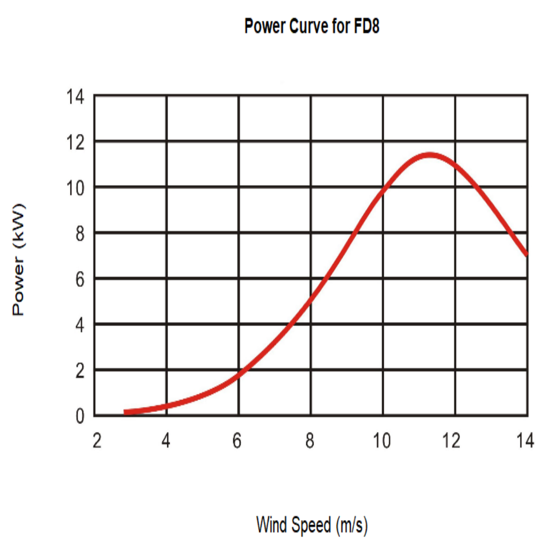
Wind turbine power curves ideally correlate wind speed, measured immediately before the rotor, with electrical power from the turbine. In practice, it is not possible to accurately measure the wind speed at such a close distance to the rotor. This might cause bias in the resulting power curve. The solution to this issue is either to apply suitable time averaging on the data (Elliott and Infield, 2014; Sharpe et al., 2013) or apply a measurement error methodology. To examine the importance of sufficient time averaging on the data, we compared 10-Min and hourly averaged data in our experiment.

Different turbine models may have different shapes of power curves. Fig.2.1 shows two different power curves related to wind turbines NM82 and FD8. One of these power curves is monotone in the whole range of the wind speed till the cut-out point, however, the other one behaves differently.

Most of the statistical approaches for wind turbine power curve modeling that are available in the literature do not incorporate some of theses intrinsic properties of power curves in their estimation process. The main focus of this article is to develop methods to incorporate some inherent, physical and/or sometimes theoretical properties of the underlying wind turbine power curves in the power curve model fitting process. For example, if according to theory, the fitted curve needs to be monotone in a specific region, practitioners might want to enforce this into their estimation or optimization process. From practical point of view, sometimes finding a model that satisfies some important features of the underlying relationship between power and speed is of interest. In practice, it is reasonable to require that employed



(a)



(b)

Figure 2.1: Theoretical power curve for two different wind turbine models showing different shape properties a) Turbine model NM82, b) Turbine model FD8.

estimates of the power curve preserve the shape property of the underlying wind turbine. In this research, we focus on the wind turbines with monotonic power curve e.g. NM82, however, our proposed methods can be utilized for other wind turbines with different power curve shapes.

Fig.2.2 shows real data of a wind turbine alongside with its corresponding manufacturer power curve as well as estimated power curve using a Nadaraya-Watson kernel estimator method. Although the data are polished and properly cleaned and the outliers are detected (using the polishing data method proposed in Shokrzadeh et al. (2014) which removes all data with negative speed or power values as well as data points that lie outside of three standard deviation of the average power value at that speed), still we see that monotonic property of estimated power curve is not preserved. Fig.2.2 also represents the estimated power curve using the spline method. when the existing outliers have not been removed from the original data. The effect of the outliers is visible as it causes the tail of the estimated power curve to decline. Another observation is that not only outliers can cause a fitted power curve to show a different behavior than the theoretical (manufacturer) power curve model, but also modern nonparametric methods might not necessarily preserve the inherent monotonicity property of the curve. One might decide to remove data that cause non-monotonicity from the training data. However, how much data we need to remove depends on the working data and might change from one turbine to another one. This makes the whole process very subjective. Removing many observations from the training data to enforce monotonicity results in a smaller dataset to train models and sometimes it reduces the goodness of fitted curves. In this paper, we enforce monotonicity as an inherent part of the power curve fitting method. This might also help to reduce the impact of outliers on estimated curves.

We show that our proposed methods do not decline the estimation accuracy of predicted power values and sometimes they can even boost the prediction accuracy of fitted curves. So our results can also be used toward obtaining efficient methods for characterizing wind turbine power curves, as well as other applications such as wind power forecasting, detecting anomalies in a wind turbine power generation process, etc.

The outline of this paper is as follows. In Section 2.2, we briefly overview some parametric and nonparametric power curve estimation methods. In Section 2.3, we introduce two methods of fitting monotone regression model and explain the theoretical basis of each method and details of their implementation. In Section 2.4, we present actual data sets of a wind farm in Canada. We also investigate the performance of each method and explain the evaluation metrics. The results of the proposed techniques are also presented. Section 2.5 provides concluding remarks and future work.

2.2 Power Curve Estimation

The power curve of a wind turbine represents the amount of electrical power output as a function of wind speeds (Gasch and Twele, 2011). Usual wind turbine power curves have three main characteristic speeds which include the cut-in (Vc);

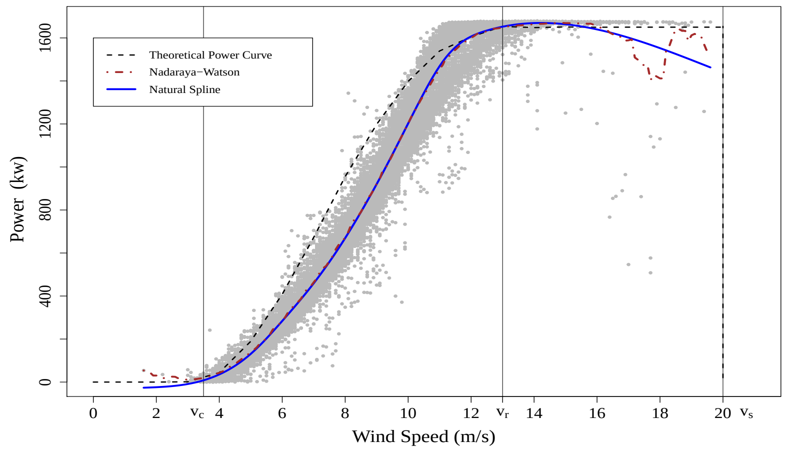


Figure 2.2: Real data of a wind turbine alongside with its corresponding manufacturer power curve as well as estimated power curves using Nadaraya-Watson kernel estimator method and natural spline regression.

rated (Vr); and cut-out (Vs) speeds. The cut-in point is the threshold speed value for generating power, which means wind speed should be at least equal or greater than this value so that wind turbine starts generating electrical power. For any wind speed value less than Vc, wind turbine is not able to produce any power. The rated speed is a speed between the cut-in and the cut-out points, in which wind turbine generates power at its maximum or rated capacity. To prevent damages, the power generation is shut down when the wind speed reaches the cut-out speed (Manwell et al., 2010). Wind turbine manufacturers release a power curve assuming ideal meteorological and topographical conditions. However, in practice ideal conditions do not occur and wind turbines often behave differently from what is expected according to the theoretical power curve. Location of the turbine, air density, wind velocity distribution, wind direction, mechanical and control issues, as well as uncertainties in measurements are some reasons which may cause empirical power curves differ from theoretical ones. The theoretical wind power is obtained using the following equation

$$P_w = \frac{1}{2}\rho A v^3, \tag{2.1}$$

where P_w is the wind power in Watt, ρ is the air density in kg/m^3, A is the turbine rotor area in m^2, and v is the wind speed in m/s (Manwell et al., 2010).

In practice (2.1) is not used to estimate the generated power of a wind turbine. Instead one might want to estimate the generated power using statistical techniques based on historical data. In the following we study some of such methods. In what follows the symbol v is used for wind speed, whereas p stands for generated power.

2.2.1 Polynomial and Spline Regression Models

Polynomial regression is often used to estimate the power curve of a wind turbine (Lydia et al., 2013). This method can be considered as a natural extension of the linear regression $p_i = \beta_0 + \beta_1 v_i + \epsilon_i, i = 1, \ldots, n$, where the linear dependency of p to v is replaced by a polynomial function as follows

$$p_i = \beta_0 + \beta_1 v_i + \beta_2 v_i^2 + \cdots + \beta_k v_i^k + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.2}$$

where ϵ_i is assumed to be a sequence of independent and identically distributed random variables with zero mean and finite variance. Also, n is the number of data points, $\beta_j, j = 0, \ldots, k$, are unknown parameters, and k is the degree of the polynomial regression. One can rewrite (2.2) in the matrix form as

$$P = V\boldsymbol{\beta} + \epsilon \tag{2.3}$$

where $P = (p_1, p_2, \cdots, p_n)^\top$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)^\top$, $\epsilon = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)^\top$, and V is a matrix with its i-th row being defined as $V_i = (1, v_i, v_i^2, ..., v_i^k)$. Using the least squares (LS) criterion and by minimizing the residual sum of squares (RSS), we obtain the explicit estimate of $\boldsymbol{\beta}$ as follows

$$\hat{\boldsymbol{\beta}} = (V^\top V)^{-1} V^\top P. \tag{2.4}$$

Polynomial regression model uses a global approach to construct the fitted value of the power at a given value v_0, which depends strongly on generated power for other v-values even those that are far from v_0. So, one might want to use a more local estimation process to estimate p_0. A solution to this problem is to employ piecewise polynomial regression models, often referred to as spline models. B-spline model is one of such models which constitutes a smooth function having continuous derivatives (Shokrzadeh et al., 2014).

A very attractive property of the B-spline model is that it can also be formulated as $P = B\boldsymbol{\beta} + \epsilon$, where B is a matrix with its (i, j)-th element being

$$B_j^k(v_i) = \frac{(v_i - \zeta_j)}{(\zeta_{j+k} - \zeta_j)} B_j^{k-1}(v_i) + \frac{(\zeta_{j+k+1} - v_i)}{(\zeta_{j+k+1} - \zeta_{j+1})} B_{j+1}^{k-1}(v_i), \tag{2.5}$$

for $j = -K, -K+1, \ldots, K$, $\zeta_0 = \zeta_{-1} = \ldots = \zeta_{-K} = min\{v_i, i = 1, \ldots, n\}$, and $\zeta_{K+1} = max\{v_i, i = 1, \ldots, n\}$. Here $\{\zeta_i, i = 1, \cdots, K\}$ is a set of knots that are selected by the user and control the flexibility of the fitted curve. Also, $B_j^0(v_i)$ are the natural basis for piecewise constant functions. The least squares estimates of $\boldsymbol{\beta}$ is then given by

$$\boldsymbol{\beta} = (B^\top B)^{-1} B^\top P, \tag{2.6}$$

which is more feasible for computational purposes.

It has been observed that spline regression compared with the corresponding global polynomial regression, tends to behave erratically beyond the boundary knots (Friedman et al., 2001). To address this issue the natural spline regressions has been proposed as useful tools to overcome this problem. Natural splines are analogous to B-Spline except they are constrained to be linear beyond the boundary knots.

Several different methods are proposed in the literature to choose the number and location of knots, see (Friedman and Silverman, 1989). In this paper, we use 10-fold cross-validation by randomly dividing our data into 10 folds (random subsets) of approximately equal sizes. In each iteration, we take one fold as test data and the remaining folds as training data. The 10-fold cross-validation is then performed and the number of knots are selected such that corresponding model minimizes

$$CV(10) = \frac{1}{10} \sum_{i=1}^{10} MSPE_i, \tag{2.7}$$

where $MSPE_i$ is the mean-squared prediction error rate associated with the i'th test group.

2.2.2 Kernel Estimators

Another approach that is widely used for wind turbine power curve modeling is based on nonparametric kernel methods (Carrillo et al., 2013; Friedman et al., 2001; Wand and Jones, 1994a; Greblicki and Pawlak, 2008). A classical nonparametric kernel estimator of a regression function is the Nadaraya-Watson (NW) kernel regression estimator (Wand and Jones, 1994a). To define this estimator, we first assume that

$$p_i = m(v_i) + \epsilon_i, \quad i = 1, \cdots, n,$$

where ϵ_i is random error and $m(\cdot)$ is an unknown regression function. We do not assume that $m(\cdot)$ has any parametric form. Under the squared error loss function, the best choice for $m(\cdot)$ is given by

$$m(v) = E(P|V = v) = \int \frac{pf(v,p)}{f(v)} dp, \tag{2.8}$$

where (V, P) denotes the random version of our variables (v, p). Hence, $m(v)$ is the regression function of P given $V = v$. Here, $f(v, p)$ is the joint density function of the wind speed and generated power and $f(v)$ stands for the marginal density function of the wind speed. A generic estimate of (2.8) is as follows

$$\hat{m}(v) = \begin{cases} \int \frac{p\hat{f}(v,p)}{\hat{f}(v)} dp & ; f(v) \neq 0, \\ 0 & ; f(v) = 0, \end{cases} \tag{2.9}$$

where $\hat{f}(v, p)$ and $\hat{f}(v)$ are some estimates of $f(v, p)$ and $f(v)$, respectively. By using the classical kernel density estimate (Wand and Jones, 1994a) for $\hat{f}(v, p)$ and

$\hat{f}(v)$ we can derive the explicit formula for $\hat{m}(v)$. In fact, $\hat{f}(v)$ is estimated by

$$\hat{f}(v) = \frac{1}{nh} \sum_{i=1}^{n} K_h(v - v_i), \tag{2.10}$$

whereas $\hat{f}(v, p)$ by

$$\hat{f}(v, p) = \frac{1}{nh^2} \sum_{i=1}^{n} K_h(v - v_i) K_h(p - p_i), \tag{2.11}$$

where h is the so-called bandwidth and $K_h(t) = K(\frac{t}{h})$ with K being a symmetric probability density function (Wand and Jones, 1994a; Greblicki and Pawlak, 2008). Plugging the above formulas into (2.9) one can define the following Nadaraya-Watson kernel estimate of $m(v)$

$$\hat{m}_{NW}(v) = \frac{\sum_{i=1}^{n} K_h(v - v_i) p_i}{\sum_{i=1}^{n} K_h(v - v_i)}. \tag{2.12}$$

It is important to note that the bandwidth parameter h plays a critical role in the performance of the kernel estimator $\hat{m}_{NW}(v)$ as it controls the level of smoothing of the estimator. There are different methods for choosing the proper value of h. Analogously to the spline method we use 10-fold cross-validation for choosing the bandwidth h in our all experiments.

As it can be concluded from Fig. 2.2 the Nadaraya-Watson kernel estimator does not necessarily preserve the monotonicity property of the curve. Hence, it is necessary to modify the kernel estimator in order to enforce the monotonicity property. In Section 2.3, we introduce a generic method for the monotonicity correction of any curve estimation method. Also we describe a monotone spline regression approach and explain how it can be used to construct monotone wind turbine power curve estimators.

2.3 Monotone Power Curve Estimation

2.3.1 Tilting Method

As we have already argued it is desirable to obtain a monotone (non-decreasing) smooth regression estimator for a wind turbine power curve in the regions that this property is expected. It has been observed that common curve estimation methods do not necessarily preserve this important property (Friedman and Tibshirani, 1984; Bloch and Silverman, 1997; Mammen, 1991; Hall et al., 2001; Dette and Pilz, 2006). In this section, we discuss a method for enforcing the monotonicity property for the Nadaraya-Watson kernel estimator. Although the proposed method is very general and applies to a larger class of kernel and orthogonal basis estimators. The method is called tilting as it is based on the idea of adjusting the observed output variables (power in our case) in order to impose monotonicity on a nonparametric regression estimate. The method was originally introduced in Hall et al. (2001), where some basic statistical properties of the estimate were established (see also (Dette and Pilz, 2006) for some further discussion). Other existing methods that enforce monotonicity reveal some undesirable properties as jumps in the fitted curve (Friedman and Tibshirani, 1984; Bloch and Silverman, 1997; Mammen, 1991).

In order to use the tilting method to monotonize the Nadaraya-Watson kernel estimator, first note that $\hat{m}_{NW}(v)$ in (2.12) can be written in the following form

$$\hat{m}_{NW}(v) = \frac{1}{n}\sum_{i=1}^{n} A_i(v)p_i, \tag{2.13}$$

with the weight functions

$$A_i(v) = \frac{K(\frac{v-v_i}{h})}{\frac{1}{n}\sum_{j=1}^{n} K(\frac{v-v_j}{h})}. \tag{2.14}$$

To impose monotonicity, we introduce a probability weight sequence $\boldsymbol{w} = (w_1, \ldots, w_n)$ to replace the factor $1/n$ in (2.13) and obtain a modified weighted kernel estimator as follows

$$\hat{m}(v|\boldsymbol{w}) = \sum_{i=1}^{n} w_i A_i(v)p_i, \tag{2.15}$$

where the the weight vector $\boldsymbol{w}$ satisfies the following restrictions

$$\sum_{i=1}^{n} w_i = 1 \quad \text{and} \quad w_i \geq 0. \tag{2.16}$$

The weights can be then adjusted so that the resulting estimator is monotonic. In other words, we wish to choose the weight sequence $\boldsymbol{w}$ such that

$$\hat{m}^{(1)}(v|\boldsymbol{w}) \geq 0, \tag{2.17}$$

where $\hat{m}^{(1)}(v|\boldsymbol{w})$ is the first derivative of $\hat{m}(v|\boldsymbol{w})$ with respect to v. It is clear that choosing $\boldsymbol{w}_{unif} = (1/n, \ldots, 1/n)$ results in the original non-monotonic estimate. In order to decide how to adjust the weights $\boldsymbol{w}$ we introduce the distance between $\boldsymbol{w}$ and $\boldsymbol{w}_{unif}$ that should be minimized with respect to v subject to the constraint (2.17). Hence, we wish to choose the least modified version of the estimate $\hat{m}_{NW}(v)$ that meets the monotonicity constraint. Let $\mathcal{D}(\boldsymbol{w}, \boldsymbol{w}_{unif})$ be some distance between $\boldsymbol{w}$ and $\boldsymbol{w}_{unif}$. Then our monotonized estimate is derived from the following constrained minimization problem

$$\min_{\boldsymbol{w}} \mathcal{D}(\boldsymbol{w}, \boldsymbol{w}_{unif}) \quad \text{subject to} \quad \hat{m}^{(1)}(v|\boldsymbol{w}) \geq \delta \tag{2.18}$$

and

$$\sum_{i=1}^{n} w_i = 1 \quad \text{and} \quad w_i \geq 0. \tag{2.19}$$

In (2.18), $\delta \geq 0$ reflects the fact whether we wish to enforce weak ($\delta = 0$) or strict ($\delta > 0$) monotonicity.

The remaining issue is to choose the proper distance $\mathcal{D}(\boldsymbol{w}, \boldsymbol{w}_{unif})$ that would reflect the aforementioned conditions and at the same time would lead to the computationally efficient optimization problem. Following Cressie and Read (1984) (see also (Hall et al., 2001)) we choose the following quadratic distance

$$\mathcal{D}(\boldsymbol{w}, \boldsymbol{w}_{unif}) = \frac{1}{2} \sum_{i=1}^{n} (n w_i)^2 - \frac{n}{2}. \tag{2.20}$$

It is clear that the minimum of this distance subject to (2.19) is obtained by

$\boldsymbol{w} = \boldsymbol{w}_{unif}$. Also since for any vector $\boldsymbol{w}$ satisfying (2.19) we have

$$\frac{1}{n} \leq \sum_{i=1}^{n} w_i^2 \leq 1,$$

therefore, we obtain the following bound for the distance $\mathcal{D}(\boldsymbol{w}, \boldsymbol{w}_{unif})$:

$$0 \leq \mathcal{D}(\boldsymbol{w}, \boldsymbol{w}_{unif}) \leq \frac{n(n-1)}{2}.$$

The lower bound is reached at $\boldsymbol{w} = \boldsymbol{w}_{unif}$, whereas the upper bound at any weight vector $\boldsymbol{w}$ that has value one on a single position and other values are zero. These facts confirm that $\mathcal{D}(\boldsymbol{w}, \boldsymbol{w}_{unif})$ can be employed as the proper distance between $\boldsymbol{w}$ and $\boldsymbol{w}_{unif}$.

The optimization problem in (2.18) equipped with the distance in (2.20) takes the form of the quadratic programming with linear constrains. In fact, the derivate constraint in (2.18) can be evaluated by recalling the definition of $\hat{m}(v|\boldsymbol{w})$ in (2.15). Hence, we have

$$\hat{m}^{(1)}(v|\boldsymbol{w}) = \sum_{i=1}^{n} w_i A_i^{(1)}(v) p_i, \tag{2.21}$$

where this should be evaluated on the pre-selected set of grid points over the range of the variable v. The derivate in (2.21) is indeed a linear function of the weight sequence $\boldsymbol{w}$. There is also an efficient way of determining the derivative of $A_i(v)$ defined in (2.14). In fact, by denoting

$$K_i(v) = K\left(\frac{v - v_i}{h}\right)$$

and after direct algebra we can show that

$$A_i^{(1)}(v) = \frac{K_i^{(1)}(v) \sum_{j=1}^n K_j(v) - K_i(v) \sum_{j=1}^n K_j^{(1)}(v)}{\frac{1}{n}(\sum_{j=1}^n K_j(v))^2}. \tag{2.22}$$

This formula can be efficiently evaluated for some smooth kernel functions. In particular, if $K(v)$ is the standard Gaussian density then we obtain the following version of (2.22)

$$A_i^{(1)}(v) = A_i(v) \left(\frac{1}{n} \sum_{j=1}^n \alpha_j(v) A_j(v) - \alpha_i(v) \right), \tag{2.23}$$

where $\alpha_i(v) = \frac{v - v_i}{h^2}$. This is due to the fact that for $K(v)$ being the standard Gaussian density we have $K^{(1)}(v) = -vK(v)$.

Solving the above mentioned optimization problem is straightforward and easy to implement. We use quadratic programming for this purpose. To implement the tilting approach we use *quadprog* library in R. You can find details of implementation in Appendix.

Although we developed our proposed method using the Nadaraya-Watson kernel estimator, the proposed monotonizing method is applicable to other general kernel estimator methods by modifying them principally in regions where they are not monotone. This method can also be used to impose shape constraints other than just monotonicity, for example, if some theoretical wind turbine power curves are decreasing in some specific regions of wind speed, this method can be used to impose such behavior for desired regions.

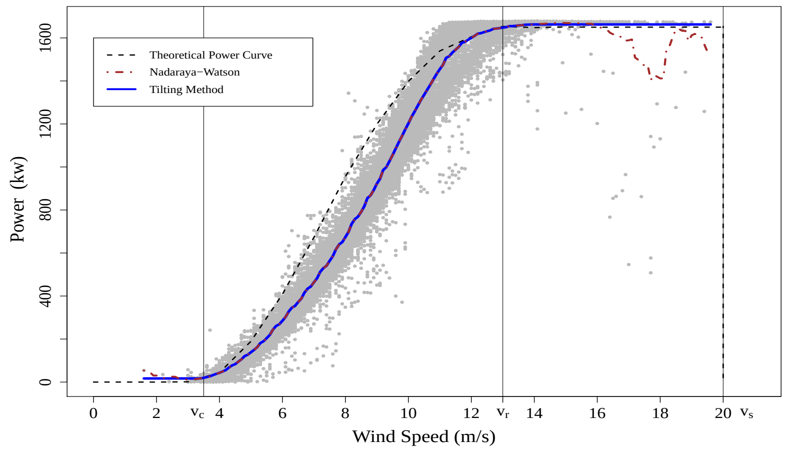


Figure 2.3: Effect of tilting method on the Nadaraya-Watson kernel estimator. Nadaraya-Watson fitted power curve is shown with dashed line. Applied tilting method power curve is shown with solid line. Long dashed line shows the theoretical power curve.

Fig. 2.3 shows the Nadaraya-Watson kernel estimator and its corresponding monotonized curve. One can notice that in monotone regions both fitted curves are essentially the same, however, in non-monotone regions there exist some differences between fitted curves.

