DATA ANALYTICS METHODS IN WIND TURBINE DESIGN AND OPERATIONS

A Dissertation

by

GIWHYUN LEE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Yu Ding
Committee Members,	Lewis Ntaimo
	Marc G. Genton
	Chanan Singh
Head of Department,	César O. Malavé

August 2013

Major Subject: Industrial Engineering

ABSTRACT

This dissertation develops sophisticated data analytic methods to analyze structural loads on, and power generation of, wind turbines. Wind turbines, which convert the kinetic energy in wind into electrical power, are operated within stochastic environments. To account for the influence of environmental factors, we employ a conditional approach by modeling the expectation or distribution of response of interest, be it the structural load or power output, conditional on a set of environmental factors. Because of the different nature associated with the two types of responses, our methods also come in different forms, conducted through two studies.

The first study presents a Bayesian parametric model for the purpose of estimating the extreme load on a wind turbine. The extreme load is the highest stress level that the turbine structure would experience during its service lifetime. A wind turbine should be designed to resist such a high load to avoid catastrophic structural failures. To assess the extreme load, turbine structural responses are evaluated by conducting field measurement campaigns or performing aeroelastic simulation studies. In general, data obtained in either case are not sufficient to represent various loading responses under all possible weather conditions. An appropriate extrapolation is necessary to characterize the structural loads in a turbine's service life. This study devises a Bayesian spline method for this extrapolation purpose and applies the method to three sets of load response data to estimate the corresponding extreme loads at the roots of the turbine blades.

In the second study, we propose an additive multivariate kernel method as a new power curve model, which is able to incorporate a variety of environmental factors in addition to merely the wind speed. In the wind industry, a power curve refers to the functional relationship between the power output generated by a wind turbine and

the wind speed at the time of power generation. Power curves are used in practice for a number of important tasks including predicting wind power production and assessing a turbine's energy production efficiency. Nevertheless, actual wind power data indicate that the power output is affected by more than just wind speed. Several other environmental factors, such as wind direction, air density, humidity, turbulence intensity, and wind shears, have potential impact. Yet, in industry practice, as well as in the literature, current power curve models primarily consider wind speed and, with comparatively less frequency, wind speed and direction. Our model provides, conditional on a given environmental condition, both the point estimation and density estimation of the power output. It is able to capture the nonlinear relationships between environmental factors and wind power output, as well as the high-order interaction effects among some of the environmental factors. To illustrate the application of the new power curve model, we conduct case studies that demonstrate how the new method can help with quantifying the benefit of vortex generator installation, advising pitch control adjustment, and facilitating the diagnosis of faults.

ACKNOWLEDGMENTS

Foremost, it is with immense gratitude that I acknowledge the help of my advisor, Prof. Yu Ding, for his continuous support of my Ph.D. study and research, for his guidance, motivation, enthusiasm, extensive knowledge, and most importantly, his kind amiability throughout my years of study. The completion of this study would not have been possible without Dr. Ding's thoughtful advice and supportive mentorship. One simply could not wish for a better or friendlier supervisor.

Besides my advisor, I owe my deepest gratitude to the rest of my thesis committee members: Dr. Lewis Ntaimo, Dr. Marc G. Genton, and Dr. Chanan Singh for their constant encouragement, insightful comments, and thought-provoking questions. This thesis would have remained a dream had it not been for them. My sincere thanks also goes to Dr. Le Xie for his helpful discussions at various times.

In my daily work I have been blessed with friendly and cheerful labmates: Eunshin Byon, Chiwoo Park, Arash Pourhabib, Shilan Jin, Anna Kotyk, Yanjun Qian, and Hoon Hwangbo, for the most stimulating discussions that led to insights which further enlightened and challenged my study, and resulted in substantially improving my research product. I also consider it an honor to have worked with Dr. Byon, who provided me with not only helpful discussions, but also many inspirations for my research.