2.3.2 Monotone Regression Spline

Polynomial regression spline is introduced as a piecewise polynomial with specific continuity constraints. The continuity characteristic of the spline and the number of independent parameters depend on the number and location of knot sequence $\boldsymbol{\zeta}$. These knots divide the interval $[L, U]$ into some regions. Spline models use

$\boldsymbol{\zeta}$ and define a set of basis functions which weighted linear combination of these basis functions can be used as spline regression. Spline regression models do not necessarily retain the monotonicity property of the wind turbine power curves (e.g., Fig. 2.2). One approach to impose monotonicity for such models is to use other basis functions such that the resulting fitted curve is monotone (e.g. I-spline).

The earliest basis function in spline regression was the simple truncated power function as: $\phi_{jk}(v) = [(v - \zeta_j)_+]^k$ where $u_+ = \max\{u, 0\}$. The most important advantage of this basis function which makes it so attractive for statistical work is its simplicity. On the other hand, its considerable rounding error for large value of K, number of knots, is a disadvantage of this basis function. In addition, truncated power functions do not seem to have natural interpretation in some applications.

Another basis function which is so attractive for statisticians is the M-spline basis function, because of its properties, e.g. normalization $\int M(v_i)dv = 1$, $i = 1, \cdots, n$, and the fact that they are defined such that they are positive in (ζ_i, ζ_{i+k}), and 0 elsewhere.

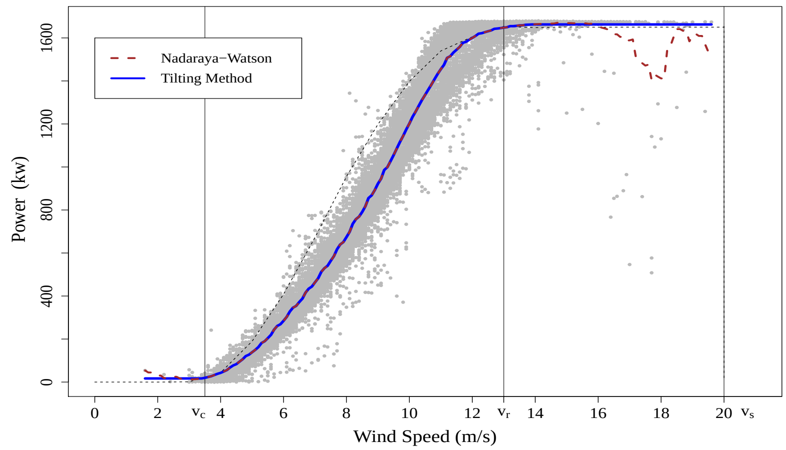
For $k = 1$

$$M_i(v|1, \zeta) = \begin{cases} \frac{1}{\zeta_{i+1} - \zeta_i}, & \zeta_i \leq v < \zeta_{i+1} \\ 0, & otherwise \end{cases} \tag{2.24}$$

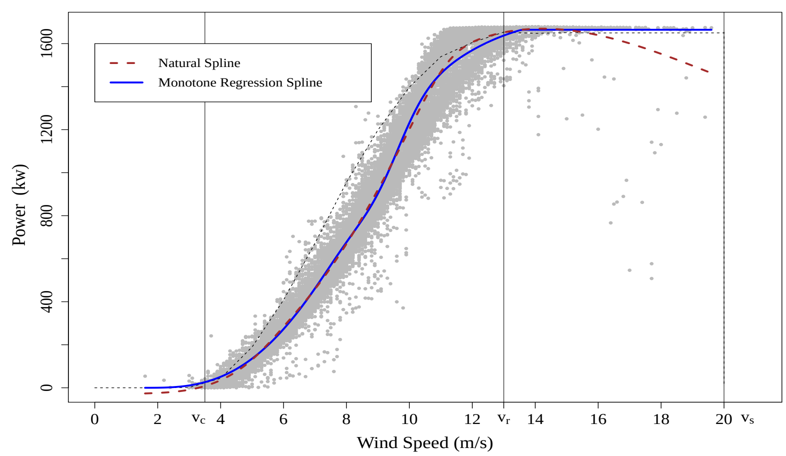
and for $k > 1$, we have

$$M_i(v|k, \zeta) = \frac{k[(v - \zeta_i)M_i(v|k-1, \zeta) + (\zeta_{i+k} - v)M_{i+1}(v|k-1, \zeta)]}{(k-1)(\zeta_{i+k} - \zeta_i)}. \tag{2.25}$$

The advantage of M-spline compared with truncated power is controlling rounding



(a)



(b)

Figure 2.4: Four power curve fitting methods on real data from wind turbines in a wind farm located in Canada a) Nadaraya-Watson Kernel Estimator (NWKE) and Tilting Method (TM) applied to NWKE b) Natural Cubic Spline (NCS) and Monotone Regression Spline method (MRS).

error in the computations. Since the value of m_i is zero outside the interval $[\zeta_i, \zeta_{i+k}]$ and positive within it, non-negativity of $f(v) = \sum a_i M_i(v)$ is satisfied by assuring

$a_i \geq 0$. However still using M-spline we can not satisfy the monotonicity requirement, so we need to define a monotone basis function. An obvious approach is to use integrated spline, I-spline, as a modification of M-spline basis function,

$$I_i(v|k,\zeta) = \int_L^v M_i(u|k,\zeta)du, \tag{2.26}$$

which satisfying $a_i \geq 0$ in $f(v) = \Sigma a_i I_i(v|k,\zeta)$ will result in a monotone spline regression curve. We should mention that if m_i is a piecewise polynomial of degree $k-1$ so I_i will be a piecewise polynomial of degree k. We can rewrite I_i in the following form:

$$I_i(v|k,\zeta) = \begin{cases} 0, & i > j, \\ \sum_{m=i}^{j}(\zeta_{m+k+1} - \zeta_m)\frac{M_m(v|k+1,\zeta)}{(k+1)}, & j-k+1 \leq i \leq j, \\ 1, & i < j-k+1. \end{cases} \tag{2.27}$$

2.4 Real Data Application

In this section, we apply our proposed methods in previous sections on proprietary wind power data of a wind farm in North America. The wind power plant (WPP) includes over five dozen identical wind turbines. Specification of these turbines are as follow: the rated capacity is 1.7 MW, the hub height is 80 m spread over an area of over 90 km^2, the cut-in speed is 3.5 m/s, rated speed is 13 m/s and cut-out speed is 20 m/s. We have selected four wind turbines (T1,...,T4) from the WPP to analyze the performance of the presented methods. Our raw data contains 30398 of

Table 2.1: Results for 10-Min and hourly averaged data for turbine T1, T2, T3 , and T4 for the methods Natural Spline (NS), Tilting Method (TM), Nadaraya-Watson Kernel Estimator (NWKE), and Monotone Regression Spline (MRS)

Time	Method	Turbine T1		Turbine T2		Turbine T3		Turbine T4		Rank
		RMSE	NMAPE	RMSE	NMAPE	RMSE	NMPAE	RMSE	NMAPE	
Hourly	NCS	76.83 (2)	3.26 (2)	84.29 (2)	3.20 (1)	103.94 (4)	3.8 (2)	78.65 (2)	3.43 (2)	(2)
	MRS	75.53 (1)	3.20 (1)	84.46 (3)	3.20 (1)	103.90 (3)	3.8 (2)	78.60 (1)	3.42 (1)	(1)
	NWKE	78.77 (4)	3.33 (4)	84.29 (2)	3.25 (3)	103.06 (1)	3.77 (1)	78.85 (4)	3.44 (3)	(3)
	TM	78.21 (3)	3.32 (3)	83.92 (1)	3.34 (4)	103.18 (2)	3.83 (3)	78.71(3)	3.47 (4)	(4)
10-Min	NCS	73.31 (1)	2.96 (2)	83.43 (2)	3.26 (2)	85.54 (2)	3.25 (1)	84.51 (1)	3.38 (1)	(2)
	MRS	73.34 (3)	2.95 (1)	83.41 (1)	3.25 (1)	85.51 (1)	3.25 (1)	84.51 (1)	3.38 (1)	(1)
	NWKE	73.32 (2)	2.96 (2)	83.41 (1)	3.26 (2)	85.61 (3)	3.26 (2)	89.94 (2)	3.52 (2)	(3)
	TM	73.39 (4)	2.95 (1)	83.60 (3)	3.32 (3)	85.7 (4)	3.27 (3)	89.94 (2)	3.55 (3)	(4)

wind speed and generated power data points with a resolution equal to 10-minutes. In addition, we took average of these 10-minutes data in order to evaluate our proposed methods on hourly data.

Since there are outlier data, data preprocessing is required to filter the invalid data points. Following Shokrzadeh et al. (2014) we polished the data using the outlier detection process. Fig. 2.4 shows the four models proposed in this paper applied on the filtered wind speed and the generated wind power data. Issue of non-monotonicity of the Nadaraya-Watson kernel estimator and natural spline is obvious in Fig. 2.4. On the other hand, Fig. 2.4 shows that the tilting method and the monotone regression method result in curves that are more similar to manufacturer power curve as both are monotone in all regions. Consequently, since these two power curve estimators are similar to the manufacturer one, it makes more sense to use these methods in practice as they are more appealing for the interpretation and analysis of wind turbine behavior.

Root-mean-squared error (RMSE), normalized mean absolute percentage error (NMAPE), symmetric mean absolute percentage error (sMAPE), the mean absolute error (MAE) and the coefficient of determination (R2) are several statistical metrics that can be used as appropriate measures of performance for the fitted power curves (Marvuglia and Messineo, 2012). In this paper, we use RMSE and NMAPE given by

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - \hat{p}_i)^2}, \tag{2.28}$$

and

$$NMAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{|p_i - \hat{p}_i|}{\max_{j=1}^{n}\{\hat{p}_j\}} \times 100, \tag{2.29}$$

where p_i is the observed wind power and $\hat{p}_i$ is the estimated value of the power. By performing the error analysis, the values of RMSE and NMAPE for the Nadaraya-Watson Kernel Estimator (NWKE), Tilting Method (TM), Natural Cubic Spline (NCS), and Monotone Regression Spline (MRS) methods are presented here. Table 2.1 shows the results of the analysis for two data sets with different time resolution. These two data sets contain 10-min and hourly averaged data since June 2006 till January 2007. Table 2.1 also ranks the performance of all four methods based on the calculated measures, where the smaller values are desirable, and also provides the overall performance ranking of each method. Numbers in the parenthesis show the rankings. As shown in Table 2.1, the monotone regression spline method outperforms all other methods addressed in this study. Table 2.1 also suggests that natural spline regression is dominating the Nadaraya-Watson kernel estimator and the tilting method. It also shows that tilting method performs almost the same as the Nadaraya-Watson kernel estimator, however, this method performs more similar to the manufacturer power curve which makes it more realistic and applicable for wind power systems and also in interpretation. It also worths to mention that we experimented the effect of bandwidth for kernel estimator and tilting methods which conclude that both methods will have best performance with the same bandwidth. It means that tilting method will perform its best with the same bandwidth as the one obtained in the kernel method.

2.5 Concluding Remarks

Some applications of wind turbine modeling include monitoring of the turbine's performance, sizing the storage capacity for wind power integration, as well as power forecasting. We have presented four regression modeling approaches for estimating wind turbine power curves with a special focus on imposing the monotonicity on fitted curves. We have shown that natural spline and polynomial regression models do not necessarily preserve the monotonicity property, however their power curve estimation could be very accurate. Tilting method and monotone spline regression are introduced to address the issues of monotonicity in power curve estimation and reduce the sensitivity to outliers within the observations. Tilting method, applied to the Nadaraya-Watson kernel estimator and monotone spline regression are then examined to achieve monotonicity property for fitted power curves. The performance of these proposed methods are evaluated based on operational wind power data for a wind farm in Canada. We analyzed the performance of the presented methods on four different wind turbines. The results of this study suggest that monotone spline regression presents a better performance over the other analyzed methods. Also results show that using tilting method compared to kernel methods results in almost equal accuracy but a shape similar to the manufacturer power curve. As discussed in Elliott and Infield (2014), applying insufficient time averaging on the data reduces the correlation between measured wind speed and measured power. This issue is called "errors in bins". In this study, we evaluated the 10-Min and hourly averaged data. Similar to Elliott and Infield (2014), our experiments show that the variation of averaging time has little overall impact upon the general shape of the fitted power curve. In practice, we recommend to fit constrained and unconstrained wind turbine power curve models and decide which model to use for further analysis by using proper performance measurements and other reasonable criteria. Some variables other than just wind speed including turbine age, wind direction as well as location of turbine may have relationship with the generated power at the location of the wind turbine. Some of these variables might indeed cause the generated power to be lower than the expected theoretical one. An interesting direction of future research is to incorporate more covariates in wind turbine power curve modeling and extend the proposed methodology to a multiple covariate case.

Appendix

To be more specific about details of the implementation of the tilting approach toward monotonicity, we show the procedure in the matrix format. First, we have a set of pairs (v_i, p_i), then we fit the Nadaraya-Watson kernel estimator as our regression function. The estimated value of the fitted method is not necessarily monotone in the whole range of the predictor. So we find the regions which contain decreasing values of the fitted response. Meanwhile we define $w_i = 1/n$ for all distinct values in the range of v. After finding decreasing regions we try to modify w_j, j in decreasing regions and small neighborhood around them, while keeping w_i related to monotone regions unchanged. This will lead to monotonicity in the whole range of v. To implement the tilting approach we use *quadprog* library in R by first writing our optimization as follow:

$$\min(\frac{1}{2}\Sigma_{i=1}^{n}(nw_i)^2 - \frac{n}{2}), \tag{2.30}$$

subject the probability constraint on w and the monotonicity constraints imposed on the weighted kernel estimate. This criterion function can be written in the matrix form as

$$\min(\frac{1}{2}\boldsymbol{w}^\top Q\boldsymbol{w}), \tag{2.31}$$

with constraints $A^T\boldsymbol{w} \geq b$, where $Q = diag[n^2]$ is the $n \times n$ diagonal matrix. Note that to define A one can add any other constraints which might be necessary, e.g. the constraint that maximum value of the fitted power curve should not be greater than the maximum possible value of the generating power related to the wind turbine.

Let us now examine the required constraints. After considering all constraints, A will be in the form: $A = [a_1, a_2, a_3, a_4]$ where a_1 is a $n \times 1$ vector such that its entries associated with indexes in decreasing regions are equal to 1 and the rest are 0. a_1 is related to the constraint $\sum_{i\in\text{DR}} w_i = 1 - \sum_{i\in\text{NDR}} w_i$ where DR is the decreasing range and NDR is the non-decreasing range. b vector according to first constraint contains the value of $1 - \sum_{i\in\text{NDR}} w_i$. a_2 is a $n \times n$ matrix related to the constraint that $w_i = \frac{1}{n}$ for all i in monotone regions. So a_2 is a diagonal matrix with all values equal to 0 except for those indexes which are in monotone regions and their corresponding values will be 1. The values of vector b for those indexes will be $\frac{1}{n}$ and the rest 0. a_3 is an $n \times (n+1)$ matrix

$$a_3 = \begin{pmatrix} -f_1 & f_2 & 0 & \cdots & 0 \\ 0 & -f_2 & f_3 & \cdots & 0 \\ \vdots & \ddots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & f_n \end{pmatrix},$$

where f_i is the fitted value of v_i based on the Nadaraya-Watson estimator. Associated value in vector b will be ϵ which means for all values the constraint $w_i f_i - w_{i-1} f_{i-1} \geq \epsilon$ should be satisfied. By choosing an $\epsilon > 0$ one can impose the strictly monotone property to the regression function and if $\epsilon = 0$ the resulting regression function will be non-decreasing.

a_4 is also an $n \times n$ diagonal matrix which satisfies the constraint $w_i \geq 0$ with their corresponding values in b vector being 0. a_5 is related to the constraint that $w_i f_i < \max(p)$ where $\max(p)$ is the maximum possible generating power value of the wind turbine. Consequently at the end, matrix A will be a $n \times (4n+2)$ matrix and b a $1 \times (4n+2)$ vector.

The final result of the quadratic programming will be a set $\boldsymbol{w}$ which satisfies all the constraints with respect to minimizing the distance of $\boldsymbol{w}$ to uniform distribution $(\frac{1}{n}, \cdots, \frac{1}{n})$ given that there is any valid solution. In summary, the tilting approach in practice can be implemented using the following steps:

1. Fit any kernel estimator method based on wind-power data.

2. Find all regions which fitted model is non-monotone.

3. Define $\boldsymbol{w} = \boldsymbol{w}_{uniform}$, matrix A and vector b as explained.

4. Using quadratic programming, try to find $\boldsymbol{w}$ which satisfies all constraints including the monotonicity property and minimizes the distance function.

5. If no valid $\boldsymbol{w}$ is available then expand the region around non-monotone regions and return to 3.

6. $\boldsymbol{w}$ is selected.

Chapter 3

Hybrid Power Curves

This chapter encloses a peer-reviewed journal paper published in *Energy*[1] Journal that consists of the author's work during his Ph.D. study.

Abstract: Wind turbine power curves are often used for monitoring the performance of wind turbines and play an important role in wind power forecasting. Various factors such as age of turbines, installed location, air density, wind direction and measurement errors cause non-homogeneity among observed data, which often influences the accuracy of fitted power curves. To overcome this problem, a hybrid estimation approach is proposed, based on weighted balanced loss functions that account for both estimation error and goodness of fit by shrinking estimates toward

[1]Mehrdad Mehrjoo, Mohammad Jafari Jozani, and Miroslaw Pawlak. "Toward hybrid approaches for wind turbine power curve modeling with balanced loss functions and local weighting schemes." Energy 218 (2020). https://doi.org/10.1016/j.energy.2020.119478

standardized target models. Two different weighting schemes are developed to incorporate the non-homogeneity of data in the estimation process. The proposed algorithm is compared with commonly used power curve estimation methods based on classical least squares, natural splines, and local linear smoothing methods. The performance of the original and proposed hybrid methods is evaluated using a historical data set collected from a wind farm located in Canada. Results show that hybrid methods can be used as an effective tool to improve the performance of existing power curve modeling approaches. Consequently, proposed methods can facilitate more robust monitoring the performance of wind turbines as well as wind power forecasting.

3.1 Introduction

Wind power is an important source of sustainable energy that has significantly been used in many countries to provide reliable, clean, and relatively cheap mechanical power and electricity. Nowadays, wind energy encompasses a large portion of power systems in several countries (Ulazia et al., 2019). Significant penetration of wind power causes uncertainties in stability of power systems. Wind power fluctuation and wind speed variability have negative impacts on the reliability of power systems (Li and Bagen, 2010; Lydia et al., 2014). In such situations, accurate modeling of the relationship between the wind speed and the power generated by a wind turbine can play an important role in improving the performance of power systems that use wind energy. Most manufacturers provide graphs or tables to describe the underlying relationship between the wind speed and output power of their wind turbines, known as the wind turbine power curve (WTPC) (Raj et al., 2011).

Even though the manufacturer's power curves (MPC) are often available for wind turbines, practitioners prefer using empirical data to estimate them. This is mostly done because MPCs are obtained in a test site with specific environmental conditions that do not necessarily hold in wind farm environment (Commission et al., 2007). Also, the performance of a turbine is affected by many other factors (e.g., it declines rapidly with age), which makes the MPCs less reliable sources for power forecasting (Staffell and Green, 2014).

Accurate WTPC modeling is a very important area in power management and plays an important role in increasing the robustness of the power systems and, ultimately, reducing the system's operational costs (Chang et al., 2014). Using accurate power curve models will also contribute to easier, more reliable and faster forecasts of the wind generation. An accurate power curve model also is required to more realistically size the storage capacity for wind energy integration (Shokrzadeh et al., 2015; Nabat et al., 2020). These models also can aid in the early identification of faults, such as blade faults and yaw and pitch system faults (Park et al., 2014). Each type of these faults will cause the generated power differs from the expected value from the power curve in a different way. Statistical analysis of the outlier data can give indications of the specific reason of anomaly if the empirical power curve is fitted correctly (Kusiak and Li, 2011).

To obtain an empirical WTPC model based on the observed data, one can use parametric or non-parametric methods (Shokrzadeh et al., 2014; Greblicki and Pawlak, 2008). Parametric methods are prevalent in the literature and include segmented linear models (Lydia et al., 2013), polynomial regression (Lydia et al., 2014), cubic power curve, models based on probabilistic distributions such as four or five parameters logistic distributions and neural networks (Saint-Drenan et al., 2020; Carrillo et al., 2013; Villanueva and Feijóo, 2016a). These methods often suffer from the inherent model misspecification error and their limited ability to describe the dynamic nature of the power curves (Ouyang et al., 2017). On the other hand, non-parametric methods do not need any prior knowledge about the shape of the power curve and make minimal assumptions about the relationship between wind speed and generated power (Pandit et al., 2019). These techniques contain kernel smoothing algorithms, regression trees, k-nearest neighbor methods, and spline regression procedures (Shokrzadeh et al., 2014; Mehrjoo et al., 2019; Marvuglia and Messineo, 2012; Karamichailidou et al., 2020).

There are a number of issues that might influence the accuracy of an empirical WTPC model. For example, the MPCs are suitable for predicting the power output of a single turbine of a specific type. In a big wind farm, a number of turbines are spread over a wide area Sohoni et al. (2016b). Wind energy production of each turbine depends on wind condition at different parts of the farm, air density, turbine condition, and location.

Another important factor is due to possible inconsistencies in the observed data that can cause serious problems in the WTPC model fitting process. For example, non-homogeneity of generated power in different regions of wind speed often results in observations that are far from what one might expect based on the MPC. Also, one should note that wind is highly stochastic in nature and changes with height. Most power curve model fitting methodologies use generated powers associated with wind speeds that sometimes are not measured at turbines' location. This will influence the accuracy of empirical WTPC models and make some of the fitted curves less reliable, especially for those that are far from meteorological stations (Sohoni et al., 2016b). To solve this issue, some researchers use filtering methods to identify and remove suspicious data before performing any model fitting approaches (Ouyang et al., 2017; Taslimi-Renani et al., 2016; Wang et al., 2020). However, even after filtering inconsistent data, often the variances of the estimated powers might significantly vary with wind speed, indicating heteroscedasticity of the observed data (Wang et al., 2018). In other words, there is no guarantee that all the outliers will be removed from the raw wind data. Consequently, designing robust regression models or finding a curve that is close to a robust target WTPC function may be considered as two effective ways to obtain accurate power curves in the presence of outliers and data non-homogeneity.

To overcome the aforementioned problems, a novel WTPC modeling approach is proposed using weighted balanced loss functions that account for both estimation error (using observed data) and goodness of fit that reflects the proximity to a target model. As it is expected, turbines that are close to each other and share similar characteristics, such as age, model, installed location (i.e., same altitude and latitude), should have comparable power output. Hence, one might want to find an empirical WTPC for a specific turbine using its measured data and, at the same time, shrink the fitted curve toward a more robust and sometimes a more reliable target function representing the relationship between output power and wind speed. The target function could be the MPC, the WTPC of the wind farm, or the one obtained using empirical data associated with wind turbines in the wind farm that are in the vicinity of the underlying turbine.

Proposed methodology results in WTPC estimates that can be represented in the form of shrinkage estimators. They are designed to provide hybrid estimates of WTPC models that can work with both parametric and nonparametric methods. This reveals another application of the balanced loss functions that are mostly used in parametric estimation problems with a target or a set of target models in mind (Zellner, 1994). Balanced loss functions are appealing as they combine the proximity of a given estimator to a target value as well as the unknown parameter of interest (Marchand and Strawderman, 2020). These loss functions have captured the interest of some researchers for regression problems (Hu and Peng, 2011), estimation and prediction problems (Jafari Jozani et al., 2012, 2006), as well as credibility theory, finance, sequential estimation (Baran and Stepien-Baran, 2013; Zhang and Chen, 2018).

WTPC modeling using balanced loss functions provides more reliable estimates of the underlying relationship between wind speed and generated power of each turbine. Also, developed methodology allows to obtain estimates that are more robust to outliers by giving more weights to more reliable target values in wind speed regions that observed data are very noisy or measurement errors are prevalent. To do this, two different weighting schemes are developed in this paper to deal with non-homogeneity and, in particular high variability of generated power of each turbine in different wind speed regions. These weights are then used to balance the importance of estimation error and proximity to target values in the construction of WTPCs. It is worth mentioning that proposed weighting scheme can be used to extend the commonly used WTPC modeling methods under the least squares methodology to a more complex one under the weighted least squares approach. This will resolve the effect of the possible data non-homogeneity issue of generated powers in different wind speed regions for the standard WTPC modeling methods developed in the literature. Also, the proposed weighting scheme is very flexible, and by taking the weight function to be zero, one can obtain the usual results under the least squares as a special case.

The outline of this paper is as follows. Section 3.2 introduces the problem of WTPC modeling and the proposed hybrid estimation method. It explains the need for taking into account the proximity of a fitted WTPC to a target model and shows how a hybrid power curve estimate can be obtained using the proposed balanced loss function. Section 3.3 defines different weighting schemes to be used in conjunction with the proposed loss function for WTPC modeling. Section 3.4 elaborates on two different target models, whereas the performance of each method is assessed in Section 3.5. Finally, Section 3.6 provides concluding remarks and future work.

3.2 Wind Turbine Power Curve Estimation

To motivate the statistical methodology proposed in this paper a real data set obtained from a wind farm in Canada was studied. Fig. 3.1 shows a MPC along with real historical wind speed-power data. Also, three important points are highlighted in Fig. 3.1: 1) Cut-in speed (V_c) being the threshold which wind turbine starts generating power when wind speed is greater than this value; 2) Rated speed (V_r) is the threshold which wind turbine increasingly generates power till wind speed reaches this threshold; 3) Cut-out speed (V_s) is the threshold which turbine operation stops when wind speed is higher than this value to avoid damages Gasch and Twele (2011); Manwell et al. (2010). Wind turbine manufacturers release power curves for existing machines, derived using field tests. The approximate shape of the power curve for a given machine can also be estimated theoretically using the power characteristics of the rotor, generator, gearbox ratio, and efficiencies of various components. The following function is often used to represent the relationship between theoretical wind power and wind speed (v in m/s).

$$P(v) = \frac{1}{2} C_p \rho A v^3, \tag{3.1}$$

where $v \in [V_c, V_s]$, C_p is the efficiency of the whole wind turbine system which is usually calculated experimentally, ρ is the air density in kg/m^3, and A is the turbine rotor area in m^2 (Manwell et al., 2010; Walker and Jenkins, 1997; Ackermann and Söder, 2000).

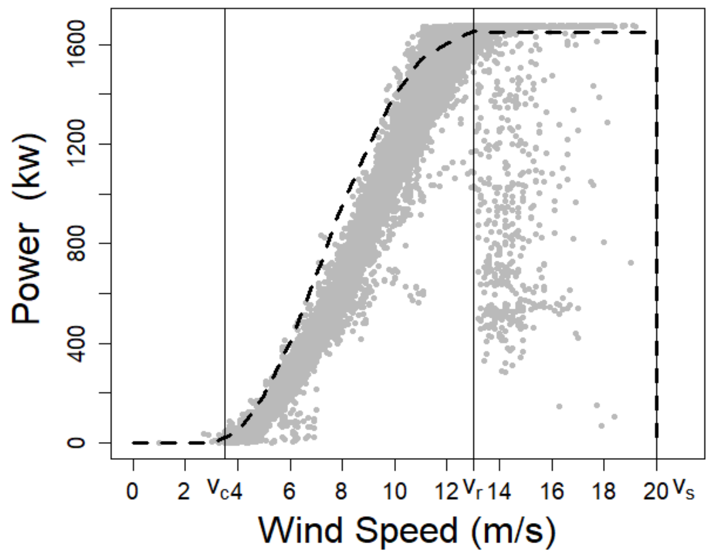


Figure 3.1: Manufacturer power curve (MPC) along with the cut-in (V_c), rated (V_r) and cut-out (V_s) thresholds as well as 10-minutes historical speed-power data.

In practice, due to several factors such as environmental conditions that are different from those in test sites, turbine aging, air density, wind velocity distribution, wind direction, and the existence of obstacles near turbines, generated wind power might differ from what is expected by MPC. Thus, empirical power curve modeling is often used instead.

Let $P = f(v)$ represent the unknown power curve. In practice, $f(v)$ is unknown, but some observations of its noisy version at some given wind speed values are available, i.e.,

$$Y_i = f(v_i) + \xi_i, \quad i = 1, \dots, n, \tag{3.2}$$

where $P_i = f(v_i)$ is the true value of the power at the speed level v_i and ξ_i is the noise process. Thus, the problem is to estimate $f(v)$ from the data set

$\{(v_1, Y_1), \ldots, (v_n, Y_n)\}$, representing the real historical power dataset, also known as training data. Existing literature on WTPC modeling often uses the goodness of the fitted curves to the observed data as the only optimality criterion. For example, under the classical least squares method, a fitted power curve is obtained by minimizing the sum of squares

$$\widehat{\mathbf{P}} = \arg \min_{\mathbf{P} \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^{n} (Y_i - P_i)^2, \tag{3.3}$$

where $\widehat{\mathbf{P}} = (\widehat{P}_1, \ldots, \widehat{P}_n)^\top$ are the values of the estimated power curve, such that $\widehat{P}_i$ is the estimated power value at the speed level v_i. In (3.3), the minimization is taken with respect to $\mathbf{P} = (P_1, \ldots, P_n)^\top$ that belongs to a restricted class $\mathcal{M}$ of power curve models. For instance, $\mathcal{M}$ can represent polynomial models of degrees q, i.e., $\mathcal{M} = \{f(v) | f(v) = \Sigma_{k=1}^{q} a_k v^k\}$.

In practice, there might exist reliable estimates P_i^* of P_i obtained from other sources. For example, P_i^* could be the estimated power curve for a turbine T' that is very close to the underlying turbine T in the wind farm such that T and T' are sharing similar characteristics. Another choice for P_i^* would be the value of P at $v = v_i$ obtained through the theoretical power curve, MPC or the average value of P generated in the wind farm at $v = v_i$. One might also want to use the estimated P_i obtained through robust methods after removing outliers using one of the standard techniques for polishing data or manually polishing by some experts. This is the motivation to define the following locally balanced loss function

$$\mathcal{L}(\mathbf{P}) = \frac{1}{n}\sum_{i=1}^{n}\left[(1-\alpha(v_i))(Y_i - P_i)^2 + \alpha(v_i)(P_i^* - P_i)^2\right], \tag{3.4}$$

where P_i^* is the target value and the weight $\alpha(v_i)$ takes values in $[0, 1]$. This loss function reflects a desire of closeness of P_i both to: (i) the observed data value Y_i in terms of the squared error loss, and (ii) the target value P_i^* in terms of the squared distance. The weight $\alpha(v_i)$ calibrates the relative importance of these two components. The minimum of the loss $\mathcal{L}(\mathbf{P})$ with respect to $\mathbf{P}$ over a class of models $\mathcal{M}$ defines the modified estimate $\widehat{\mathbf{P}}$ that is able to take into account a prior knowledge defined by the target model P_i^* (Jafari Jozani et al., 2012).

Setting $\alpha(v_i) = 0$ for each i results in the usual power curve models under the sum of squared errors that totally ignores the target model in the estimation process. On the other hand, by setting $\alpha(v_i) = 1$ for each i, power curve estimator will be the same as the target model. By decreasing the value of $\alpha(v_i)$, the impact of the target model on the hybrid power curve estimator will be reduced.

One can employ commonly used model fitting approaches to obtain optimum hybrid WTPCs under the balanced loss function in (3.4). Polynomial regression and some of its generalizations such as spline regression functions are widely used for power curve modeling. A spline regression function is a piecewise polynomial regression with the constraints that the first $K-1$ derivatives of the fitted curve be continuous. K denotes the degree of the polynomial fitted on each piece. Spline regression methods can be constructed using different basis functions. In this paper, a B-Spline regression model is utilized. In other words, suppose $\mathcal{M}_B = \{\mathbf{P} \in \mathbb{R}^n : \mathbf{P} = \mathbf{B}\boldsymbol{\beta}\}$, is the underlying class of B-Spline models, where $\mathbf{B}$ is the $n \times n$ matrix with its (i, j)'th element being $B_j^k(v_i)$ given by

$$B_j^k(v_i) = \frac{(v_i - \zeta_j)}{(\zeta_{j+k} - \zeta_j)} B_j^{k-1}(v_i) + \frac{(\zeta_{j+k+1} - v_i)}{(\zeta_{j+k+1} - \zeta_{j+1})} B_{j+1}^{k-1}(v_i), \tag{3.5}$$

for $j = -k, -k+1, ..., K$, $\zeta_0 = \zeta_{-1} = \cdots = \zeta_{-k} = \min\{v_i, i = 1, \cdots, n\}$, $\zeta_{K+1} = \max\{v_i, i = 1, \cdots, n\}$, and $\{B_j^0(v_i)\}$ are the natural basis functions for piecewise constant approximation (Friedman et al., 2001).

The Observed data set contains samples $\{(v_i, Y_i), i = 1, \ldots, n\}$, where v_i is the wind speed and $Y_i = P_i + \xi_i$ is the corresponding noisy version of power. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ be the vector of the observed powers. Consider the usual quadratic loss function and suppose that one is interested in the solution $\mathbf{P} \in \mathcal{M}_B$ that minimizes $\frac{1}{n}\sum_{i=1}^n (Y_i - P_i)^2$. This is equivalent for finding the following estimate of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^n (Y_i - \boldsymbol{B}_i^\top \boldsymbol{\beta})^2, \tag{3.6}$$

where $\boldsymbol{B}_i$ is the $i-$row of the matrix $\mathbf{B}$. This results in the solution

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top \boldsymbol{Y}, \tag{3.7}$$

yielding the prediction power model $\widehat{\mathbf{P}} = \boldsymbol{B}\widehat{\boldsymbol{\beta}}$ (Friedman et al., 2001).

Suppose $\boldsymbol{P}^* = (P_1^*, \ldots, P_n^*)^\top$ is a vector of given target power values. This additional information can be employed to obtain a modified estimate of $\boldsymbol{\beta}$ under the balanced loss function resulting in power estimates that are close to both the target curve and the observed data. Hence, one can calculate the coefficients of the spline model by minimizing the following locally balanced loss function:

$$\mathcal{L}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\{(1-\alpha(v_i))(Y_i - \boldsymbol{B}_i^\top\boldsymbol{\beta})^2 + \alpha(v_i)(P_i^* - \boldsymbol{B}_i^\top\boldsymbol{\beta})^2\}. \tag{3.8}$$

The estimate $\widehat{\boldsymbol{\beta}}$ that minimizes this criterion yields the modified predictive power model $\widehat{P}_i = \boldsymbol{B}_i^\top\widehat{\boldsymbol{\beta}}$. In (3.8), P_i^* is the target value and the weight $\alpha(v_i) \in [0,1]$ will be chosen based on a suitable weighting scheme. By dropping the factor $1/n$, the formula in (3.8) can be re-written in the matrix form as follows

$$\mathcal{L}(\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{\beta})^\top(\mathbb{I} - \boldsymbol{\alpha})(\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{\beta}) + (\boldsymbol{P}^* - \boldsymbol{B}\boldsymbol{\beta})^\top\boldsymbol{\alpha}(\boldsymbol{P}^* - \boldsymbol{B}\boldsymbol{\beta}), \tag{3.9}$$

where $\boldsymbol{\alpha}$ is the diagonal matrix $\boldsymbol{\alpha} = \text{diag}[\alpha(v_1), \cdots, \alpha(v_n)]$ and $\mathbb{I}$ is the diagonal unit matrix. The formula in (3.9) can be easily minimized with respect to $\boldsymbol{\beta}$ by first taking the derivative of $\mathcal{L}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$,

$$\frac{d\mathcal{L}(\boldsymbol{\beta})}{d\boldsymbol{\beta}} = (\mathbb{I} - \boldsymbol{\alpha})(2\boldsymbol{B}^\top\boldsymbol{B}\boldsymbol{\beta} - 2\boldsymbol{B}^\top\boldsymbol{Y}) + \boldsymbol{\alpha}(2\boldsymbol{B}^\top\boldsymbol{B}\boldsymbol{\beta} - 2\boldsymbol{B}^\top\boldsymbol{P}^*). \tag{3.10}$$

and then solving $\frac{d\mathcal{L}(\boldsymbol{\beta})}{d\boldsymbol{\beta}} = 0$, to get:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{B}^\top\boldsymbol{B})^{-1}\boldsymbol{B}^\top[(\mathbb{I} - \boldsymbol{\alpha}) \times \boldsymbol{Y} + \boldsymbol{\alpha} \times \boldsymbol{P}^*]. \tag{3.11}$$

This leads to the following locally combined prediction of the generated powers

$$\begin{aligned}\widehat{\boldsymbol{P}}_{\boldsymbol{\alpha}} &= \boldsymbol{B}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top[(\mathbb{I}-\boldsymbol{\alpha}) \times \boldsymbol{Y} + \boldsymbol{\alpha} \times \boldsymbol{P}^*]. \end{aligned} \tag{3.12}$$

Consequently, using the balanced loss function $\mathcal{L}(\boldsymbol{\beta})$ in (3.8), the estimate $\widehat{\boldsymbol{P}}_{\boldsymbol{\alpha}}$ does not only depend on the observed data $\boldsymbol{Y}$ through the usual B-spline regression fit $\widehat{\boldsymbol{P}} = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top \boldsymbol{Y}$ but also reflects a desire of closeness to the target value $\boldsymbol{P}^*$.

It is known that the B-spline method might behave erratically beyond the boundary knots. To overcome this issue, one can use the natural spline regression models that are similar to B-spline models except that they are constrained to be linear beyond the boundary knots (Friedman et al., 2001). Also, the choice of the right number of knots and their locations play important roles in spline regression modeling. The most common degree of K for spline regression is $K = 3$, which is called cubic splines regression. There exist different methods in the statistical literature for selecting the number of knots and finding the proper location of knots (Friedman and Silverman, 1989).

As wind turbines are expected to generate similar powers at wind speeds around a given point, one can also use a local regression approach to obtain an estimated power at a point v_0 by giving more weights to nearby training observations. The parameter span s defines the number of nearby training observations to be considered for estimating the power at $v = v_0$. Span s plays an important role in the fitting process since it controls the flexibility of the fitted curve. A small value of s leads to a local and wiggly fitted power curve, however large values of s cause the fit to be more smooth yielding a global fit.

Fitting a local regression model for WTPC includes 4 steps:

1. Selecting the portion $s = \frac{k}{n}$ of closest training data v_i to target point v_0.

2. Assigning a weight $W(v_i, v_0)$ to each point v_i in the neighborhood of v_0. Further points from v_0 get less weights and the closest point to v_0 gets the highest weight. The weights of all data points other than the nearest k points will be zero. There are number of different choices to use for the weighting function $W(v_i, v_0)$. In this study, a tricubic weighting function $W(v_i, v_0) = (1 - (\frac{v_i - v_0}{r_k})^3)^3$, where r_k defines the distance from v_0 to the furthest point amongst k points in the neighborhood of v_0, was used.

3. Finding local linear coefficients $\widehat{\beta}_0(v_0)$ and $\widehat{\beta}_1(v_0)$ that minimize the localized least square criterion

   $\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1(v_0 - v_i))^2 W(v_i, v_0)$.

4. The fitted value at v_0 is given by $\widehat{f}(v_0) = \widehat{\beta}_0(v_0)$. It is worth noting that the local coefficient $\widehat{\beta}_1(v_0)$ estimates the derivative of the regression function at $v = v_0$ (Fan and Gijbels, 1996).

It is worth noting that the hybrid nonparametric/parametric regression estimates have already been examined in Burman and Chaudhuri (2012). Nevertheless, only the global combination of nonparametric and parametric estimates have been proposed.

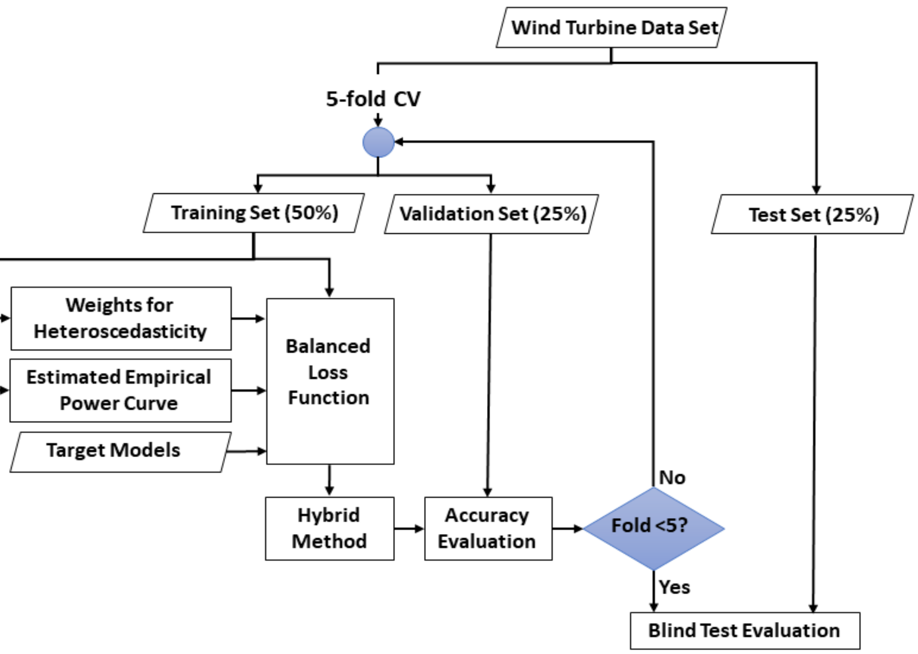


Figure 3.2: A flowchart describing the process of fitting a hybrid power curve using 5-fold cross-validation, and evaluating its performance using test data.

Fig. (3.2) summarizes the whole procedure of fitting a hybrid WTPC model using the proposed approach in this paper, based on a 5-fold Cross-Validation to find the optimum values for hyper-parameters and evaluating the performance of fitted model using a set aside test data.

3.3 Weighting Schemes

Thus far, the focus was on the regression function estimation for the mean generated power as a function of the wind speed based on the observation model in

(3.2), where the information on the noise structure was ignored. In real-life cases, the assumption that the noise variance is constant is unrealistic. For example, consider the power-speed data in Fig. 3.3, which shows the variability in the generated power is much smaller when wind speed is less than 8 m/s compared with the variability when wind speed is larger than 12 m/s. Fig. 3.4 also confirms that noise variance (estimated unbiasedly by squared residual error) is not constant and changes as a function of wind speed. One can observe that the variance of residuals is low when wind speed is low and it increases as wind speed increases. This is known as heterogeneity or heteroscedasticity, whereas constant noise variance is known as homogeneity. There are two main reasons for trying to understand how the variability changes with the predictor in this case. The first reason is that considering non-constant variance in model fitting results in a more accurate estimation. Another reason is that non-constant variance property violates the important homogeneity assumption when using least squares to fit regression models, hence influencing the validity of statistical inferences. In particular, constructed confidence intervals are not reliable. Fig. 3.3 makes it clear, i.e., when trying to draw a confidence interval for the true regression function, one should be less certain about what will happen for large values of the wind speed than for small values (Ruppert et al., 2003).

In order to include the non-constant variance in the regression model, the formula in (3.2) can be re-formulated as follows

$$Y_i = f(v_i) + \sqrt{g(v_i)}\epsilon_i, \tag{3.13}$$

where ϵ_i is a zero mean, unit variance noise and the non-negative function $g(v)$ represents the lack of homogeneity of the observed data. In fact, one can define $Var(Y|v) = g(v)$. The model in (3.13) is referred to as a heteroscedastic regression model.

In order to incorporate the data heteroscedasticity represented by the function $g(v)$, this paper proposes the shrinkage procedure that makes the weighted estimate in (3.2) closer to the target power curve model which is more robust in the region with the high variance. In this paper, the MPC and wind farm average power curves (WFAPC) are considered as target models. Two different weighting schemes which will be used in the proposed hybrid methods for WTPC modeling are introduced in the following sections. These methods are based on estimates of $g(v)$ that are then used to design suitable weighting schemes.

3.3.1 Weighting Scheme 1

Fig. 3.3 shows historical data from one turbine in a wind farm. Also, the MPC is shown in dashed line, and the natural spline regression estimate under the squared error loss function fitted to the data is shown in solid line. One can easily see that the natural spline estimate provides a better fit in the range that wind speed values are low. However, for high wind speed values, the variance of the generated power is high, which will inversely affect the precision of the empirical WTPC fitted to the observed data. This motivates to define a weighting scheme so that it assigns higher

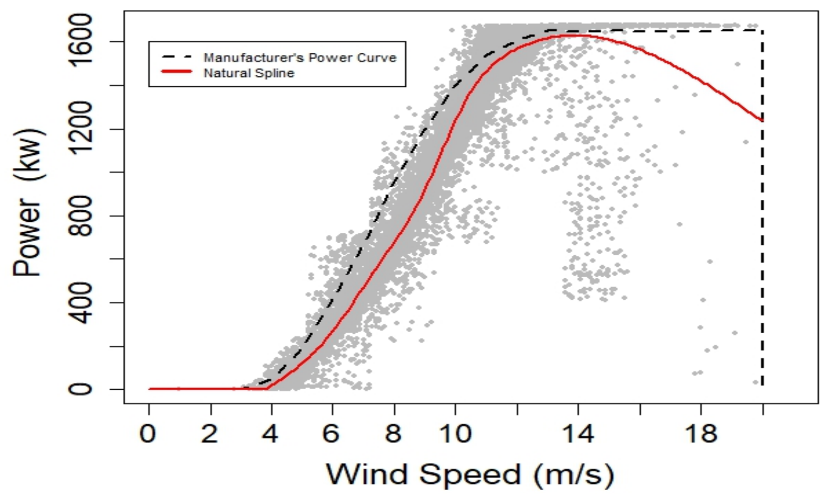


Figure 3.3: Natural spline power curve regression model compared with the MPC shows that the difference between these two models increases as the wind speed increases.

weights to the target power curve when wind speed is high and lower weights to the natural spline estimator. The opposite strategy is applied when wind speed is low.

In order to introduce a data-driven weighting scheme, the natural spline method was utilized to estimate the conditional variance function $g(v)$. The proposed method can also be extended to other methods discussed in Section II. To this end, the natural spline regression estimator was fitted to the training data, and the squared error between the observed values $\{Y_i\}$ of generated power and the predicted values $\{\widehat{P}_i\}$ was calculated (Uyanto, 2019), i.e.,

$$e_i^2 = (Y_i - \widehat{P}_i)^2, \quad i = 1, \ldots, n. \tag{3.14}$$

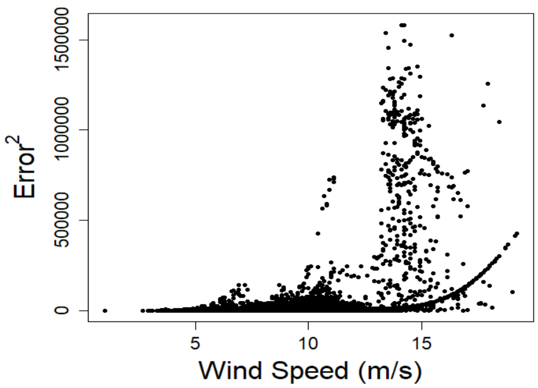


Figure 3.4: Squared errors of the fitted power curve show an exponential relationship between the variance of the squared error and wind speed.