Last but not the least, my deepest appreciation goes to my family. Especially, I owe my sincerest gratitude to my husband, Yongjoo Lee, for his endless support, love, and patience. For about four years, I remember his constant support when encountering difficulties, and he has always been by my side not only as a husband, but as a friend, consultant, and adviser throughout my study. My precious child, Geon, has been a strong force during my doctoral studies and given me momentum

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xii

CHAPTER I

INTRODUCTION

I.1. Motivation

Wind energy is one of the fastest growing renewable energy sources. According to a report issued by American Wind Energy Association (AWEA), wind power installations in the U.S. increased to 46,919 MW by the end of 2011 (AWEA, 2012). The U.S. Department of Energy advocates working toward the goal for wind power to account for 20% of the U.S. total electricity generation by 2030 (DOE, 2008) as shown in Figure 1.
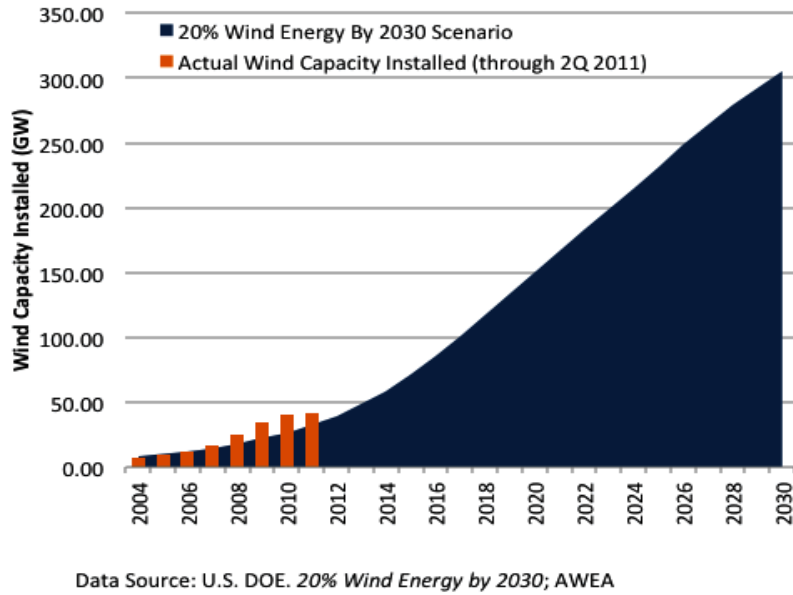
Fig. 1. Graphic illustration of U.S. DOE 20% Wind Energy by 2030 [Source: DOE (2008)]

Wind energy has numerous advantages compared to conventional sources in terms of environmental and economic benefits, such as reduced greenhouse gas emissions, and free fuel. Despite the advantages, a number of challenges arise in the development of wind energy, including:

- Operational costs of electricity systems: Efficient integration of wind energy into electric power system is an important issue in the wind industry. For example, to balance fluctuations and uncertainties in wind output, system operators must dynamically schedule other types of generation as operational reserves. Consequently, the variability and uncertainty in wind power generation leads to higher costs in energy scheduling and dispatching.

- Operation and maintenance (O&M) costs: A key factor for enhancing the marketability of wind energy is to reduce O&M costs. According to Walford (2006), O&M costs can account for 10 - 20% of the total energy production costs. As a wind turbine comes to the end of its service life, it is expected that failure rates or O&M costs will increase drastically.

- Wind turbine system reliability: Poor reliability of wind turbines leads directly to reduced wind power generation as well as increased O&M costs. Catastrophic failures due to careless design parameter selections can result in immense economic loss. Thus, manufacturers are interested in assessing extreme wind turbine loads for more reliable designs.

Therefore, studies concerning structural reliability analysis, wind power forecasting, and performance assessment play a critical role in addressing the aforementioned challenges.

Wind turbines convert the kinetic energy in the wind into electrical power, so wind speed is the most important factor influencing the power generation of a wind

turbine. At the same time, it is also the most essential element in quantifying structural loads on a wind turbine. Hence, when evaluating structural loads and wind power generation, wind speed must always be measured. In order to better understand the relationship between wind speed and wind turbine operation, we present two scatter plots obtained from an observed dataset. Figures 2(a) and (b) show that both loads and power outputs have nonlinear relationships with wind speed. We also observe the presence of a great amount of uncertainty in both relationships. Although the shape or pattern of these curves may change depending on the size or type of the wind turbines, the nonlinearity in the relationship with wind speed and the presence of uncertainty are common characteristics in wind energy research. In addition, a significant portion of uncertainty arises from other environmental factors such as wind direction, turbulence intensity, air density, wind shear, and the like. Because of this uncertainty, we are interested in the conditional distribution of y given $\boldsymbol{x}$, $p(y|\boldsymbol{x})$ or the conditional expectation of y given $\boldsymbol{x}$, $E(y|\boldsymbol{x})$. Here, y represents a dependent variable of interest such as structural loads or power outputs and $\boldsymbol{x}$ refers to one or more of the previously mentioned environmental factors.