In other words, e_i^2 estimates the variance of the noise associated with the power generated at a given wind speed v_i. Fig. 3.4 shows the values of $\{e_i^2\}$, the estimated residuals variances of the fitted natural spline regression model with respect to wind speeds for one of the turbines in the data set. It is clear that e_i^2 is changing exponentially with v_i. Fig. 3.4 also represents high variance of generated power when wind speed is high. Owing to (3.13), one can write

$$\mathbb{E}[e_i^2] = \mathbb{E}[\widehat{P}_i - P_i]^2 + g(v_i), \tag{3.15}$$

where $P_i = f(v_i)$. It should be noted that the prediction error $\mathbb{E}[\widehat{P}_i - P_i]^2$ is at least of order $O(n^{-4/5})$ for most parametric/nonparametric estimation methods (Li and Racine, 2007). Hence, one can approximately write

$$e_i^2 = g(v_i) + u_i, \tag{3.16}$$

for u_i being a zero-mean noise process, i.e., $\mathbb{E}[e_i^2] = g(v_i)$. By virtue of some prior information, see Fig. 3.4, one can make use of the following parametric model for $g(v_i)$:

$$g(v_i) = \gamma_0 v_i^{\gamma}, \tag{3.17}$$

for some positive constants γ_0, γ.

To recover the coefficients γ_0 and γ, one can write

$$e_i^2 = \gamma_0 v_i^{\gamma} \left[1 + \frac{u_i}{\mathbb{E}[e_i^2]}\right].$$

Then

$$\log(e_i^2) = \log(\gamma_0) + \gamma \log(v_i) + \log\left[1 + \frac{u_i}{\mathbb{E}[e_i^2]}\right].$$

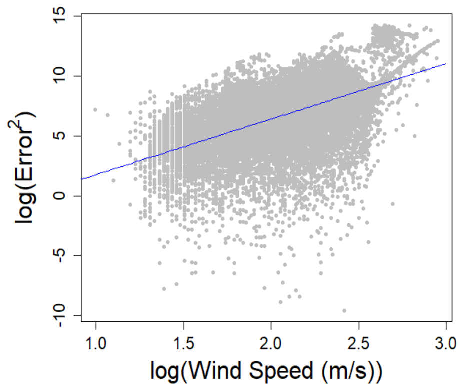
Since the assumption that u_i is small and the fact that $\log(1+x) \approx x$ for small x implies that one can obtain the following linearized regression model

$$e_i' = \log(\gamma_0) + \gamma \log(v_i) + \eta_i, \tag{3.18}$$

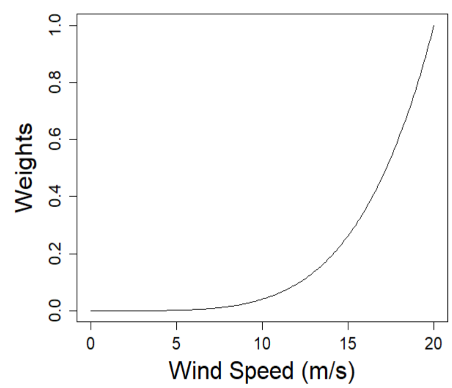
where $e_i' = \log(e_i^2)$ and $\eta_i \approx \frac{u_i}{\mathbb{E}[e_i^2]}$ is the zero-mean noise process (Wasserman, 2006; Wand and Jones, 1994b).

Thus, the estimates $\widehat{\gamma}_0$ and $\widehat{\gamma}$ of γ_0 and γ, respectively, are obtained by minimizing the sum of squares criterion

$$(\widehat{\gamma}_0, \widehat{\gamma}) = \arg \min_{(\gamma_0, \gamma)} \sum_{i=1}^{n} (e_i' - \log(\gamma_0) - \gamma \log(v_i))^2. \tag{3.19}$$



(a)



(b)

Figure 3.5: a) Log transformation of squared error and the fitted linear regression model. b) Weights $\alpha(v_j)$ for each wind speed value v_j.

Fig. 3.5a shows the log transformation and the fitted linear regression model. Using $\widehat{\gamma}_0$ and $\widehat{\gamma}$ in (3.17) one can define

$$w_j = \widehat{\gamma}_0 v_j^{\widehat{\gamma}}, \tag{3.20}$$

where w_j is the weight of the hybrid method for each value of v_j. As w_j may be larger than 1, it is required to scale them to be in the range $[0, m]$ by defining the corrected weights $\alpha(v_j)$

$$\alpha(v_j) = m \times \frac{w_j}{\max(w_j)}, \tag{3.21}$$

where $0 < m < 1$ is a constant that controls the maximum value of $\alpha(v_j)$. Fig. 3.5b shows $\alpha(v_j)$ for different v_j. $m = 0$ leads to $\alpha(v_j) = 0$ for all j, which makes the hybrid method to be the same as the original estimation method. In other words, the contribution of the target model to the aggregated estimator will be

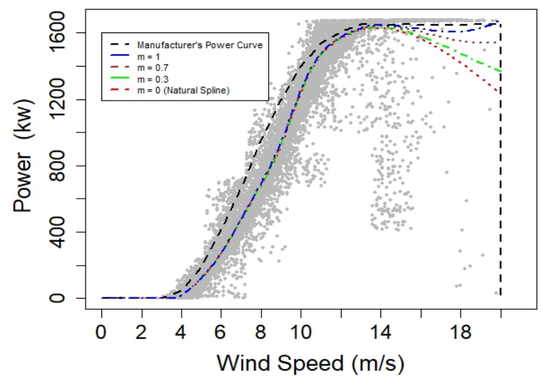


Figure 3.6: Hybrid power curves based on natural spline and MPC for different values of m using weighting scheme 1.

null. Fig. 3.6 shows different power curves for different values of m. It shows that by increasing m, the weights $\alpha(v_j)$ increase which leads to having hybrid curves that are more close to target values in those regions with high weights. To find the proper value of m, different values for m ranging from 0 to 1 by an increment of 0.1 were evaluated.

3.3.2 Weighting Scheme 2

In the second approach, the weight associated with each wind speed range was defined using the standard deviation of the generated power in that speed range. To this end, the whole speed range is divided into R bins. In each bin, the standard

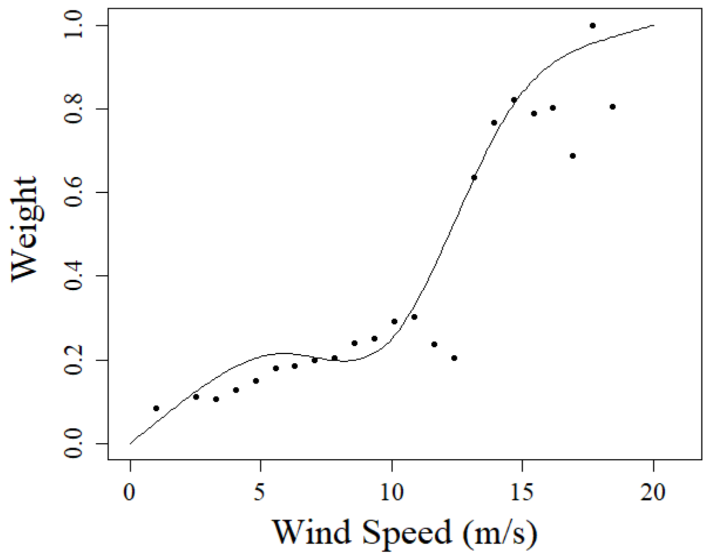


Figure 3.7: Standard deviation of generated power in R bins for a sample turbine and weighting scheme for a sample turbine. Standard deviations are shown with the points alongside with the fitted curve showing the weights.

deviation of the generated power is calculated as follows

$$SD_r = \sqrt{\frac{\Sigma_{v_i \in bin_r}(Y_i - \bar{Y}_r)^2}{|n_r|}}, \quad r \in 1, \ldots, R, \tag{3.22}$$

where n_r is the number of data in bin r, SD_r is the standard deviation of generated power in bin r, and $\bar{Y}_r$ is mean of the generated powers corresponding to wind speed values in bin r. Fig. 3.7 shows values of SD_r and the mid-point of each bin for the one of the examined wind turbines. A natural spline model was fitted on pairs (v_r, SD_r) where v_r is the middle value of each wind speed bin. In this approach, scaling the weights is required to be in the range $[0, m]$ so that $m < 1$. Fig. 3.7

shows weights $\{\alpha(v_r)\}$ by a fitted line on SD_r points, as an example of the weighting scheme for one of the sample turbines.

3.4 Target Models

In most WTPC modeling methodologies, the focus of study is on fitting a curve that provides the goodness of fit with respect to the classical least squares criterion defined in (3.3). However, for non-homogeneous data, like wind data case in this study, emphasizing the goodness of fit may not be sufficient, especially in those regions with high variance in the data. In this situation, one may want to rely on some robust estimates of power for a given wind speed. Using a balanced loss function (3.4), one can define a general framework to make locally weighted hybrid methods to focus more on the goodness of fit in areas with low variance as well as closeness to robust target models in regions with high variance. The ultimate goal is to improve the robustness and accuracy of the fitted curve. In this paper, two different target models based on MPC and WFAPC are utilized.

3.4.1 Manufacturer Power Curve

The International Standard IEC 61400-12-1 is prepared by the International Electrotechnical Commission (IEC) technical committee as a standard methodology for measuring the power performance characteristics of a single wind turbine.

According to this standard, simultaneous measurements of wind speed and power output are recorded at a test site for a sufficiently long duration to provide a significant database. The IEC-based power curve is determined by applying the "method of bins" (Commission et al., 2007).

Wind turbine manufacturers provide a table of pairs data set, which shows the amount of electrical power output according to the wind speed value based on the IEC procedure. For example, in this research, a data set containing 20 pairs of (v_i, p_i) is available, which shows the expected value of output power p_i when the wind speed is equal to v_i. One can calculate the power curve for the whole range of wind speed between cut-in and cut-out points, by interpolation using a polynomial regression on the data set of 20 pairs.

Although the generated power is usually different from the MPC value, still the MPC may contain worthy information about the general performance of the wind turbine. For example, one may consider it a reliable resource when empirical estimation is not robust, e.g., in regions with high variance data.

3.4.2 Wind Farm Average Power Curve

Another target model that can be utilized in the proposed hybrid modeling of WTPC is the WFAPC. Each wind farm contains a different number of wind turbines. Although the performance of the same model of wind turbines is similar to each other, it also depends on different properties such as the location, turbines' age, and maintenance history of the turbine.

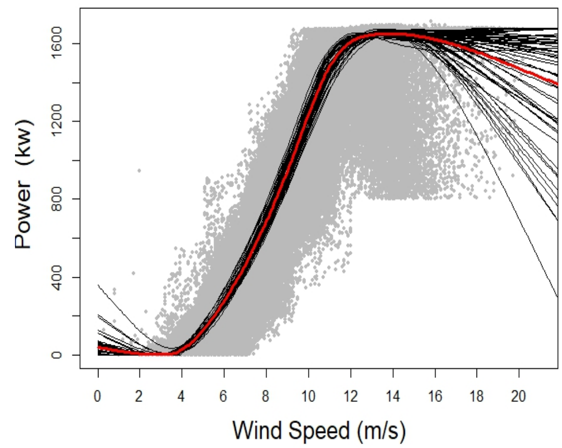


Figure 3.8: WTPC of each 63 wind turbines in a farm located in Canada, alongside with theWFAPC, highlighted one, which calculated by taking average of all 63 power curves.

Generally, one can estimate the average power curve of wind turbines in a wind farm by taking the average of estimated power curves of wind turbines in the farm. To this end, the power curve of each turbine is estimated one by one, using one of the commonly used power curve estimators, following standard data polishing techniques such as three standard deviation around the mean (Shokrzadeh et al., 2014). Then, by taking the average of estimated power curves one can get the WFAPC to be used as a target model for the hybrid method. For example, in this research, the WTPC for each 63 wind turbine in the wind farm is estimated using the natural spline method. Fig. 3.8 shows 63 power curves related to wind turbines in the farm alongside with the highlighted one, which is the WFAPC. The WFAPC contains virtuous information about the general performance of turbines in the farm so that it can be considered as a trustworthy resource.

3.5 Results

In this section, the proposed hybrid methods on proprietary wind power data of a wind farm are evaluated. The wind power plant (WPP) includes 63 identical wind turbines with the rated capacity of 1.7 MW, and the hub height of 80 m spread over an area of over 90 km^2. The cut-in speed, rated speed, and cut-out speed of the turbines are 3.5, 13, and 20 m/s, respectively. Four wind turbines (T1,$\cdots$,T4) of the WPP are selected to analyze the performance of the proposed methods. Raw data contains 30398 samples of wind speed and generated power with a resolution equal to 10-minutes from June 2006 to January 2007. To polish the data, the outlier detection process described in Shokrzadeh et al. (2014) is used. To do so, after removing data with negative values of wind speed or wind power, data points with power values that are at least three standard deviation away from the average power value at that speed were removed.

Statistical metrics that can be used for evaluating the performance of the fitted power consist the root-mean-squared error (RMSE), normalized mean absolute percentage error (NMAPE), symmetric mean absolute percentage error (sMAPE), the mean absolute error (MAE), and the coefficient of determination (R2) (Kusiak et al., 2012; Marvuglia and Messineo, 2012). In this paper, MAE index was evaluated by the following formula

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{P}_i|, \tag{3.23}$$

where Y_i is the observed wind power and $\widehat{P}_i$ is the predicted value of the power.

To compare the MAE value of different methods, data set was divided into 3 different subsets: training set (50%), validation set (25%) and test set (25%). Training data and validation set were utilized to find the best model. Then, the performance of fitted model was evaluated on the test data. The process was repeated 5 times and the average and standard deviation of MAE for each model are reported. This process is represented for comparison purposes.

3.5.1 Experiment 1

In this section, the performance of polynomial regression, natural spline regression, and local regression as commonly used power curve estimation methods are investigated when MPC and WFAPC are used as target models.

Table 3.1 shows the mean of MAE values of each method on 4 different turbines. Table 3.1 also shows rankings of methods, where (1) is the best and (5) is the worst rank. The last column of Table 3.1 shows the overall performance ranking of each method. As expected, it shows that the performance of target models (MPC and WFAPC) individually is worse than the other three empirical estimation methods (PR, NS, and LR). It is also clear that natural spline outperforms other methods.

In the following sections, the effect of hybrid approaches using natural spline and local regression are investigated when MPC and WFAPC are used as target models and when 2 different weighting schemes are utilized.

Method	Turbine T1	Turbine T2	Turbine T3	Turbine T4	Rank
	MAE	MAE	MAE	MAE	
PR	74.71 (3) [±1.3]	65.94 (3) [±1.22]	67.77 (3) [±0.99]	69.64 (3) [±1.51]	(3)
NS	73.38 (1) [±1.5]	63.51 (1) [±1.27]	66.57 (2) [±0.87]	68.56 (2) [±1.73]	(1)
LR	73.95 (2) [±1.41]	64.03 (2) [±1.17]	66.44 (1) [±0.92]	68.03 (1) [±1.56]	(2)
MPC	217.46 (5) [±1.36]	193.52 (5) [±1.79]	150.85(5) [±1.81]	198.46 (5) [±1.53]	(5)
WFAPC	88.84 (4) [±1.41]	67.49 (4) [±1.27]	73.29 (4) [±1.39]	77.5 (4) [±1.4]	(4)

Table 3.1: Mean and standard deviation (in [ ]) of MAE for 10-MIN Data from Turbines T1, T2, T3 and T4 using the Methods Including: Polynomial Regression (PR), Natural Spline (NS), Local Regression (LR), Manufacturer's Power Curve (MPC) and Wind Farm Average Power Curve (WFAPC)

3.5.2 Experiment 2

In this experiment, hybrid models using natural spline as an empirical estimator were evaluated. Four different hybrid methods using two different target models (MPC and WFAPC) based on two various weighting schemes were implemented. In this experiment, the same data and turbines were used as in Experiment 1. Table 3.2 shows the results of this experiment. It also shows that the best hybrid method is the combination of natural spline and WFAPC using the weighting scheme 2.

Method	Turbine T1	Turbine T2	Turbine T3	Turbine T4	Rank
	MAE	MAE	MAE	MAE	
NS	73.38 (5) [±1.5]	63.51 (5) [±1.27]	66.57 (5) [±0.87]	68.56 (5) [±1.73]	(5)
NS-MPC-1	71.71 (3) [±1.43]	62.16 (2) [±1.26]	65.21 (3) [±1.1]	66.56 (3) [±1.57]	(3)
NS-MPC-2	69.73 (2) [±1.31]	62.96 (4) [±1.31]	63.99 (2) [±1.7]	64.34 (2) [±2.75]	(2)
NS-WFAPC-1	71.95 (4) [±1.44]	62.32 (3) [±1.25]	65.49 (4) [±1.12]	66.82 (4) [±1.51]	(4)
NS-WFAPC-2	68.13 (1) [±1.43]	60.51 (1) [±1.51]	62.99 (1) [±1.35]	63.61 (1) [±1.76]	(1)

Table 3.2: Mean and standard deviation (in [ ]) of MAE for 10-MIN Data from Turbines T1, T2, T3 and T4 using the Methods Including: Natural Spline (NS), Natural Spline Combined with MPC based on Weighting Scheme 1 (NS-MPC-1), Natural Spline Combined with MPC based on Weightin Scheme 2 (NS-MPC-2), Natural Spline Combined with WFAPC based on Weighting Scheme 1 (NS-WFAPC-1) and Natural Spline Combined with WFAPC based on Weighting Scheme 2 (NS-WFAPC-2)

However, one can figure out all different combinations of hybrid methods outperform the original method. One can conclude that hybrid method performs better than original estimators since hybrid method gives more weights to target models when the variance of data is high.

3.5.3 Experiment 3

In this section, the previous experiment was repeated except that a local regression power curve modeling was used to form an empirical estimator. To this end, four different hybrid methods were implemented using two different target

Method	Turbine T1	Turbine T2	Turbine T3	Turbine T4	Rank
	MAE	MAE	MAE	MAE	
LR	73.95 (5) [±1.41]	64.03 (5) [±1.17]	66.44 (5) [±0.92]	68.03 (5) [±1.56]	(5)
LR-MPC-1	69.96 (3) [±1.31]	63.5 (4) [±1.4]	66.07 (4) [±1.6]	65.3 (3) [±1.9]	(4)
LR-MPC-2	69.68 (2) [±1.34]	63.41 (3) [±1.36]	63.94 (2) [±1.7]	64.14 (2) [±2.6]	(2)
LR-WFAPC-1	72.53 (4) [±1.34]	62.76 (2) [±1.15]	65.24 (3) [±1.07]	66.33 (4) [±1.46]	(3)
LR-WFAPC-2	68.04 (1) [±1.48]	60.62 (1) [±1.6]	62.96 (1) [±1.38]	63.43 (1) [±1.76]	(1)

Table 3.3: Mean and standard deviation (in [ ]) of MAE for 10-MIN Data from Turbines T1, T2, T3 and T4 using the Methods Including: Local Regression (LR), Local Regression Combined with MPC based on Weighting Scheme 1 (LR-MPC-1), Local Regression Combined with MPC based on Weightin Scheme 2 (LR-MPC-2), Local Regression Combined with WFAPC based on Weighting Scheme 1 (LR-WFAPC-1), Local Regression Combined with WFAPC based on Weighting Scheme 2 (LR-WFAPC-2)

models based on two different weighting schemes discussed in Section 3.3. Table 3.3 shows the results of this experiment. It also confirms that the hybrid method using WFAPC as the target model outperforms the original methods.

To sum up all the results, the four best methods are presented in Table 3.4. It shows that the hybrid method using local regression and WFAPC using the second weighting scheme based on variance performs the best on almost all turbines studied in this paper. Also, it shows using the second weighting scheme results in hybrid methods that improve the performance of fitted power curve. One can conclude that using WFAPC is a better option to be used as target model according to the results. It worth mentioning that there is no unique best method for all different

Method	Turbine T1	Turbine T2	Turbine T3	Turbine T4	Rank
	MAE	MAE	MAE	MAE	
NS-WFAPC-2	68.13 (2) [±1.43]	60.51 (1) [±1.51]	62.99 (2) [±1.35]	63.61 (2) [±1.76]	(2)
NS-MPC-2	69.73 (4) [±1.31]	62.96 (3) [±1.31]	63.99 (4) [±1.7]	64.34 (4) [±2.75]	(4)
LR-MPC-2	69.68 (3) [±1.34]	63.41 (4) [±1.36]	63.94 (3) [±1.7]	64.14 (3) [±2.6]	(3)
LR-WFAPC-2	68.04 (1) [±1.48]	60.62 (2) [±1.6]	62.96 (1) [±1.38]	63.43 (1) [±1.76]	(1)

Table 3.4: Mean and standard deviation (in [ ]) of MAE for 10-MIN Data from Turbines T1, T2, T3 and T4 for the Best Methods in the Previous Experiments

data sets, but one can conclude that hybrid methods perform better than classic methods in power curve modeling.

3.6 Concluding remarks

WTPC modeling is not only an important part of wind power forecasting but also wind energy power system management. Heteroscedasticity of data is one of the important issues to deal with in power curve modeling which impacts the quality and robustness of the fitted power curves. In this paper, two different locally defined weighting schemes were proposed to make hybrid nonparametric/parametric methods for WTPC modeling and increase the goodness of fitted curve as well as its closeness to some target models. A balanced loss function was defined to achieve this goal, and a WTPC model was constructed for its minimization. The proposed approach is very general and can be used for any target power curve model.

However, results were presented when WFAPC and MPC were used as target power curves. Experimental results in this paper confirm that although these target models may not perform accurate enough to be utilized individually, they contain worthy properties such as robustness which causes hybrid methods to perform better in terms of accuracy and robustness than classical power curve estimators. Results show that the local regression method combined with the WFAPC using a local weighting scheme based on the variance of real data in different bins of the wind speed value, outperforms other methods. Results also show that hybrid methods improved the MAE values alongside keeping the standard deviation of MAE at the same level. It confirms that using hybrid methods, the bias will be reduced by keeping the variance at the same level. The outcome of this study can be used in wind farms data sets which suffer from heteroscedasticity issue in different applications including turbine performance monitoring, power forecasting, fault detection, and estimating the required storage capacity for integrating to power system.

Chapter 4

Multilevel Modeling

This chapter encloses a section of the author's PhD work that is accepted to published at *IEEE Transaction on Sustainable Energy*[1]. In previous chapters, we discussed how one can improve the accuracy of fitted power curve for each wind turbine. In this chapter, we focus on improving the wind farm aggregated power curves instead of just one wind turbine power curve.

[1]© 2021 IEEE. Reprinted, with permission, from
M. Mehrjoo, M. Jafari Jozani, M. Pawlak and B. Bagen, "A Multilevel Modeling Approach towards Wind Farm Aggregated Power Curve," in IEEE Transactions on Sustainable Energy, https://doi.org/10.1109/TSTE.2021.3087018

Abstract: Wind farm multiple aggregated power curve modeling plays an important role in reducing the complexity of analyses in wind farm management and annual power prediction. There is a trade-off between the complexity and accuracy of aggregated power curves. In this paper, K-Means clustering is utilized to classify turbines in a wind farm into homogeneous groups according to a new set of features based on the overall performance of turbines. We apply multilevel modeling methods, including random intercept and random slope models on turbine clusters, to take into account the hidden correlation among different clusters. Results show that the accuracy of our proposed methods are higher than the single aggregated method alongside an equal complexity. The proposed multiple aggregated power curve model can be utilized to analyze wind farm behavior and wind farm power simulations to forecast wind power.

4.1 Introduction

The shortage of fossil fuels and their associated problems such as pollution have forced countries to replace them with clean and renewable energies such as wind Lindenberg et al. (2008). Rapid growth in wind farms' size has helped wind energy to cover a massive portion of electricity production sources. However, the stochastic nature of wind and its volatility have raised important challenges pertinent to wind farms and power system stability. Therefore, it is necessary to analyze wind farms' behavior by accurately modeling the wind farm power function and accordingly monitoring power system stability Mehrjoo et al. (2020). Modeling power curves of tens or hundreds of individual wind turbines make the mathematical model of the power system very complex and increase the overall time and effort required to simulate and evaluate the performance of the underlying power system Ghaedi et al. (2013); Hur (2018). Any reasonable reduction in the mathematical model of each component of the power system may reduce the computational cost and simulation time. One solution is to present the wind turbine power curve models of the wind farm by a single or multiple aggregated models while maintaining a sufficient level of accuracy Ni et al. (2016); Liu et al. (2019).

A single aggregated model may be suitable for small wind farms as it can retain the accuracy requirements, but it often suffers from significant errors when there exist a large number of wind turbines of different types or control parameters Liu et al. (2019); Kim et al. (2012); Perdana et al. (2008). For such situations, multiple aggregated wind farm power curves are suggested at the cost of an increase in the complexity of the fitted models Zha et al. (2019); Ghosh and Senroy (2013).

In Fernandez et al. (2006), an aggregation model by clustering based on the similarity of wind speed has been proposed. In this approach one needs to choose an appropriate number of clusters and divide wind turbines into different clusters (i.e., homogeneous groups) based on the similarity of wind speed recorded at each turbine's location. However, in most wind farms (including the data set in this paper), the number of meteorological masts in the farm is much less than the number of wind turbines, which makes wind speed not accurate enough to be used as a suitable feature in the clustering process. Later on, in González-Longatt et al.

(2012); Dou et al. (2019); Gao et al. (2019); Ali et al. (2012); Zhang and Liu (2019), wake effect and wind farm layout were added to the clustering process. However, taking the wake effect into account for clustering turbines is challenging and raises complex issues.

As the wake effect changes by wind direction and impacts turbines' performance and their generated power, one should always consider wind direction as an important feature. To take into account the impact of the wake effect and avoid different clustering groups for different wind direction values, and to reduce the complexity of the wind farm aggregated model, this paper develops a clustering approach based on a set of features derived from wind turbines' generating power performance in a long term duration. Since the wake effect impacts wind turbines' performance, clustering based on wind turbines' performance inherently considers the wake effect. So, clustering based on statistical metrics of historical data, especially generated power performance, would be the proper option to partially resolve the issue of wind direction impact on wake effect. In addition, the proposed feature selection not only considers the wake effect from other turbines indirectly but also the effects pertinent to turbines' age, capacity, and model, among many others.

The mean of generated power in different wind speed bins was used as a reasonable feature for clustering turbines into several groups with similar performance, and this can be used even in the absence of turbines' exact location. K-means clustering based on the Euclidean distance was used to assign turbines in a wind farm to different groups such that in each group generating powers are homogeneous, and between groups, they are non-homogeneous.

An appropriate clustering approach allows multiple aggregated power curves to reduce the order of fitted power curves from the number of turbines in a wind farm to the number of clusters and accordingly reduce the model complexity, the number of parameters, and simulation time. In this paper, multilevel models are also developed to take into account the correlation among fitted power curves for each cluster and improve the accuracy of the wind farm aggregated model Grajeda et al. (2016); Chen et al. (2020). Random intercept and random slope B-spline aggregated models are developed as competitors to a single aggregated model.

This study aims to:

- cluster a relatively large number of turbines in a wind farm into a small number of homogeneous groups by introducing a novel set of features based on the performance of turbines in generating power,
- proposing a solution for handling the important trade-off between complexity and accuracy in developing wind farm multiple aggregated power curve models,
- incorporating the hidden correlation that exists between turbines in different clusters for fitting wind farm multiple aggregated power curves using advance statistical tools based on mixed effect models,
- utilizing statistical tests to confirm the effectiveness of our proposed wind farm multiple aggregated power curve.

The outline of the paper is as follows. In Section 4.2, the proposed clustering method and the statistical metrics used for choosing the correct number of clusters are discussed. Section 4.3 introduces multilevel modeling and shows how this can be used to reduce the order of wind farm aggregated power curve. In Section 4.4, the proposed methodology is implemented on an actual data set retained from a wind farm in Canada. Also, using different criteria, the performance of the aggregation method is evaluated. Section 4.5 is devoted to concluding remarks and future work.

4.2 Clustering Methods

Clustering methods are utilized to assign wind turbines into homogeneous groups. Features that affect the turbines' performance can be classified into those inherent to the turbine itself (e.g., turbine's model, height, etc.) or to environmental effects (e.g., wake effect, the direction of the turbine, terrain condition). In practice, it is desirable to develop an algorithm that accurately simulates turbine's performance. Our proposed approach below uses a set of features based on the average turbine's generated power in different wind speed intervals (bins) that represents those two classes in order to perform the necessary clustering before developing the proposed multilevel models to construct a wind farm aggregated model.

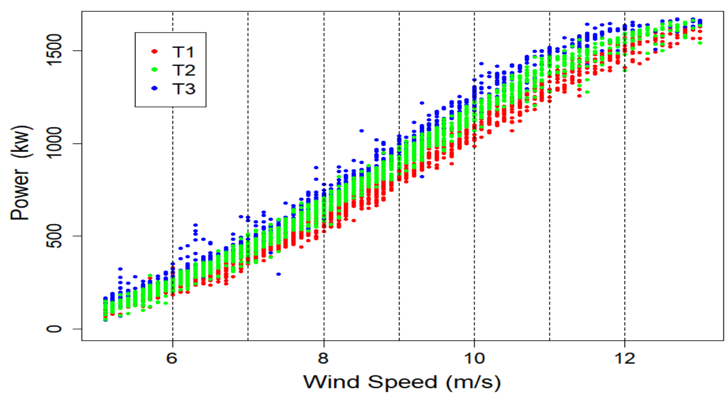
4.2.1 Clustering Features

Different clustering features with different separability powers were investigated in the literature, such as wind speed, wind direction, and turbines' location. In

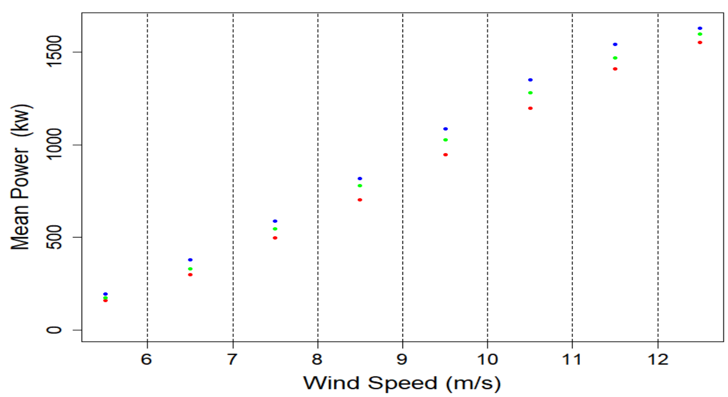
Liu et al. (2014), the mean and standard deviation of generated power for each turbine were considered as clustering features. However, they do not contain enough information about the performance of turbines and ignore the wake effects in the clustering procedure. This paper uses the whole turbines' semi-empirical power curve model as the underlying feature for clustering. To evaluate the semi-empirical power curve for each turbine, the wind speed range between the cut-in and rated points is divided into M regions $[R_m, R_{m+1}], m = 1, \ldots, M$ and the mean of generated power in each region is calculated, where $\{R_i\}$ denotes the speed knot point of the i-th partition. For turbine t, $t = 1, \ldots, T$, this results in a set of M features describing the mean of generated power in that wind speed range, given by

$$\overline{p}_{t,m} = \frac{\sum_k p_t(v_k)\mathbf{1}(R_m \leq v_k < R_{m+1})}{\sum_k \mathbf{1}(R_m \leq v_k < R_{m+1})}, \quad m = 1, \ldots, M, \tag{4.1}$$

where $\mathbf{1}(A)$ is the indicator function of the set A. Also $p_t(v_k)$ is the generated power by the t-th turbine at the wind speed v_k. The equation in (4.1) defines the M-dimensional feature vector $\overline{\boldsymbol{p}}_t = (\overline{p}_{t,1}, \ldots, \overline{p}_{t,M})^\top$ that characterizes the generated power of the t-th turbine over the assumed range of wind speed. Fig. 4.1 shows clustering features for three sample turbines. Fig. 4.1(a) represents generated power by three different turbines in three different colors. Fig 4.1(b) represents the mean of generated powers in each bin for these turbines, which are considered as clustering features. So, for each turbine, the empirical turbine's power curve in the range between the cut-in and rated speeds are simulated by a set of M numbers equal to means of generated powers. To avoid any bias in wind power curve modeling,



(a)



(b)

Figure 4.1: a) Speed-Power pairs of data from 3 different turbines in the wind farm represented by 3 different colors b) Average of generated power in different bins (M=8) of speed values for each turbine.

a portion of data (one month) was used for clustering and the rest was used for training and validity of fitted wind power curve models.

4.2.2 K-Means Clustering Method

This is a commonly used clustering method that clusters data into K groups based on some distance measures. In this paper, the distance between two M-dimensional feature vectors $\overline{\boldsymbol{p}}_i$ and $\overline{\boldsymbol{p}}_j$ is determined by the classical Euclidean distance

$$d(\overline{\boldsymbol{p}}_i, \overline{\boldsymbol{p}}_j) = \left(\sum_{m=1}^{M} |\overline{p}_{i,m} - \overline{p}_{j,m}|^2 \right)^{1/2}. \tag{4.2}$$

Prior to performing the clustering method, it is necessary to specify the desired number of clusters K. Upon deciding the number of clusters, K-means clustering approach will assign each data to an appropriate cluster. Section 4.2.3 describes how to find K in practice. The K-means clustering algorithm is as follow:

1) Choose an initial set of K centers and assign each turbine feature vector to a cluster with the nearest center.

2) Repeat the following steps continuously until there is no change in the cluster assignments:

- For each cluster, compute the cluster centers by taking the average of each turbine's feature set assigned to this cluster.
- Assign turbines to the clusters with closest center.

The K-means criterion is based on minimizing the pairwise distance of data points within the same cluster. Formally, the K-means criterion with respect to the partition $\{c_1, \ldots, c_K\}$, such that $\bigcup_{r=1}^{K} c_r = \{\overline{\boldsymbol{p}}_1, \ldots, \overline{\boldsymbol{p}}_T\}$, is defined as follows

$$\sum_{r=1}^{K} \frac{1}{|c_r|} \sum_{i,j \in c_r} d^2(\overline{\boldsymbol{p}}_i, \overline{\boldsymbol{p}}_j), \tag{4.3}$$

where $|c_r|$ is the size of the cluster c_r and the notation $i, j \in c_r$ denotes that the feature vectors $\overline{\boldsymbol{p}}_i, \overline{\boldsymbol{p}}_j$ fall into the cluster c_r. The optimal partition is the one that minimizes (4.3); however the algorithm may return its local minimum. To avoid that, it is required to repeat the K-means algorithm multiple times with different random initial cluster assignments.

4.2.3 Elbow Method

Selecting the correct number of K clusters plays an essential role in clustering methods as well as in validating the results of the wind farm aggregate model. One of the most commonly used methods to determine the optimal number of clusters is the elbow method that evaluates the within-cluster dissimilarity W_k as a function of the number of clusters K (Friedman et al., 2001).

Let $\{\overline{\boldsymbol{p}}_1, \ldots, \overline{\boldsymbol{p}}_T\}$ be data set representing the $M-$dimensional feature vectors corresponding to T turbines and suppose the data is partition into K clusters $\{c_1, \ldots, c_K\}$.

The total distances D_r for all turbines in the group c_r is defined as

$$D_r = \sum_{i,j \in c_r} d(\overline{\boldsymbol{p}}_i, \overline{\boldsymbol{p}}_j). \tag{4.4}$$

There are $\alpha_r = (|c_r| - 1)|c_r|/2$ non-zero different distances $d(\overline{\boldsymbol{p}}_i, \overline{\boldsymbol{p}}_j)$ for the group c_r, where $|c_r|$ is the size of the group c_r. Let W_k be the mean within-cluster sum of distances :

$$W_K = \sum_{r=1}^{K} \frac{1}{\alpha_r} D_r. \tag{4.5}$$

Different values will be obtained for $K \in \{1, 2, \cdots, K_{max}\}$. Generally, W_K decreases by increasing K. A simple mathematical derivation shows that W_K is monotonically decreasing with K. Hence, W_K is not informative in choosing the optimal number of clusters by itself.

Considering that the actual distinct number of turbine clusters is K^*, the intuition underlying the approach is that for $K < K^*$, there exists a subset of the true underlying groups in each or some of the K groups returned by the algorithm. Consequently, the solution criterion value tends to decrease substantially with each successive increase in the number of clusters and $W_{K+1} < W_K$. In other words for $K < K^*$, it is expected to have a fast reduction in consecutive differences in the criterion values $W_K - W_{K+1}$ and $\{W_K - W_{K+1} | K < K^*\} >> \{W_K - W_{K+1} | K > K^*\}$. An estimate of K^* is obtained by specifying the point that there will be a sharp elbow in the graph of W_K versus the number of clusters (Friedman et al., 2001). A more formal method is proposed in Tibshirani et al. (2001) using the GAP statistic.

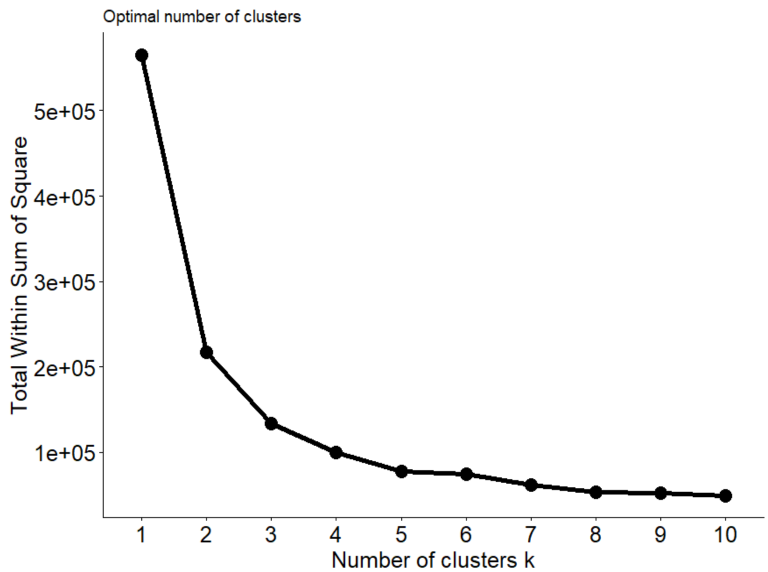


Figure 4.2: Elbow method for finding the optimal number of clusters for 63 turbines in a wind farm in Canada. $K = 3$ is selected as the optimal number of clusters according to the sharp decreases prior to $K = 3$ and smooth decrease after $K = 3$.

Table 4.1: Number of turbines in each cluster

Cluster	C1	C2	C3
#Turbines	18	18	27

Fig. 4.2 shows the elbow metric for $T = 63$ turbines in a wind farm in Canada based on clustering feature set explained in Section 4.2.1. It is clear that the optimal number of clusters based on the elbow method is 3 since the sharp decreasing condition is not satisfied after $K = 3$ number of clusters. Table 4.1 shows the number of turbines in each cluster.

4.3 Multilevel Modeling

Multilevel models are commonly used for problems with multilevel, hierarchical, or longitudinal data structures. They are also called mixed models since they contain fixed effects and random effects. One should expect that utilization of information pertinent to turbines within and between clusters would be beneficial in developing a more accurate wind farm aggregated power curve. Multilevel models enable us to study the effect of wind speed on generated power in each level and account for common characteristics of observations on turbines such as irregularly spaced observations, time-varying variability and correlated errors within clusters.

Three different approaches are developed for wind farm aggregated model fitting in this study, including 1- Single level model, which fits just one power curve on the whole data from all turbines in the wind farm, 2- Random intercept model, which is the simplest multilevel model, 3- Random slope model.

4.3.1 Single level B-spline wind farm aggregated model

The wind farm's power curve shows the electrical power output ratings of the turbines in a wind farm for different wind speeds (Gasch and Twele, 2011). Each turbine's manufacturer provides a table or graph showing the relationship between generated power and wind speed under some specific conditions (Commission et al., 2007). Manufacturers' power curve has three main characteristic speeds: 1) cut-in (Vc); 2) rated (Vr), and 3) cut-out (Vs) speeds. For wind speed less than Vc, a wind turbine can not generate power. The turbine generates power increasingly by incrementing wind speed in (Vc, Vr) until wind speed reaches the rated power. To avoid damages, wind turbine stops generating power as wind speed reaches Vs (Manwell et al., 2010).

Various parametric and nonparametric statistical models can be used to obtain a single level aggregated empirical power curve of a wind farm (Mehrjoo et al., 2019; Carrillo et al., 2013; Shokrzadeh et al., 2014). One of such methods is the polynomial regression (Lydia et al., 2013). Under a single level model, the observed power p_{it} at the wind speed v_i for the t-th turbine is written as

$$p_{it} = f(v_i) + \epsilon_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T, \tag{4.6}$$

when $f(v)$ represents the single power curve for all turbines. Let $p_{i.} = \frac{1}{T}\Sigma_{t=1}^{T} p_{it}$, and $\epsilon_{i.} = \frac{1}{T}\Sigma_{t=1}^{T}\epsilon_{it}$ and write

$$p_{i.} = f(v_i) + \epsilon_{i.}, \quad i = 1, \ldots, N. \tag{4.7}$$

Using a polynomial regression model $f(v_i) = \Sigma_{j=0}^{l}\beta_j v_i^j$, the least squares (LS) estimate of the aggregated wind farm power curve is given by

$$\hat{p}_{i.} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}^\top V_i, \tag{4.8}$$

where

$$\hat{\boldsymbol{\beta}} = (V^\top V)^{-1} V^\top \tilde{P} \quad \text{and} \quad \hat{\beta}_0 = p_{..} - \hat{\boldsymbol{\beta}}^\top \overline{V}. \tag{4.9}$$

Here $p_{..} = \frac{1}{N}\Sigma_{i=1}^{N} p_{i.}$ and $\tilde{P} = (p_{1.}, \ldots, p_{N.})^\top$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_l)^\top$, and V is a matrix with its i-th row being defined as $V_i = (v_i, v_i^2, ..., v_i^l)$.
Polynomial regression models are global in nature and are not robust against the presence of outliers. Piecewise polynomial regression models such as B-splines are often used to resolve this issue by fitting smooth functions with continuous derivatives on different wind speed ranges (Shokrzadeh et al., 2014). A B-spline model is represented as

$$\tilde{P} = \boldsymbol{B}\boldsymbol{\beta} + \epsilon, \tag{4.10}$$

where $\boldsymbol{B}$ is a matrix with its (i, j)-th element being

$$\boldsymbol{B}_j^l(v_i) = \frac{(v_i - \zeta_j)}{(\zeta_{j+l} - \zeta_j)} \boldsymbol{B}_j^{l-1}(v_i) + \frac{(\zeta_{j+l+1} - v_i)}{(\zeta_{j+l+1} - \zeta_{j+1})} \boldsymbol{B}_{j+1}^{l-1}(v_i), \tag{4.11}$$

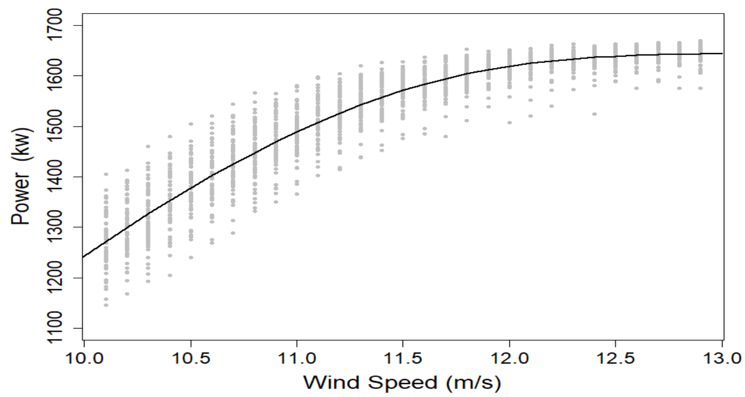
for $j = -L, -L+1, \ldots, L$, $\zeta_0 = \zeta_{-1} = \ldots = \zeta_{-L} = min\{v_i, i = 1, \ldots, N\}$, and $\zeta_{L+1} = max\{v_i, i = 1, \ldots, N\}$. Here $\{\zeta_i, i = 1, \cdots, Q\}$ is a set of knots that controls the flexibility of the fitted curve. Also, $\boldsymbol{B}_j^0(v_i)$ are the natural basis for piecewise constant functions. One can estimate $\boldsymbol{\beta}$ using the LS criterion as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top \tilde{P}. \tag{4.12}$$

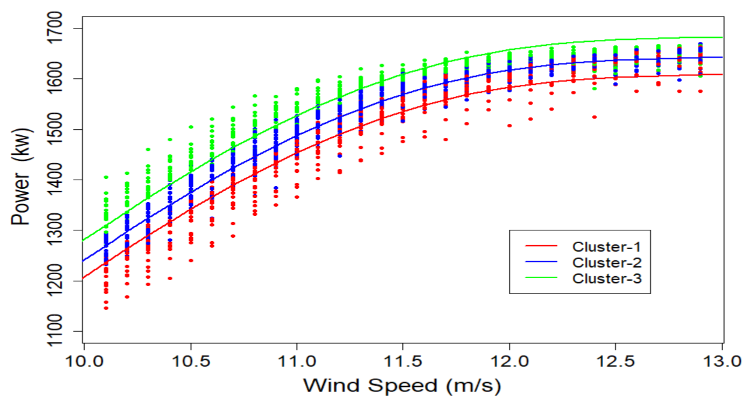
In our experiments, we utilized the B-spline method for fitting the wind farm aggregated power curve in the range of wind speed between the cut-in and rated points. To this end, we aggregated all the data from each turbine in the farm (after separating the portion of data used for clustering), divided the remaining data into training and test data using a 10-fold cross-validation approach. We then fitted one B-spline model on all turbines' data, called single aggregated model. Fig. 4.3(a) shows the B-spline single aggregated model curve fitted to our data set. We compared the results of the single aggregated model with our proposed models using different accuracy and complexity metrics.

4.3.2 Random intercept B-spline wind farm aggregated model

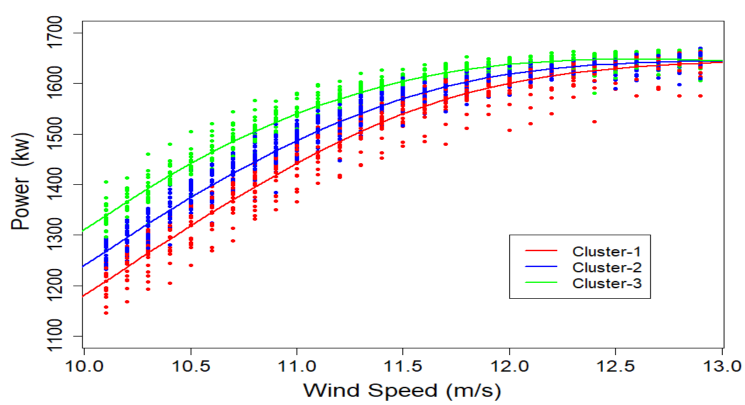
To construct a wind farm aggregated power curve using B-spline models when there are clusters of turbines, multilevel models can be used in order to take into account two sources of variabilities at the turbine-level (level 1) and cluster-level (level 2). From this point of view, a multilevel model can be considered as a variance components model since the total variance can be decomposed into a between-cluster and within-cluster variance components and account for the correlation on within- and between-cluster effects.



(a)



(b)



(c)

Figure 4.3: (a) A single B-spline wind farm aggregated model (b) A random intercept B-spline wind farm aggregated model and (c) a random slope B-spline wind farm aggregated mode fitted to the wind farm data set shown for the wind speeds between 10 and 13 m/s.