I.2. Overview of wind turbine design and operations

In this section, we introduce some basic concepts and terminologies related to wind turbine design and operations. Then, the research problems are identified and defined.

I.2.1. Extreme load estimation

A wind turbine operates under various loading conditions in stochastic weather environments. The increase in size, weight, and length of components of utility-scale wind turbines escalates the stresses/loads/responses imposed on the structure. As a
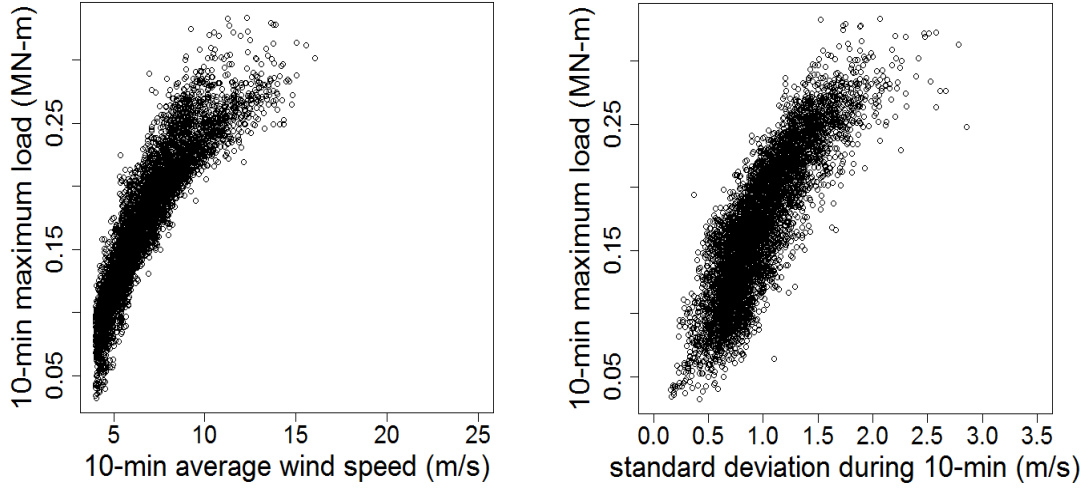
(a) Structural load and wind speed   (b) Power output and wind speed

Fig. 2. Scatter plots of the response variables versus average wind speed: structural load and power output

result, modern wind turbines are prone to structural failures. Of particular interest in a wind turbine system are the extreme events under which loads exceed a threshold called a "nominal design load" or "extreme load." Upon the occurrence of a load higher than the nominal design load, a wind turbine could experience catastrophic structural failures.

Mathematically, an extreme load is defined as an extreme quantile value in a load distribution corresponding to a turbine's service time of T years (Sørensen and Nielsen, 2007). Let y denote the maximum load, in the unit of mill-Newton-meters (MN-m), during a specific time interval. Then, we define the load exceedance probability as follows:

$$P_T = P[y > l_T], \tag{1.1}$$

4

where P_T is the target probability of exceeding the load level l_T (in the same units as that of y).

In structural reliability analysis of wind turbines, people collect load response data and arrange them in 10-minute intervals because wind speeds are considered stationary over a 10-minute duration (Fitzwater and Winterstein, 2001). Given this data arrangement in the wind industry, y commonly denotes the maximum load during a 10-minute interval. The unconditional distribution of y, $p(y)$, is called the *long-term* distribution and is used to calculate $P[y > l_T]$ in (1.1).

In (1.1), the extreme event, $\{y > l_T\}$, takes place with the exceedance probability P_T. During the service time of a wind turbine, it is considered ideal to prevent the occurrence of such an extreme event. Differently stated, the optimal amount of waiting time before the occurrence of an extreme event should be longer than, or equal to, the service time. Therefore, a reasonable estimation of P_T can be done in the following way (IEC, 2005a, Peeringa, 2003):

$$P_T = \frac{10 \text{ min}}{T \text{ years} \times 365.25 \text{ days} \times 24 \text{ hours} \times 60 \text{ min}}. \tag{1.2}$$

Note that P_T is the reciprocal of the number of 10-minute intervals in T years. For example, when T is 50 years, P_T becomes 3.8×10^{-7}.