Let p_{itk} be the observed generated power at wind speed v_i for turbine t which belongs to cluster k. Note that $T = \Sigma_{k=1}^{K}|c_k|$, where $|c_k|$ is the number of turbines in cluster k. Let

$$p_{ik} = \frac{1}{|c_k|}\Sigma_{t\in c_k} p_{itk}, \tag{4.13}$$

be the average value of the power generated by turbines in cluster k at the wind speed $v_i, i = 1, \ldots, N$ and $k = 1, \ldots, K$.

In a random intercept aggregated model p_{ik}s are models as follows

$$p_{ik} = \beta_0 + U_{0k} + \boldsymbol{\beta}^\top \boldsymbol{B}_i + \epsilon_{ik}, \tag{4.14}$$

for $i = 1, .., N$, where U_{0k} is a random variable and $\boldsymbol{\beta}$ is the l-dimensional vector of parameters. The model in (4.14) is often referred to as the random intercept model as U_{0k} plays the role of a random intercept along with its deterministic counterpart β_0. U_{0k} are assumed to follow Gaussian distribution $\mathbb{N}(0, \sigma_U^2)$ with a noise process ϵ_{ik} distributed as $\mathbb{N}(0, \sigma_\epsilon^2)$. Furthermore, the random variables $\{U_{0k}\}$ and $\{\epsilon_{ik}\}$ are independent. In (4.14) the term $\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{B}_i$ is the same for all clusters, whereas the random intercept U_{0k} varies from cluster to cluster. Thus, the random intercept aggregated model predictor is obtained as

$$\hat{p}_{ik.} = \hat{\beta}_0 + \widehat{U}_{0k} + \hat{\boldsymbol{\beta}}^\top \boldsymbol{B}_i, \tag{4.15}$$

for some estimates $\hat{\beta}_0$, $\hat{\boldsymbol{\beta}}$ of β_0, $\boldsymbol{\beta}$ and the predictor $\widehat{U}_{0k}$ of U_{0k}. Besides finding the estimates $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}$, one also needs to evaluate the uncertainty of the model by estimating σ_U^2 and σ_ϵ^2, the variances of U_{0k} and ϵ_{ik}, respectively.

To simplify the model fitting problem and reduce the number of unknown parameters, the B-spline basis functions are assumed to be the same for all clusters. Fig 4.3(b) shows the plot of an estimated random intercept aggregated power curve consisting of a set of parallel curves, one for each cluster.

4.3.3 Random slope B-spline wind farm aggregated model

In a wind farm aggregated model fitting, one can use a more complex model than a random intercept B-spline aggregated model by using different slopes for models fitted inside each cluster. A random slope B-spline aggregated model is more flexible than the one with random intercept and is defined as

$$p_{ik} = \beta_0 + U_{0k} + \boldsymbol{\beta}^\top \boldsymbol{B}_i + \boldsymbol{U}_k^\top \boldsymbol{B}_i + \epsilon_{ik}, \tag{4.16}$$

for $i = 1, \ldots, N$. Here besides the random intercept U_{0k}, one has the random slope component $\boldsymbol{U}_k$ being the l-dimensional vector distributed as an $l-$variate Gaussian process $\mathbb{N}_l(\mathbf{0}, \boldsymbol{\Sigma}_U)$ with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

The model in (4.16) has the fixed part $\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{B}_i$ that is universal for all clusters and the corresponding random effect part $U_{0k} + \boldsymbol{U}_k^\top \boldsymbol{B}_i$ represented by the random vector $(U_{0k}\ , \boldsymbol{U}_k)$ that specifies the given cluster. Hence, the model can be re-written as

$$p_{ik} = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{B}_i + U_{0k} + \boldsymbol{U}_k^\top \boldsymbol{B}_i + \epsilon_{ik}, \quad i = 1, \ldots, N. \tag{4.17}$$

The equivalent matrix form of (4.17) is as follows

$$\boldsymbol{p}_k = \boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{B}\mathcal{U}_k + \boldsymbol{\epsilon}_k, \tag{4.18}$$

where $\boldsymbol{p}_k = (p_{1k}, p_{2k}, \ldots, p_{Nk})^\top$ and with some abuse of the notation $\boldsymbol{\beta}$ stands now for the augmented $l+1$-dimensional vector $(\beta_0, \boldsymbol{\beta}^\top)^\top$. Also $\mathcal{U}_k$ denotes $(U_{0k}, \boldsymbol{U}_k^\top)^\top$. The $N \times (l+1)$-dimensional matrix $\boldsymbol{B}$ has the row structure such that the i-row is $(1, \boldsymbol{B}_i^\top)$. One needs to estimate the unknown parameters $\boldsymbol{\beta}$, $\sigma_{\mathcal{U}}^2$, $\boldsymbol{\Sigma}_{\mathcal{U}}$, σ_ϵ^2 and predict the random effect $\mathcal{U}_k$. This is done by using the joint distribution of $(\mathcal{U}_k, \boldsymbol{p}_k)^\top$ given by an $(l+1+N)$-dimensional Gaussian distribution (4.18),

$$\mathbb{N}_{l+1+N}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{B}\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \overline{\boldsymbol{\Sigma}}_{\mathcal{U}} & \overline{\boldsymbol{\Sigma}}_{\mathcal{U}}\boldsymbol{B}^\top \\ \boldsymbol{B}\overline{\boldsymbol{\Sigma}}_{\mathcal{U}} & \boldsymbol{\mathcal{F}} \end{bmatrix} \right), \tag{4.19}$$

where

$$\boldsymbol{\mathcal{F}} = \boldsymbol{B}\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}\boldsymbol{B}^\top + \boldsymbol{\Sigma}_\epsilon, \tag{4.20}$$

denotes the covariance of the data vector $\boldsymbol{p}_k$, and

$$\overline{\boldsymbol{\Sigma}}_{\mathcal{U}} = \begin{bmatrix} \sigma_{\mathcal{U}}^2 & 0 \\ 0 & \boldsymbol{\Sigma}_{\mathcal{U}} \end{bmatrix}, \tag{4.21}$$

is the covariance of the augmented vector $(U_{0k}, \boldsymbol{U}_k^\top)^\top$. The form of the covariance in (4.19) can be confirmed by observing first that

$$cov[\boldsymbol{p}_k] = E[\boldsymbol{p}_k\boldsymbol{p}_k^\top] - \boldsymbol{B}\boldsymbol{\beta}\boldsymbol{\beta}^\top\boldsymbol{B}^\top.$$

Then, owing to (4.18), one can write $\boldsymbol{p}_k^\top = \boldsymbol{\beta}^\top \boldsymbol{B}^\top + \mathcal{U}_k^\top \boldsymbol{B}^\top + \boldsymbol{\epsilon}_k^\top$, and by a simple algebra obtain

$$cov[\boldsymbol{p}_k] = \boldsymbol{B}\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}\boldsymbol{B}^\top + \boldsymbol{\Sigma}_\epsilon,$$

where $\boldsymbol{\Sigma}_\epsilon = \mathrm{diag}(\sigma_\epsilon^2, \ldots, \sigma_\epsilon^2)$ is the covariance matrix of the noise process and $\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}$ is the covariance matrix of the augmented random vector $(U_{0k}, \boldsymbol{U_k}^\top)^\top$. Finally note that

$$cov[\mathcal{U}_k, \boldsymbol{p}_k] = cov[\mathcal{U}_k, \boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{B}\mathcal{U}_k + \epsilon_k] = \overline{\boldsymbol{\Sigma}}_{\mathcal{U}}\boldsymbol{B}^\top,$$

whereas $cov[\boldsymbol{p}_k, \mathcal{U}_k] = \boldsymbol{B}\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}$, while $\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}\boldsymbol{B}^\top$ is the $(l+1) \times N$ matrix. The result in (4.19) is used to find the marginal distribution of $\boldsymbol{p}_k$ as a Gaussian distribution of the following form

$$\boldsymbol{p}_k \sim \mathbb{N}_N(\boldsymbol{B}\boldsymbol{\beta}, \boldsymbol{\mathcal{F}}). \tag{4.22}$$

Specializing this to the individual component of $\boldsymbol{p}_k$, i.e., p_{ik}, results in $p_{ik} \sim \mathbb{N}(\beta_0 + \beta^\top \boldsymbol{B}_i, \sigma_{\mathcal{U}}^2 + \boldsymbol{B}_i^\top \boldsymbol{\Sigma}_{\mathcal{U}} \boldsymbol{B}_i + \sigma_\epsilon^2)$, where now the original (not-augmented) notations $(\beta_0, \boldsymbol{\beta}^\top)^\top$ and $(U_{0k}, \boldsymbol{U}_k^\top)^\top$ are employed.

Predicting the random effects $\mathcal{U}_k$ from the observed data $\boldsymbol{p}_k$ needs the projection distribution of $\mathcal{U}_k$ onto $\boldsymbol{p}_k$. Owing to the standard multivariate normal theory (Rao and Toutenburg, 1995), it is known this is the normal distribution characterized by the conditional mean

$$E[\mathcal{U}_k|\boldsymbol{p}_k] = \overline{\boldsymbol{\Sigma}}_{\mathcal{U}}\boldsymbol{B}^\top\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{p}_k - \boldsymbol{B}\boldsymbol{\beta}), \tag{4.23}$$

and the conditional covariance

$$cov[\mathcal{U}_k|\boldsymbol{p}_k] = \overline{\boldsymbol{\Sigma}}_{\mathcal{U}} - \overline{\boldsymbol{\Sigma}}_{\mathcal{U}}\boldsymbol{B}^\top\boldsymbol{\mathcal{F}}^{-1}\boldsymbol{B}\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}.$$

Hence the conditional mean represents the optimal predictor of $\mathcal{U}_k$ from the observed data $\boldsymbol{p}_k$. However, to use this in practice, one needs to estimate $\boldsymbol{\beta}$, $\boldsymbol{\mathcal{F}}$, $\sigma^2_{\mathcal{U}}$, and $\boldsymbol{\Sigma}_{\mathcal{U}}$. The latter, i.e., $\sigma^2_{\mathcal{U}}$ and $\boldsymbol{\Sigma}_{\mathcal{U}}$ uniquely define the matrix $\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}$ in (4.21). The maximum likelihood estimators of these parameters will be presented in the following section.

Fig 4.3(c) shows the random slope B-spline aggregated model fitted to our real data set. Although it is expected to achieve more accurate power curves using random slope models, such models are more complex. So, it is suggested to compare both the accuracy and complexity of models in any experiment and decide to choose the best model.

4.3.4 Parameter Estimation

The distribution theory developed in the previous section plays a critical role in obtaining estimates of the unknown model parameters. Using (4.22) one can form the likelihood function $L(\boldsymbol{p}_k; \boldsymbol{\beta}, \boldsymbol{\mathcal{F}}) = f(\boldsymbol{p}_k; \boldsymbol{\beta}, \boldsymbol{\mathcal{F}})$, where $f(\boldsymbol{p}_k; \boldsymbol{\beta}, \boldsymbol{\mathcal{F}})$ stands for the N-dimensional Gaussian density with the parameters $\boldsymbol{\beta}, \boldsymbol{\mathcal{F}}$ defined in (4.22). Then, the maximum log-likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\mathcal{F}}$ are obtained by maximizing

$$l(\boldsymbol{p}_k; \boldsymbol{\beta}, \boldsymbol{\mathcal{F}}) = \log|\boldsymbol{\mathcal{F}}^{-1}| - (\boldsymbol{p}_k - \boldsymbol{B}\boldsymbol{\beta})^\top \boldsymbol{\mathcal{F}}^{-1} (\boldsymbol{p}_k - \boldsymbol{B}\boldsymbol{\beta}), \tag{4.24}$$

where the constant terms are dropped. The direct maximization of $l(\boldsymbol{p}_k; \boldsymbol{\beta}, \boldsymbol{\mathcal{F}})$ in (4.24) with respect to $\boldsymbol{\beta}, \boldsymbol{\mathcal{F}}$ can be a complex task and one can divide this into the two steps by employing the concept of the profiled log-likelihood. The profiled log-likelihood for the parameter $\boldsymbol{\mathcal{F}}$ is the maximized value of $l(\boldsymbol{p}_k; \boldsymbol{\beta}, \boldsymbol{\mathcal{F}})$ with respect to $\boldsymbol{\beta}$, i.e.,

$$l(\boldsymbol{p}_k; \boldsymbol{\mathcal{F}}) = \max_{\boldsymbol{\beta}} l(\boldsymbol{p}_k; \boldsymbol{\beta}, \boldsymbol{\mathcal{F}}). \tag{4.25}$$

The result of this maximization is the value of $\boldsymbol{\beta} = \boldsymbol{\beta}(\boldsymbol{\mathcal{F}})$ being the maximum likelihood estimate of $\boldsymbol{\beta}$ for fixed $\boldsymbol{\mathcal{F}}$. The solution of (4.25) is equivalent to minimization of the second term in (4.24)

$$(\boldsymbol{p}_k - \boldsymbol{B}\boldsymbol{\beta})^\top \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{p}_k - \boldsymbol{B}\boldsymbol{\beta}).$$

This represents the weighted least squared criterion for which the minimum is achieved (Ruppert et al., 2003) by

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\mathcal{F}}) = (\boldsymbol{B}^\top \boldsymbol{\mathcal{F}}^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}^\top \boldsymbol{\mathcal{F}}^{-1}\boldsymbol{p}_k. \tag{4.26}$$

This plugged back into (4.24) would give the profile log-likelihood function depending merely on $\boldsymbol{\mathcal{F}}$, i.e., the formula in (4.25) becomes

$$l(\boldsymbol{p}_k; \boldsymbol{\mathcal{F}}) = \log|\boldsymbol{\mathcal{F}}^{-1}| - (\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}(\boldsymbol{\mathcal{F}}))^\top \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}(\boldsymbol{\mathcal{F}})), \tag{4.27}$$

where $\hat{\boldsymbol{\beta}}(\boldsymbol{\mathcal{F}})$ is defined in (4.26). Maximization of this with respect to $\boldsymbol{\mathcal{F}}$ would be rather involved as $l(\boldsymbol{p}_k; \boldsymbol{\mathcal{F}})$ in (4.27), is a highly nonlinear function of $\boldsymbol{\mathcal{F}}$. To ease the computational complexity on the obtained nonlinear optimization problem, an iterative method using the semi-likelihood strategy can be implemented. Hence, let $\widehat{\boldsymbol{\mathcal{F}}}_0$ be an initial estimate of $\boldsymbol{\mathcal{F}}$. Then from (4.26), one can get the preliminary estimate of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}}_0 = \hat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\mathcal{F}}}_0) = (\boldsymbol{B}^\top \widehat{\boldsymbol{\mathcal{F}}}_0^{-1} \boldsymbol{B})^{-1} \boldsymbol{B}^\top \widehat{\boldsymbol{\mathcal{F}}}_0^{-1} \boldsymbol{p}_k. \tag{4.28}$$

The semi-likelihood strategy is to form the following version of (4.27)

$$l(\boldsymbol{p}_k; \boldsymbol{\mathcal{F}}; \widehat{\boldsymbol{\mathcal{F}}}_0) = \log |\boldsymbol{\mathcal{F}}^{-1}| - (\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0))^\top \boldsymbol{\mathcal{F}}^{-1} (\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0). \tag{4.29}$$

Hence, only the regression part in (4.27), i.e., $\hat{\boldsymbol{\beta}}(\boldsymbol{\mathcal{F}})$ is fixed, whereas the remaining part of the criterion is still a function of $\boldsymbol{\mathcal{F}}$. As a result, it suffices to find $\boldsymbol{\mathcal{F}}$ that maximizes the semi-likelihood criterion in (4.29). The following lemma plays the important role here.

Lemma 1 *The criterion $l(\boldsymbol{p}_k; \boldsymbol{\mathcal{F}}; \widehat{\boldsymbol{\mathcal{F}}}_0)$ in (4.29) is maximized by*

$$\widehat{\boldsymbol{\mathcal{F}}}_1 = (\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0)(\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0)^\top. \tag{4.30}$$

You can find the proof in the Appendix.

The aforementioned result reveals that the next value of the estimate of $\boldsymbol{\mathcal{F}}$ is $\widehat{\boldsymbol{\mathcal{F}}}_1$ given in (4.30). This updates the estimate of $\boldsymbol{\beta}$ in (4.28) to the new value $\hat{\boldsymbol{\beta}}_1$, where $\widehat{\boldsymbol{\mathcal{F}}}_0$ is replaced by $\widehat{\boldsymbol{\mathcal{F}}}_1$. With the updated $\hat{\boldsymbol{\beta}}_1$, we return back to the criterion in (4.29) and find the new version $\widehat{\boldsymbol{\mathcal{F}}}_2$ of $\widehat{\boldsymbol{\mathcal{F}}}_1$ that maximizes $l(\boldsymbol{p}_k; \boldsymbol{\mathcal{F}}; \widehat{\boldsymbol{\mathcal{F}}}_1)$. This iterative process can continue till some stopping rule is met. The above profiled likelihood iterative procedure produces the maximum likelihood estimates $\hat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\mathcal{F}}}$, of $\boldsymbol{\beta}$ and $\boldsymbol{\mathcal{F}}$, respectively. This can be in turn use in (4.23) to form the empirical counterpart of the linear predictor of the random effect. Hence, one can write

$$\hat{\mathcal{U}}_k = \widehat{\overline{\boldsymbol{\Sigma}}}_{\mathcal{U}} \boldsymbol{B}^\top \widehat{\boldsymbol{\mathcal{F}}}^{-1} (\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}),$$

where $\widehat{\overline{\boldsymbol{\Sigma}}}_{\mathcal{U}}$ is an estimate of the matrix $\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}$ given in (4.21). The estimate of $\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}$ can be obtained by recalling that, see (4.20), $\boldsymbol{\mathcal{F}} = \boldsymbol{B}\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}\boldsymbol{B}^{\top} + \boldsymbol{\Sigma}_{\epsilon}$. First of all, the noise variance σ_{ϵ}^2 defining the diagonal matrix $\boldsymbol{\Sigma}_{\epsilon}$ can be estimated by the residual process or any method being independent on a regression function form (Dette et al., 1998). This leads to the model free estimate $\widehat{\boldsymbol{\Sigma}}_{\epsilon}$ of $\boldsymbol{\Sigma}_{\epsilon}$.

Next, the aforementioned estimates $\widehat{\boldsymbol{\mathcal{F}}}$ and $\widehat{\boldsymbol{\Sigma}}_{\epsilon}$ can form the estimating equation $\widehat{\boldsymbol{\mathcal{F}}} = \boldsymbol{B}\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}\boldsymbol{B}^{\top} + \widehat{\boldsymbol{\Sigma}}_{\epsilon}$. The solution of this yields the required estimate of $\overline{\boldsymbol{\Sigma}}_{\mathcal{U}}$.

4.4 Real Data Application

In this section, the empirical data on a wind farm in Canada was used to evaluate the performance of random intercept and random slope B-spline aggregated wind farm models compared with a single level model. The complexity of fitted models is also studied. The wind farm includes 63 wind turbines the rated capacity equals of 1.7 MW, the hub height of 80 m spread over an area of over 90 km^2, the cut-in speed of 3.5 m/s, rated speed of 13 m/s, and cut-out speed of 20 m/s, respectively. The data set contains more than 1.2 million pairs of wind speed and generated power data points with a resolution equal to 10-minutes from June 2006 to January 2007. The data set for June was only used for the clustering purpose and not the power curve fitting process to avoid adding bias in estimating wind farm aggregated models. Features for each turbine were obtained using the data in June and were used for clustering turbines into three different groups. Finally, single level, random

Table 4.2: Results for 10-min data from a wind farm for Single level, Random Intercept, and Random Slopes B-spline wind farm aggregated models

Model	MAE (±SD)	RMSE (±SD)	Rank
Single Model	64.9 (±0.08) (3)	88.87 (±0.22) (3)	(3)
Random Intercept	60.83 (±0.05) (2)	83.43 (±0.19) (2)	(2)
Random Slopes	59.9 (±0.06) (1)	82.42 (±0.19) (1)	(1)

Table 4.3: Degree of Freedom and likelihood ratio tests for testing Random Intercept, and Random Slopes B-spline wind farm aggregated models against a single level model

Model	Degree of Freedom	ChiSq	DF	Pr(>ChiSq)
Single Model	7			
Random Intercept	8	56625	1	$< 2.2e^{-16}$
Random Slopes	28	67409	21	$< 2.2e^{-16}$

intercept, and random slope B-spline aggregated models were developed using the remaining portion of the data set.

Mean absolute error (MAE) and root mean squared error (RMSE) were used as two statistical metrics to evaluate the accuracy of fitted models:

$$RMSE = \sqrt{\frac{1}{n} \sum \sum (p_{ik} - \hat{p}_{ik})^2}, \tag{4.31}$$

$$MAE = \frac{1}{n} \sum \sum |p_{ik} - \hat{p}_{ik}|, \tag{4.32}$$

where p_{ik} and $\hat{p}_{ik}$ are defined as in Section 4.3. Table 4.2 shows the results of the analysis as well as the rank of each method in the parenthesis, where the smaller values are desirable. As shown in Table 4.2, the random slope B-spline aggregated model is the most accurate one with 7.2% RMSE improvement over the single model. Random intercept model also improves the RMSE metric by 6.2%. According to

Table 4.2, one can conclude that all mixed effect models outperform the single level model. The complexity of each model is shown based on its degree of freedom in Table 4.3. Note that the random intercept model's degree of freedom is just one more than the single level model. However, the accuracy of the random intercept model is significantly more than the single model. It is worth mentioning that the degree of freedom for fitting separate models for each turbine will be equal to the number of turbines in the farm times fitted model's degree of freedom, which equals to $63 \times 7 = 441$ in our data set.

To investigate the significance of the improvement of mixed effect models, these models were tested against a single level model using the ANOVA as well as likelihood ratio tests. The values represented under Chisq in Table 4.3 show the difference between the log-likelihood for each mixed effect model and the single model. The DF column shows the difference between the number of parameters in mixed effect models and the single model. According to Table 4.3, one can conclude that the removal of the random intercept parameter for clusters is not justified (ChiSquare : 56625; p-value:$< 2.2e^{-16}$). The likelihood ratio tests show that the random intercept has the right level of complexity and accuracy for this data set. Consequently, the ANOVA test shows that the random intercept model is outperforming the single level model. The same p-value resulted in the random slope model as well, which shows mixed effect models outperform the single level model. However, the random slope model's complexity is much more than the random intercept models. One can conclude that utilizing the mixed effect models, especially the random intercept model, for wind farm aggregated model can be beneficial for power forecasting, maintenance management, power system reliability assessment, and sizing the storage capacity for wind power integration since the accuracy improves with the negligible increase of model's complexity.

4.5 Concluding Remarks

A significant challenge in managing wind farms is simulating the wind farm power curve to forecast power in the future. Using a separate wind power curve for each turbine will make the model complex and causes a substantial computation cost. On the other hand, considering just one wind farm aggregated power curve is not accurate enough. In this paper, a novel clustering method was proposed to assign wind turbines into several similar groups and then develop mixed effect models for wind farms to take into account the correlation between different groups. To this end, a clustering method was developed based on the turbines' performance in terms of generating power. The feature set, which is based on the turbines' performance, takes into account not only the turbine's age and location but also considers the wake effect of other surrounding turbines in the wind farm. In addition to the clustering method, mixed effect models were used to develop wind farm aggregate models and improve the accuracy of power forecasting. The performance of the presented methods was evaluated on a wind farm located in Canada. Experimental results show that utilizing the random intercept model on clustered turbines significantly outperforms a single model by keeping the model's complexity at the same level. The proposed methods in this paper can be used for wind farms with different layouts and diverse wake effects, operation conditions, or geographical characteristics of terrain. It can also be used in power system reliability assessment Bagen and Billinton (2008). Results show that such aggregated models provide better representations of the power generated by the wind farm studied in this paper. Results also show that the random intercept model reduces the time-computing cost more than 90% comparing to fitting separate wind turbine power curves for each turbine in the wind farm.

Appendix

Proof of Lemma 1: The proof is based on the standard matrix calculus. First let us substitute $\boldsymbol{Q} = \boldsymbol{\mathcal{F}}^{-1}$. Then, (4.27) becomes

$$l(\boldsymbol{p}_k; \boldsymbol{Q}; \widehat{\boldsymbol{\mathcal{F}}}_0) = \log|\boldsymbol{Q}| - Tr[\boldsymbol{Q}(\boldsymbol{p}_k - B\hat{\boldsymbol{\beta}}_0)(\boldsymbol{p}_k - B\hat{\boldsymbol{\beta}}_0)^\top], \tag{4.33}$$

where the second term in (4.29) was represented in terms of the trace operator. This is based on the known fact that $\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} = Tr[\boldsymbol{Q}\boldsymbol{x}\boldsymbol{x}^\top]$ for any vector $\boldsymbol{x}$. Next, we use the following facts (Rao and Toutenburg, 1995) that

$$\frac{\partial \log|\boldsymbol{Q}|}{\partial \boldsymbol{Q}} = 2\boldsymbol{Q}^{-1} - diag[\boldsymbol{Q}^{-1}]$$

and $\frac{\partial Tr[\boldsymbol{Q}(\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0)(\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0)^\top]}{\partial \boldsymbol{Q}}$ given by

$$2(\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0)(\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0)^\top - diag[(\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0)(\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0)^\top]$$

where $diag[\boldsymbol{A}]$ denotes the diagonal matrix consisting of the diagonal elements of the matrix $\boldsymbol{A}$. The above facts combined with Eq. (4.33) and $\boldsymbol{Q} = \boldsymbol{\mathcal{F}}^{-1}$ show that

$$\frac{\partial l(\boldsymbol{p}_k; \boldsymbol{Q}; \hat{\boldsymbol{\mathcal{F}}}_0)}{\partial \boldsymbol{Q}} = 2\boldsymbol{\Lambda} - diag[\boldsymbol{\Lambda}],$$

where $\boldsymbol{\Lambda} = \boldsymbol{\mathcal{F}} - (\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0)(\boldsymbol{p}_k - \boldsymbol{B}\hat{\boldsymbol{\beta}}_0)^\top$. The proof is completed by noting that $\frac{\partial l(\boldsymbol{p}_k; \boldsymbol{Q}; \widehat{\boldsymbol{\mathcal{F}}}_0)}{\partial \boldsymbol{Q}} = 0$ is equivalent to $2\boldsymbol{\Lambda} - diag[\boldsymbol{\Lambda}] = 0$ and this holds if $\boldsymbol{\Lambda} = 0$.

Chapter 5

Conclusions and Future Work

This chapter represents the accomplishments of the research work studied in the thesis. In addition to concluding remarks from each chapter, suggested future works will also be presented.

5.1 Summary of Accomplishments

Several approaches were proposed to improve wind turbine power curve modeling, and consequently increase the reliability of wind power integrated into the power system. First, we proposed two different types of power curves estimators that preserve the monotonicity constraint. We compared them with other commonly used power curves with a special focus on the similarity of the shape of fitted curves to the manufacturer's power curve. We proposed the tilting method and monotone spline regression to address the non-monotonicity issue of the fitted power curve and reduce the sensitivity of the fitted curves to outliers. The proposed tilting method is a generic approach that can be used to enforce the monotonicity on any kernel estimator, such as the Nadaraya-Watson kernel estimator. In addition to the tilting method, we developed a monotone spline regression and compared the tilting method's performance, monotone spline regression, Nadaraya-Watson kernel estimator, and natural spline regression on data obtained from four different wind turbines. We concluded that monotone spline regression performs the best since it preserves the monotonicity and is an accurate power curve estimator compared to other analyzed methods. Also, our results confirm that using the tilting method compared to kernel methods performs almost the same in terms of accuracy, whereas its shape is similar to the manufacturer's power curve, which makes it less sensitive to anomalies within the set of observations. We also concluded that both proposed methods are robust to outliers which make them accurate models to be used as reference power curve in the development of algorithms to detect anomalies. Since monotone power curves are not sensitive to outliers, one can easily detect all data which are further than a threshold as anomalies. This early detection is helpful to reduce unscheduled downtime and operational costs. We also conclude that, in contrast to most of other studies which manually remove a high portion of data as outliers, it is not required to do so for monotone empirical power curve. However, it is recommended to just remove outliers which are obvious such as data with negative wind power or speed, or data with wind power equal to almost zero at high wind speed.

Furthermore, we developed a statistical algorithm to calculate the uncertainty of a fitted power curve on the data according to wind power variance in different wind speed ranges. The developed algorithm contributes to the goal of handling the heteroscedasticity issue of wind data. The heteroscedasticity effect of wind data causes a high variance in fitted power curves. The introduced algorithm presents a hybrid method to combine an empirical power curve fitted on the data and a target power curve selected as a reliable resource. Two different weighting schemes are defined according to the wind power variance to give appropriate weights to the contribution of empirical and target power curves on the resulted hybrid method. The results show that utilizing target models in hybrid methods can reduce the variance of the hybrid fitted curve. It is effective since the target models shrink the hybrid method toward itself in the regions where the variance of data is high. High variance of data occurs more for those turbines which are far from meteorological masts and have more outliers in their speed-power data set. We concluded that hybrid methods is an effective solution to handle the high variance issue for these turbines. Also, hybrid methods are capable of reducing the effect of outliers. To summarize, hybrid methods reduce both estimation bias and variance compared to commonly used regression methods such as the spline method and local regression method.

So, it is suggested to take advantage of hybrid methods in those wind farms which suffer from non-homogeneous data set.

Finally, a novel feature selection is proposed to consider turbines' overall performance for clustering turbines into similar groups. Using this clustering method, turbines with similar power production would be clustered into the same group. Applying the elbow method, which is a well accepted approach for selecting the appropriate number of clusters, turbines were clustered into three groups. Consequently, a novel multiple aggregated model was proposed for wind farm power curve modeling. Multiple aggregated models studied in the literature have ignored the hidden correlation between fitted curves in each cluster of turbines. To consider the correlation between the aggregated power curve in each cluster, we applied a multilevel modeling approach by considering cluster as the parent level and turbines as the child level. Random intercept and random slope approaches were compared on a data set containing 63 turbines. The random slope model performed better in terms of accuracy; however, its complexity was higher than the random intercept. It was concluded that random intercept, which reduces the complexity in terms of the degree of freedom by 71%, compared to random slope, and with accuracy, almost the same, would be a better option for wind farm power curve modeling. It is worth mentioning that both random intercept and random slope methods outperform the single aggregated method by at least 6.7% in terms of MAE and RMSE accuracy metrics. Also, ANOVA tests confirmed that mixed effect models would significantly outperform the single aggregated model for this wind farm. In conclusion, it is recommended to utilize clustering methods based on the power output performance of turbines and applying multilevel modeling for multiple aggregated models to not only reduce the complexity of the model but also take into account the hidden correlation between power curves of each cluster. Finally, it is concluded to utilize multilevel modeling approach as an accurate and fast multiple aggregate model for fitting wind farm power curve in applications such as annual wind energy prediction.

5.2 Future Work

The suggested future works in this thesis are presented in the following directions:

5.2.1 Monotonic Power Curve with Multiple Auxiliary Information for Fault Detection

Available methods in the literature for wind turbine power curve modeling primarily focus on parametric and nonparametric approaches using wind speed as the only auxiliary information. However, with the new system such as SCADA data, more auxiliary information is available to consider in fitting empirical power curves. Recently, some approaches in the literature were proposed to take into account this auxiliary information. As an extension to this work, we suggest to consider multiple auxiliary information and enforce manufacturer power curve shape in fitting empirical power curves by including some other environmental factors such as air density, humidity, wind direction, and turbulence intensity in the estimation process to more realistically capture the nonlinear relationships between these environmental factors and the generated wind power. As we showed in chapter 2, enforcing shape properties of the manufacturer's power curve in fitting the empirical power curve can reduce outliers impact. Since the monotone power curve is similar to the manufacturer's power curve, it is expected that the turbine normal generating power would not be far from the monotone power curve. When any fault occurs, the generated power differs from the expectation based on the empirical power curve. An interesting future research direction is to use monotonic fitted power curves for fault detection and condition monitoring. Fault detection and condition monitoring has direct impact on power system reliability evaluation. Consequently, one can evaluate the effectiveness of utilizing monotonic wind power curve modeling for power system reliability as a future research.

5.2.2 Cross-Cluster Multilevel Modeling

An immediate future work regarding our proposed multilevel modeling approach is to cluster turbines based on multiple feature sets and clustering models. This will lead to a situation that one turbine belongs to a cluster based on a feature set, and that turbine belongs to a different cluster based on another feature set. In this case, there would be a cross-cluster multilevel structure. In other words, turbines t_1 and t_2 may be inside the same cluster according to one feature set (e.g., wind direction) and be in a different cluster according to another feature set (e.g., wind turbine location). Fig 5.1 represents such a case where turbines may belong to different clusters based on different features. Cross-cluster multilevel models can be investigated as wind farm multiple aggregated model for improving the accuracy of

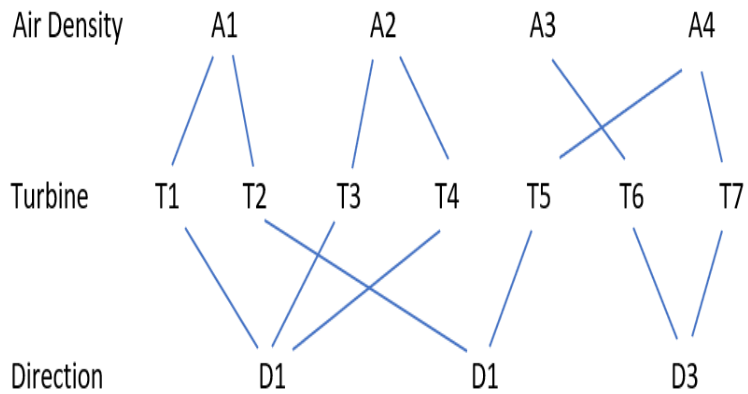


Figure 5.1: Cross-cluster shows how a turbine belongs to multiple clusters based on different features such as direction and air density in this example.

the aggregated power curve by taking into account the hidden correlation of power curves in different clusters.

5.2.3 Asymmetric Loss Function for Power Curve Modeling

The reliability associated with a power system is a measure of the system's ability to perform its intended function. The main goal of a power system is to supply its customers with electrical energy with an acceptable degree of reliability and economic costs. One can elaborate on different penalties for over and under predicting the available power in the future. The risk of underestimating the power is that it can imperil the property of supplying energy as economically as possible. On the other hand, over predicting will jeopardize the whole reliability of the system. So it is clear that different penalties should be used for over prediction as opposed to under prediction. A possible future research direction is to use an asymmetric prediction error function to deal with different penalties for under or over prediction of generated power at a specific wind speed. New methodologies need to be developed to study the performance of the fitted power curve under an asymmetric prediction error function.

Bibliography

Ackermann, T. (2005). *Wind Power in Power Systems*. John Wiley & Sons. (Cited on pages 11 and 13.)

Ackermann, T. and L. Söder (2000). Wind energy technology and current status: a review. *Renewable and Sustainable Energy Reviews 4*(4), 315–374. (Cited on pages 11 and 78.)

Ai, B., H. Yang, H. Shen, and X. Liao (2003). Computer-aided design of pv/wind hybrid system. *Renewable Energy 28*(10), 1491–1512. (Cited on page 43.)

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics 21*(1), 243–247. (Cited on page 8.)

Al-Shammari, E. T., S. Shamshirband, D. Petković, E. Zalnezhad, L. Yee, R. S.

Taher, and Ž. Ćojbašić (2016). Comparative study of clustering methods for wake effect analysis in wind farm. *Energy 95*, 573–579. (Cited on page 28.)

Ali, M., I.-S. Ilie, J. V. Milanovic, and G. Chicco (2012). Wind farm model aggregation using probabilistic clustering. *IEEE Transactions on Power Systems 28*(1), 309–316. (Cited on pages 27, 28 and 108.)

Alotto, P., M. Guarnieri, and F. Moro (2014). Redox flow batteries for the storage of renewable energy: A review. *Renewable and Sustainable Energy Reviews 29*, 325–335. (Cited on page 18.)

Amjady, N., F. Keynia, and H. Zareipour (2011). Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization. *IEEE Transactions on Sustainable Energy 2*(3), 265–276. (Cited on page 8.)

Association, E. W. E. et al. (2010). Powering europe: wind energy and the electricity grid. *European Wind Energy Association, Brussels, Belgium*. (Cited on page 5.)

Association, E. W. E. et al. (2012). *Wind energy-the facts: a guide to the technology, economics and future of wind power*. Routledge. (Cited on page 12.)

Bagen, B. and R. Billinton (2008). Reliability cost/worth associated with wind energy and energy storage utilization in electric power systems. In *Probabilistic Methods Applied to Power Systems, 2008. PMAPS'08. Proceedings of the 10th International Conference on*, pp. 1–7. IEEE. (Cited on page 42.)

Banitalebi, B., S. S. Appadoo, and A. Thavaneswaran (2020). Data driven approach for reduced value at risk forecasts in renewable power supply systems. In *Proceedings of Canadian Conference on Electrical and Computer Engineering*. IEEE. (Cited on page 3.)

Baran, J. and A. Stepien-Baran (2013). Sequential estimation of a location parameter and powers of a scale parameter from delayed observations. *Statistica Neerlandica 67*(3), 263–280. (Cited on page 76.)

Ben-Gal, I. (2005). Outlier detection. In *Data Mining and Knowledge Discovery Handbook*, pp. 131–146. Springer. (Cited on page 23.)

Bencherif, M., B. Brahmi, and A. Chikhaoui (2014). Optimum selection of wind turbines. *Sci. J. Energy Eng 2*(4), 36. (Cited on page 16.)

Besnard, F., M. Patrikssont, A.-B. Strombergt, A. Wojciechowskit, and L. Bertling (2009). An optimization framework for opportunistic maintenance of offshore wind power system. In *2009 IEEE Bucharest PowerTech*, pp. 1–7. IEEE. (Cited on page 6.)

Betz, A. (1920). Das maximum der theoretisch möglichen ausnutzung des windes durch windmotoren. *Zeitschrift fur das gesamte Turbinenwesten 20*. (Cited on page 11.)

Bhaskar, K. and S. Singh (2012). Awnn-assisted wind power forecasting using feed-forward neural network. *IEEE Transactions on Sustainable Energy 3*(2), 306–315. (Cited on page 8.)

Billinton, R. et al. (2006). Reliability considerations in the utilization of wind energy, solar energy and energy storage in electric power systems. In *Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on*, pp. 1–6. IEEE. (Cited on page 6.)

Bloch, D. A. and B. W. Silverman (1997). Monotone discriminant functions and their applications in rheumatology. *Journal of the American Statistical Association 92*(437), 144–153. (Cited on pages 54 and 55.)

Botterud, A., J. Wang, V. Miranda, and R. J. Bessa (2010). Wind power forecasting in us electricity markets. *The Electricity Journal 23*(3), 71–82. (Cited on page 15.)

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons. (Cited on page 8.)

Burman, P. and P. Chaudhuri (2012). On a hybrid approach to parametric and nonparametric regression. In *Nonparametric Statistical Methods and Related Topics: A Festschrift in Honor of Professor PK Bhattacharya on the Occasion of his 80th Birthday*, pp. 233–256. World Scientific. (Cited on page 85.)

Carrillo, C., A. O. Montaño, J. Cidrás, and E. Díaz-Dorado (2013). Review of power curve modelling for wind turbines. *Renewable and Sustainable Energy Reviews 21*, 572–581. (Cited on pages 43, 52, 74 and 118.)

Carroll, R. J. and D. Ruppert (1982). Robust estimation in heteroscedastic linear models. *The Annals of Statistics*, 429–441. (Cited on page 25.)

Castillo, A. and D. F. Gayme (2014). Grid-scale energy storage applications in renewable energy integration: A survey. *Energy Conversion and Management 87*, 885–894. (Cited on page 18.)

Catalão, J. P. d. S., H. M. I. Pousinho, and V. M. F. Mendes (2011). Short-term wind power forecasting in portugal by neural networks and wavelet transform. *Renewable Energy 36*(4), 1245–1251. (Cited on page 27.)

Chang, T.-P., F.-J. Liu, H.-H. Ko, S.-P. Cheng, L.-C. Sun, and S.-C. Kuo (2014). Comparative analysis on power curve models of wind turbine generator in estimating capacity factor. *Energy 73*, 88–95. (Cited on page 73.)

Changshui, Z., H. Guangdong, and W. Jun (2011). A fast algorithm based on the submodular property for optimization of wind turbine positioning. *Renewable Energy 36*(11), 2951–2958. (Cited on page 28.)

Chen, Y., H. Li, K. Jin, and Q. Song (2013). Wind farm layout optimization using genetic algorithm with different hub height wind turbines. *Energy Conversion and Management 70*, 56–65. (Cited on page 29.)

Chen, Z., S. Zhu, Q. Niu, and T. Zuo (2020). Knowledge discovery and recommendation with linear mixed model. *IEEE Access 8*, 38304–38317. (Cited on page 109.)

Chowdhury, H., I. Mustary, B. Loganathan, and F. Alam (2015). Adjacent wake effect of a vertical axis wind turbine. *Procedia Engineering 105*, 692–697. (Cited on page 30.)

Chowdhury, M., W. Shen, N. Hosseinzadeh, and H. Pota (2013). A novel aggregated dfig wind farm model using mechanical torque compensating factor. *Energy Conversion and Management 67*, 265–274. (Cited on page 27.)

Cocina, V., P. Di Leo, M. Pastorelli, and F. Spertino (2015). Choice of the most suitable wind turbine in the installation site: A case study. In *2015 International Conference on Renewable Energy Research and Applications (ICRERA)*, pp. 1631–1634. IEEE. (Cited on page 16.)

Commission, I. E. et al. (2005). Wind turbines-part 12-1: Power performance measurements of electricity producing wind turbines. *IEC 61400-12-1*. (Cited on pages 13 and 26.)

Commission, I. E. et al. (2007). Wind energy generation systems—part 12-1: Power performance measurements of electricity producing wind turbines. Technical report, IEC 61400-12-1: 2017. (Cited on pages 73, 96 and 118.)

Cressie, N. and T. R. Read (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B (Methodological) 46*(3), 440–464. (Cited on page 56.)

Daki, H., A. El Hannani, A. Aqqal, A. Haidine, and A. Dahbi (2017). Big data management in smart grid: concepts, requirements and implementation. *Journal of Big Data 4*(1), 1–19. (Cited on page 23.)

De Gooijer, J. G. and R. J. Hyndman (2006). 25 years of time series forecasting. *International Journal of Forecasting 22*(3), 443–473. (Cited on page 8.)

Denholm, P. and M. Hand (2011). Grid flexibility and storage required to achieve very high penetration of variable renewable electricity. *Energy Policy 39*(3), 1817–1830. (Cited on page 18.)

Deshmukh, M. and S. Deshmukh (2008). Modeling of hybrid renewable energy systems. *Renewable and Sustainable Energy Reviews 12*(1), 235–249. (Cited on page 20.)

Dette, H., A. Munk, and T. Wagner (1998). Estimating the variance in nonparametric regression—what is a reasonable choice? *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(4), 751–764. (Cited on page 129.)

Dette, H. and K. F. Pilz (2006). A comparative study of monotone nonparametric kernel estimates. *Journal of Statistical Computation and Simulation 76*(1), 41–56. (Cited on page 54.)

Diaf, S., D. Diaf, M. Belhamel, M. Haddadi, and A. Louche (2007). A methodology for optimal sizing of autonomous hybrid pv/wind system. *Energy Policy 35*(11), 5708–5718. (Cited on page 43.)

Diaf, S., G. Notton, M. Belhamel, M. Haddadi, and A. Louche (2008). Design and techno-economical optimization for hybrid pv/wind system under various meteorological conditions. *Applied Energy 85*(10), 968–987. (Cited on page 20.)

Domingues, R., M. Filippone, P. Michiardi, and J. Zouaoui (2018). A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern Recognition 74*, 406–421. (Cited on page 23.)

Dörenkämper, M., B. Witha, G. Steinfeld, D. Heinemann, and M. Kühn (2015). The impact of stable atmospheric boundary layers on wind-turbine wakes within offshore wind farms. *Journal of Wind Engineering and Industrial Aerodynamics 144*, 146–153. (Cited on page 30.)

Dou, B., M. Guala, L. Lei, and P. Zeng (2019). Experimental investigation of the performance and wake effect of a small-scale wind turbine in a wind tunnel. *Energy 166*, 819–833. (Cited on page 108.)

Ekonomou, L., S. Lazarou, G. E. Chatzarakis, and V. Vita (2012). Estimation of wind turbines optimal number and produced power in a wind farm using an artificial neural network model. *Simulation Modelling Practice and Theory 21*(1), 21–25. (Cited on page 29.)

Elliott, D. and D. Infield (2014). An assessment of the impact of reduced averaging time on small wind turbine power curves, energy capture predictions and turbulence intensity measurements. *Wind Energy 17*(2), 337–342. (Cited on pages 44 and 66.)

Eroğlu, Y. and S. U. Seçkiner (2012). Design of wind farm layout using ant colony algorithm. *Renewable Energy 44*, 53–62. (Cited on page 29.)

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*, Volume 90. M. Dekker New York. (Cited on page 20.)

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. CRC Press. (Cited on page 85.)

Fernandez, L., C. Garcia, J. Saenz, and F. Jurado (2009). Equivalent models of wind farms by using aggregated wind turbines and equivalent winds. *Energy conversion and management 50*(3), 691–704. (Cited on page 27.)

Fernández, L. M., F. Jurado, and J. R. Saenz (2008). Aggregated dynamic model for wind farms with doubly fed induction generator wind turbines. *Renewable Energy 33*(1), 129–140. (Cited on page 27.)

Fernandez, L. M., J. R. Saenz, and F. Jurado (2006). Dynamic models of wind farms with fixed speed wind turbines. *Renewable Energy 31*(8), 1203–1230. (Cited on page 107.)

Friedman, J., T. Hastie, and R. Tibshirani (2001). *The Elements of Statistical*

Learning, Volume 1. Springer Series in Statistics New York. (Cited on pages 20, 51, 52, 82, 84, 114 and 115.)

Friedman, J. and R. Tibshirani (1984). The monotone smoothing of scatterplots. *Technometrics 26*(3), 243–250. (Cited on pages 54 and 55.)

Friedman, J. H. and B. W. Silverman (1989). Flexible parsimonious smoothing and additive modeling. *Technometrics 31*(1), 3–21. (Cited on pages 51 and 84.)

Gao, X., T. Wang, B. Li, H. Sun, H. Yang, Z. Han, Y. Wang, and F. Zhao (2019). Investigation of wind turbine performance coupling wake and topography effects based on lidar measurements and scada data. *Applied Energy 255*, 113816. (Cited on page 108.)

Gasch, R. and J. Twele (2011). *Wind Power Plants: Fundamentals, Design, Construction and Operation*. Springer Science & Business Media. (Cited on pages 9, 47, 78 and 118.)

Ghaedi, A., A. Abbaspour, M. Fotuhi-Firuzabad, and M. Moeini-Aghtaie (2013). Toward a comprehensive model of large-scale dfig-based wind farms in adequacy assessment of power systems. *IEEE Transactions on Sustainable Energy 5*(1), 55–63. (Cited on page 107.)

Ghosh, S. and N. Senroy (2013). Balanced truncation based reduced order modeling of wind farm. *International Journal of Electrical Power & Energy Systems 53*, 649–655. (Cited on page 107.)

Gill, S., B. Stephen, and S. Galloway (2011). Wind turbine condition assessment through power curve copula modeling. *IEEE Transactions on Sustainable Energy 3*(1), 94–101. (Cited on page 17.)

Giorsetto, P. and K. F. Utsurogi (1983). Development of a new procedure for reliability modeling of wind turbine generators. *IEEE Transactions on Power Apparatus and Systems* (1), 134–143. (Cited on page 20.)

González-Longatt, F., P. Wall, and V. Terzija (2012). Wake effect in wind farm performance: Steady-state and dynamic behavior. *Renewable Energy 39*(1), 329–338. (Cited on page 107.)

Grajeda, L. M., A. Ivanescu, M. Saito, C. Crainiceanu, D. Jaganath, R. H. Gilman, J. E. Crabtree, D. Kelleher, L. Cabrera, V. Cama, et al. (2016). Modelling subject-specific childhood growth using linear mixed-effect models with cubic regression splines. *Emerging themes in Epidemiology 13*(1), 1. (Cited on page 109.)

Greblicki, W. and M. Pawlak (2008). *Nonparametric System Identification*. Cambridge University Press Cambridge. (Cited on pages 52, 53 and 74.)

Hall, P., L.-S. Huang, et al. (2001). Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics 29*(3), 624–647. (Cited on pages 54 and 56.)

Hassan, G. G. (2013). A guide to uk offshore wind operations and maintenance. *Scottish Enterprise and The Crown Estate 21*. (Cited on page 1.)

Hollander, M., D. A. Wolfe, and E. Chicken (2013). *Nonparametric Statistical Methods*, Volume 751. John Wiley & Sons. (Cited on page 20.)

Hu, G. and P. Peng (2011). All admissible linear estimators of a regression coefficient under a balanced loss function. *Journal of Multivariate Analysis 102*(8), 1217–1224. (Cited on page 76.)

Hur, S.-h. (2018). Modelling and control of a wind turbine and farm. *Energy 156*, 360–370. (Cited on page 107.)

Jafari Jozani, M., É. Marchand, and A. Parsian (2006). On estimation with weighted balanced-type loss function. *Statistics & Probability Letters 76*(8), 773–780. (Cited on page 76.)

Jafari Jozani, M., É. Marchand, and A. Parsian (2012). Bayesian and robust bayesian analysis under a general class of balanced loss functions. *Statistical Papers 53*(1), 51–60. (Cited on pages 76 and 81.)

Jangamshetti, S. H. and V. G. Rau (2001). Normalized power curves as a tool for identification of optimum wind turbine generator parameters. *IEEE Transactions on Energy Conversion 16*(3), 283–288. (Cited on page 16.)

Jin, T. and Z. Tian (2010). Uncertainty analysis for wind energy production with dynamic power curves. In *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*, pp. 745–750. IEEE. (Cited on page 15.)

Jónsson, T., P. Pinson, and H. Madsen (2010). On the market impact of wind energy forecasts. *Energy Economics 32*(2), 313–320. (Cited on page 6.)

Jung, J. and R. P. Broadwater (2014). Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews 31*, 762–777. (Cited on pages 8 and 43.)

Karamichailidou, D., V. Kaloutsa, and A. Alexandridis (2020). Wind turbine power curve modeling using radial basis function neural networks and tabu search. *Renewable Energy*. (Cited on page 74.)

Khalfallah, M. G. and A. M. Koliub (2007). Suggestions for improving wind turbines power curves. *Desalination 209*(1-3), 221–229. (Cited on page 19.)

Kim, H., C. Singh, and A. Sprintson (2012). Simulation and estimation of reliability in a wind farm considering the wake effect. *IEEE Transactions on Sustainable Energy 3*(2), 274–282. (Cited on page 107.)

Kitajima, T. and T. Yasuno (2010). Output prediction of wind power generation system using complex-valued neural network. In *Proceedings of SICE Annual Conference 2010*, pp. 3610–3613. IEEE. (Cited on page 8.)

Kusiak, A. and W. Li (2011). The prediction and diagnosis of wind turbine faults. *Renewable Energy 36*(1), 16–23. (Cited on pages 17 and 74.)

Kusiak, A. and A. Verma (2012). Monitoring wind farms with performance curves. *IEEE Transactions on Sustainable Energy 4*(1), 192–199. (Cited on page 17.)

Kusiak, A., A. Verma, and X. Wei (2012). Wind turbine capacity frontier from scada. *Wind Syst. Mag 3*(9), 36–39. (Cited on page 98.)

Kusiak, A., H. Zheng, and Z. Song (2009). Models for monitoring wind farm power. *Renewable Energy 34*(3), 583–590. (Cited on pages 14, 17 and 24.)

Lange, M. and U. Focken (2006). *Physical Approach to Short-term Wind Power Prediction*, Volume 208. Springer. (Cited on pages 5 and 8.)

Lee, T.-S., C.-C. Chiu, Y.-C. Chou, and C.-J. Lu (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis 50*(4), 1113–1130. (Cited on page 8.)

Li, G. and J. Shi (2010). On comparing three artificial neural networks for wind speed forecasting. *Applied Energy 87*(7), 2313–2320. (Cited on page 5.)

Li, H., C. Yang, B. Zhao, H. Wang, and Z. Chen (2012). Aggregated models and transient performances of a mixed wind farm with different wind turbine generator systems. *Electric Power Systems Research 92*, 1–10. (Cited on page 27.)

Li, Q. and J. S. Racine (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press. (Cited on page 90.)

Li, S., D. C. Wunsch, E. A. O'Hair, and M. G. Giesselmann (2001). Using neural networks to estimate wind turbine power generation. *IEEE Transactions on Energy Conversion 16*(3), 276–282. (Cited on page 20.)

Li, W. and B. Bagen (2010). Reliability evaluation of integrated wind/diesel/storage systems for remote locations. In *IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 791–795. (Cited on page 72.)

Lignarolo, L., D. Ragni, C. Krishnaswami, Q. Chen, C. S. Ferreira, and G. Van Bussel (2014). Experimental analysis of the wake of a horizontal-axis wind-turbine model. *Renewable Energy 70*, 31–46. (Cited on page 30.)

Lindenberg, S., B. Smith, and K. O'Dell (2008). 20% wind energy by 2030. *National renewable energy laboratory (NREL), US department of energy, renewable energy consulting services, energetics incorporated*. (Cited on pages 42 and 106.)

Liu, F., J. Ma, W. Zhang, and M. Wu (2019). A comprehensive survey of accurate and efficient aggregation modeling for high penetration of large-scale wind farms in smart grid. *Applied Sciences 9*(4). (Cited on page 107.)

Liu, Y., X. Gao, J. Yan, S. Han, and D. G. Infield (2014). Clustering methods of wind turbines and its application in short-term wind power forecasts. *Journal of Renewable and Sustainable Energy 6*(5), 053119. (Cited on page 111.)

Lou, J., J. Xu, H. Lu, Z. Qu, S. Li, and R. Liu (2016). Wind turbine data-cleaning algorithm based on power curve. *Automation of Electric Power Systems 40*(10), 116–121. (Cited on page 24.)

Lydia, M., S. S. Kumar, A. I. Selvakumar, and G. E. P. Kumar (2014). A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews 30*, 452–460. (Cited on pages 4, 7, 72 and 74.)

Lydia, M., S. S. Kumar, A. I. Selvakumar, and G. E. P. Kumar (2015). Wind resource estimation using wind speed and power curve models. *Renewable Energy 83*, 425–434. (Cited on page 20.)

Lydia, M., A. Selvakumar, S. Kumar, and G. Kumar (2013). Advanced algorithms for wind turbine power curve modeling. *IEEE Transactions on Sustainable Energy 4*(3), 827–835. (Cited on pages 43, 49, 74 and 118.)

Mammen, E. (1991). Estimating a smooth monotone regression function. *The Annals of Statistics*, 724–740. (Cited on pages 54 and 55.)

Manobel, B., F. Sehnke, J. A. Lazzús, I. Salfate, M. Felder, and S. Montecinos (2018). Wind turbine power curve modeling based on gaussian processes and artificial neural networks. *Renewable Energy 125*, 1015–1020. (Cited on page 20.)

Manwell, J. F., J. G. McGowan, and A. L. Rogers (2010). *Wind Energy Explained: Theory, Design and Application*. John Wiley & Sons. (Cited on pages 15, 48, 49, 78 and 118.)

Marchand, É. and W. E. Strawderman (2020). On shrinkage estimation for balanced loss functions. *Journal of Multivariate Analysis 175*, 104558–104569. (Cited on page 76.)

Marčiukaitis, M., I. Žutautaitė, L. Martišauskas, B. Jokšas, G. Gecevičius, and A. Sfetsos (2017). Non-linear regression model for wind turbine power curve. *Renewable Energy 113*, 732–741. (Cited on page 43.)

Marinelli, E. M. (2011). Wind turbine and electrochemical based storage modeling and integrated control strategies to improve renewable energy integration in the grid. (Cited on pages 6, 10, 12 and 26.)

Marvuglia, A. and A. Messineo (2012). Monitoring of wind farms' power curves using machine learning techniques. *Applied Energy 98*, 574–583. (Cited on pages 8, 17, 21, 43, 64, 74 and 98.)

Mehrjoo, M., M. Jafari Jozani, and M. Pawlak (2019). Wind turbine power curve modeling for reliable power prediction using monotonic regression. *Renewable Energy 147*, 214–222. (Cited on pages 74 and 118.)

Mehrjoo, M., M. J. Jozani, and M. Pawlak (2020). Toward hybrid approaches for wind turbine power curve modeling with balanced loss functions and local weighting schemes. *Energy*, 119478. (Cited on page 107.)

Nabat, M. H., M. Zeynalian, A. R. Razmi, A. Arabkoohsar, and M. Soltani (2020). Energy, exergy, and economic analyses of an innovative energy storage system; liquid air energy storage (laes) combined with high-temperature thermal energy storage (htes). *Energy Conversion and Management 226*, 113486. (Cited on pages 4 and 73.)

Ni, Y., C. Li, Z. Du, and G. Zhang (2016). Model order reduction based dynamic equivalence of a wind farm. *International Journal of Electrical Power & Energy Systems 83*, 96–103. (Cited on page 107.)

Norgaard, P. and H. Holttinen (2004). A multi-turbine power curve approach. In *Nordic wind power conference*, Volume 1, pp. 1–2. Chalmers. (Cited on page 15.)

Oh, K.-Y., J.-Y. Kim, J.-K. Lee, M.-S. Ryu, and J.-S. Lee (2012). An assessment of wind energy potential at the demonstration offshore wind farm in korea. *Energy 46*(1), 555–563. (Cited on page 6.)

Olaofe, Z. O. and K. A. Folly (2013). Wind energy analysis based on turbine and developed site power curves: A case-study of darling city. *Renewable Energy 53*, 306–318. (Cited on page 15.)

Ouyang, T., A. Kusiak, and Y. He (2017). Modeling wind-turbine power curve: A data partitioning and mining approach. *Renewable Energy 102*, 1–8. (Cited on pages 43, 74 and 75.)

Pallabazzer, R. (2003). Parametric analysis of wind siting efficiency. *Journal of Wind Engineering and Industrial Aerodynamics 91*(11), 1329–1352. (Cited on page 16.)

Pandit, R., D. Infield, and A. Kolios (2019). Comparison of advanced nonparametric models for wind turbine power curves. *IET Renewable Power Generation 13*(9), 1503–1510. (Cited on pages 44 and 74.)

Pandit, R. K. (2018). *Gaussian process models for SCADA data based wind turbine performance/condition monitoring.* Ph. D. thesis, University of Strathclyde. (Cited on page 16.)

Pandit, R. K. and D. Infield (2018). Scada-based wind turbine anomaly detection using gaussian process models for wind turbine condition monitoring purposes. *IET Renewable Power Generation 12*(11), 1249–1255. (Cited on pages 7 and 17.)

Park, J.-Y., J.-K. Lee, K.-Y. Oh, and J.-S. Lee (2014). Development of a novel power curve monitoring method for wind turbines and its field tests. *IEEE Transactions on Energy Conversion 29*(1), 119–128. (Cited on page 73.)

Park, R. E. (1966). Estimation with heteroscedastic error terms. *Econometrica (pre-1986) 34*(4), 888. (Cited on page 25.)

Pelletier, F., C. Masson, and A. Tahan (2016). Wind turbine power curve modelling using artificial neural network. *Renewable Energy 89*, 207–214. (Cited on pages 20 and 43.)

Perdana, A., S. Uski-Joutsenvuo, O. Carlson, and B. Lemström (2008). Comparison of an aggregated model of a wind farm consisting of fixed-speed wind turbines with field measurement. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology 11*(1), 13–27. (Cited on page 107.)

Petković, D., Ž. Ćojbašić, V. Nikolić, S. Shamshirband, M. L. M. Kiah, N. B. Anuar, and A. W. A. Wahab (2014). Adaptive neuro-fuzzy maximal power extraction of wind turbine with continuously variable transmission. *Energy 64*, 868–874. (Cited on page 29.)

Raj, M. M., M. Alexander, and M. Lydia (2011). Modeling of wind turbine power curve. In *ISGT2011-India*, pp. 144–148. IEEE. (Cited on page 73.)

Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association 65*(329), 161–172. (Cited on page 25.)

Rao, C. R. and H. Toutenburg (1995). Linear models. In *Linear Models*, pp. 3–18. Springer. (Cited on pages 125 and 133.)

Rašuo, B. P. and A. Č. Bengin (2010). Optimization of wind farm layout. *FME Transactions 38*(3), 107–114. (Cited on page 29.)

Rogers, T., P. Gardner, N. Dervilis, K. Worden, A. Maguire, E. Papatheou, and E. Cross (2020). Probabilistic modelling of wind turbine power curves with application of heteroscedastic gaussian process regression. *Renewable Energy 148*, 1124–1136. (Cited on page 25.)

Rousseeuw, P. J. and A. M. Leroy (2005). *Robust Regression and Outlier Detection*, Volume 589. John wiley & sons. (Cited on page 23.)

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Number 12. Cambridge university press. (Cited on pages 87 and 127.)

Saavedra-Moreno, B., S. Salcedo-Sanz, A. Paniagua-Tineo, L. Prieto, and A. Portilla-Figueras (2011). Seeding evolutionary algorithms with heuristics for optimal wind turbines positioning in wind farms. *Renewable Energy 36*(11), 2838–2844. (Cited on page 29.)

Saint-Drenan, Y.-M., R. Besseau, M. Jansen, I. Staffell, A. Troccoli, L. Dubus, J. Schmidt, K. Gruber, S. G. Simões, and S. Heier (2020). A parametric model for wind turbine power curves incorporating environmental conditions. *Renewable Energy*. (Cited on page 74.)

Shamshirband, S., D. Petković, N. B. Anuar, and A. Gani (2014). Adaptive neuro-fuzzy generalization of wind turbine wake added turbulence models. *Renewable and Sustainable Energy Reviews 36*, 270–276. (Cited on page 28.)

Sharpe, D., T. Burton, N. Jenkins, and E. Bossanyi (2013). *Wind Energy Handbook*. Wiley. (Cited on page 44.)

Shen, X., X. Fu, and C. Zhou (2018). A combined algorithm for cleaning abnormal data of wind turbine power curve based on change point grouping algorithm and quartile algorithm. *IEEE Transactions on Sustainable Energy 10*(1), 46–54. (Cited on pages 23 and 24.)

Shokrzadeh, S. (2014). Battery repurposing of plug-in electric vehicles: a framework for the integration of renewable energy and electrified transportation. (Cited on pages 6 and 18.)

Shokrzadeh, S. and E. Bibeau (2012). Repurposing batteries of plug-in electric vehicles to support renewable energy penetration in the electric grid. Technical report, SAE Technical Paper. (Cited on page 18.)

Shokrzadeh, S., M. Jafari Jozani, and E. Bibeau (2014). Wind turbine power curve modeling using advanced parametric and nonparametric methods. *IEEE Transactions on Sustainable Energy 5*(4), 1262–1269. (Cited on pages 20, 43, 46, 50, 64, 74, 97, 98, 118 and 119.)

Shokrzadeh, S., M. J. Jozani, E. Bibeau, and T. Molinski (2015). A statistical algorithm for predicting the energy storage capacity for baseload wind power generation in the future electric grids. *Energy 89*, 793–802. (Cited on pages 14 and 73.)

Simic, Z. and V. Mikulicic (2007). Small wind off-grid system optimization regarding wind turbine power curve. In *AFRICON 2007*, pp. 1–6. IEEE. (Cited on page 16.)

Sohoni, V., S. Gupta, and R. Nema (2016a). A comparative analysis of wind speed probability distributions for wind power assessment of four sites. *Turkish Journal of Electrical Engineering & Computer Sciences 24*(6), 4724–4735. (Cited on page 20.)

Sohoni, V., S. Gupta, and R. Nema (2016b). A critical review on wind turbine power curve modelling techniques and their applications in wind based energy systems. *Journal of Energy*. (Cited on pages 74 and 75.)

Staffell, I. and R. Green (2014). How does wind farm performance decline with age? *Renewable Energy 66*, 775–786. (Cited on page 73.)

Subramanian, B., N. Chokani, and R. S. Abhari (2016). Aerodynamics of wind turbine wakes in flat and complex terrains. *Renewable Energy 85*, 454–463. (Cited on page 30.)

Swapna, S., P. Niranjan, B. Srinivas, and R. Swapna (2016). Data cleaning for data quality. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 344–348. IEEE. (Cited on page 23.)

Taieb, S. B., G. Bontempi, A. F. Atiya, and A. Sorjamaa (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. *Expert Systems with Applications 39*(8), 7067–7083. (Cited on page 8.)

Taslimi-Renani, E., M. Modiri-Delshad, M. F. M. Elias, and N. A. Rahim (2016). Development of an enhanced parametric model for wind turbine power curve. *Applied energy 177*, 544–552. (Cited on page 75.)

Thapar, V., G. Agnihotri, and V. K. Sethi (2011). Critical analysis of methods for mathematical modelling of wind turbines. *Renewable Energy 36*(11), 3166–3177. (Cited on page 20.)

Tibshirani, R., G. Walther, and T. Hastie (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(2), 411–423. (Cited on page 115.)

Tong, W. (2010). *Wind Power Generation and Wind Turbine Design*. WIT press. (Cited on page 5.)

Trivellato, F., L. Battisti, and G. Miori (2012). The ideal power curve of small wind turbines from field data. *Journal of Wind Engineering and Industrial Aerodynamics 107*, 263–273. (Cited on page 14.)

Ulazia, A., J. Sáenz, G. Ibarra-Berastegi, S. J. González-Rojí, and S. Carreno-Madinabeitia (2019). Global estimations of wind energy potential considering seasonal air density changes. *Energy 187*, 115938. (Cited on page 72.)

Uyanto, S. S. (2019). Monte carlo power comparison of seven most commonly used heteroscedasticity tests. *Communications in Statistics-Simulation and Computation*, 1–18. (Cited on page 89.)

Villanueva, D. and A. Feijóo (2016a). Normal-based model for true power curves of wind turbines. *IEEE Transactions on Sustainable Energy 7*(3), 1005–1011. (Cited on page 74.)

Villanueva, D. and A. Feijoo (2018). Comparison of logistic functions for modeling wind turbine power curves. *Electric Power Systems Research 155*, 281–288. (Cited on page 43.)

Villanueva, D. and A. E. Feijóo (2016b). Reformulation of parameters of the logistic function applied to power curves of wind turbines. *Electric Power Systems Research 137*, 51–58. (Cited on page 20.)

Walker, J. F. and N. Jenkins (1997). *Wind Energy Technology*. John Wiley & Sons Incorporated. (Cited on pages 11 and 78.)

Wand, M. P. and M. C. Jones (1994a). *Kernel Smoothing*. Chapman and Hall/CRC. (Cited on pages 52 and 53.)

Wand, M. P. and M. C. Jones (1994b). *Kernel Smoothing*. Crc Press. (Cited on page 91.)

Wang, Y., Q. Hu, L. Li, A. M. Foley, and D. Srinivasan (2019). Approaches to wind power curve modeling: A review and discussion. *Renewable and Sustainable Energy Reviews 116*, 109422. (Cited on page 4.)

Wang, Y., Q. Hu, D. Srinivasan, and Z. Wang (2018). Wind power curve modeling and wind power forecasting with inconsistent data. *IEEE Transactions on Sustainable Energy 10*(1), 16–25. (Cited on pages 25 and 75.)

Wang, Y., Y. Li, Z. Runmin, A. M. Foley, D. Alkez, D. Song, Q. Hu, and D. Srinivasan (2020). Sparse heteroscedastic multiple spline regression models for wind turbine power curve modeling. *IEEE Transactions on Sustainable Energy*. (Cited on page 75.)

Wang, Y., Y. Yu, S. Cao, X. Zhang, and S. Gao (2020). A review of applications of artificial intelligent algorithms in wind farms. *Artificial Intelligence Review 53*(5), 3447–3500. (Cited on page 4.)

Wasserman, L. (2006). *All of nonparametric statistics.* Springer Science & Business Media. (Cited on page 91.)

Yesilbudak, M., S. Sagiroglu, and I. Colak (2013). A new approach to very short term wind speed prediction using k-nearest neighbor classification. *Energy Conversion and Management 69*, 77–86. (Cited on page 8.)

Yin, P.-Y. and T.-Y. Wang (2012). A grasp-vns algorithm for optimal wind-turbine placement in wind farms. *Renewable Energy 48*, 489–498. (Cited on page 29.)

Yongqian, L., P. Jinji, and H. Shuang (2011). Study on two neural network algorithms to predict wind power [j]. *Modern Electric Power 2*. (Cited on page 27.)

Zellner, A. (1994). Bayesian and non-bayesian estimation using balanced loss functions. In *Statistical Decision Theory and Related Topics V*, pp. 377–390. Springer. (Cited on page 76.)

Zha, X., S. Liao, M. Huang, Z. Yang, and J. Sun (2019). Dynamic aggregation modeling of grid-connected inverters using hamilton's-action-based coherent equivalence. *IEEE Transactions on Industrial Electronics 66*(8), 6437–6448. (Cited on page 107.)

Zhang, B. and J. Liu (2019). Wind turbine clustering algorithm of large offshore wind farms considering wake effects. *Mathematical Problems in Engineering 2019*. (Cited on pages 28 and 108.)

Zhang, Q. and P. Chen (2018). Credibility estimators with dependence structure over risks and time under balanced loss function. *Statistica Neerlandica 72*(2), 157–173. (Cited on page 76.)

Zhang, Z., Q. Zhou, and A. Kusiak (2014). Optimization of wind power and its variability with a computational intelligence approach. *IEEE Transactions on Sustainable Energy 5*(1), 228–236. (Cited on page 43.)

Zhao, Y., L. Ye, W. Wang, H. Sun, Y. Ju, and Y. Tang (2017). Data-driven correction approach to refine power curve of wind farm under wind curtailment. *IEEE Transactions on Sustainable Energy 9*(1), 95–105. (Cited on pages 23 and 24.)

Zhao, Y., L. Ye, and Q. Zhu (2014). Characteristics and processing method of abnormal data cluster caused by wind curtailments in wind farms. *Automation of Electric Power Systems 21*, 39–46. (Cited on page 24.)

Zheng, L., W. Hu, and Y. Min (2014). Raw wind data preprocessing: a data-mining approach. *IEEE Transactions on Sustainable Energy 6*(1), 11–19. (Cited on page 24.)

Zhou, Y., L. Zhao, I. B. Matsuo, and W.-J. Lee (2019). A dynamic weighted aggregation equivalent modeling approach for the dfig wind farm considering the weibull distribution for fault analysis. *IEEE Transactions on Industry Applications 55*(6), 5514–5523. (Cited on page 4.)

Zou, J., C. Peng, H. Xu, and Y. Yan (2015). A fuzzy clustering algorithm-based dynamic equivalent modeling method for wind farm with dfig. *IEEE Transactions on Energy Conversion 30*(4), 1329–1337. (Cited on page 28.)