Estimating the extreme load implies finding an extreme quantile l_T in the 10-minute maximum load distribution, given a target service period T, such that (1.1) is satisfied. Wind turbines should be designed to resist the l_T load level to avoid structural failures during its desired service life.

Since loads are highly affected by wind profiles, we consider the marginal distribution of y obtained by using the distribution of y conditional on a wind profile as

follows:

$$p(y) \;=\; \int p(y|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}. \tag{1.3}$$

Here, $p(\boldsymbol{x})$ is the joint probability density function of wind characteristics in vector $\boldsymbol{x}$. The conditional distribution of y given $\boldsymbol{x}$, $p(y|\boldsymbol{x})$ in (1.3), is called the *short-term* distribution. The long-term distribution can be computed by integrating out wind characteristics in the short-term distribution.

For inland turbines, the wind characteristic vector $\boldsymbol{x}$ in general is comprised of two elements: (1) a steady state mean of wind speed and (2) the stochastic variability of wind speed (Bottasso *et al.*, 2010, Manuel *et al.*, 2001, Ronold and Larsen, 2000). The first element can be measured by the average wind speed (in the unit of meters per second, or m/s) during a 10-minute interval, and the second element can be represented by the standard deviation of wind speed, or the turbulence intensity, also during a 10-minute interval. Here, turbulence intensity is defined as the standard deviation of wind speed divided by the average wind speed for the same duration. For offshore turbines, weather characteristics other than wind may be needed, such as the wave height (Agarwal and Manuel, 2008).

I.2.2. Power curve estimation and application

To manage wind turbines and to plan wind energy production, it is critical to assess wind power generation under a given weather profile. The so-called power curve plays a central role in this task (Giebel *et al.*, 2011, Monteiro *et al.*, 2009). In the wind industry, the power curve measures the relationship between the power output of a turbine and the wind speed.

We first explain the basics of the power curve. Denote by y the power output from a wind turbine and by $\boldsymbol{x}$ the vector of explanatory variables. v is the wind speed.

In wind power production, as illustrated in the left panel of Figure 3, a turbine starts to produce power after the wind reaches the cut-in speed, v_{ci}. A nonlinear relationship between y and v then ensues, until the wind reaches the rated wind speed, v_r. When the wind speed is beyond v_r, the turbine's output power will be restricted at the rated power output, y_r, also known as the nominal power capacity of the turbine, using control mechanisms such as pitch control and rotor speed regulation. The turbine will be halted when the wind reaches the cut-out speed, v_{co}, because high wind is deemed harmful to the safety of a turbine. For the power curve shown in the left panel of Figure 3, $\boldsymbol{x} := (v)$.
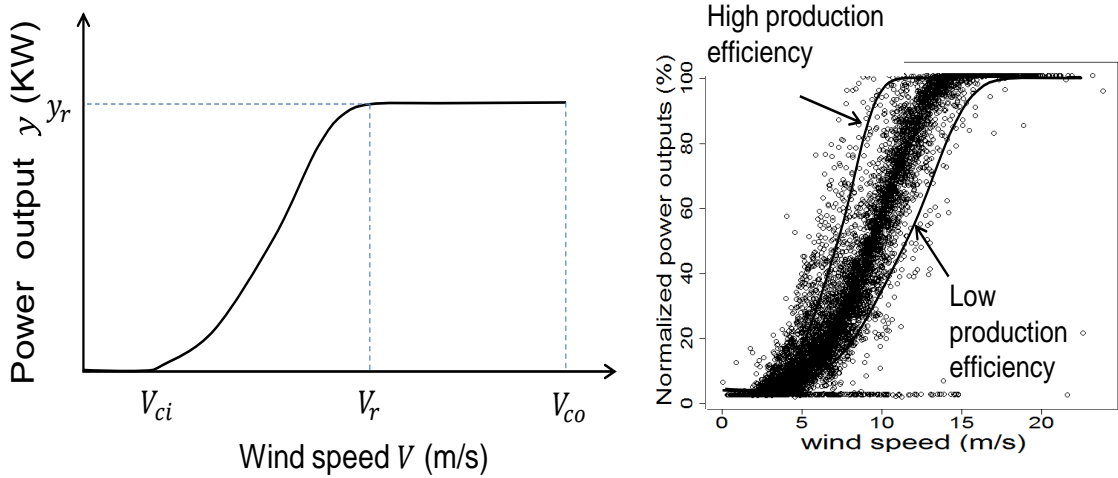
Fig. 3. An example of a power curve: V_{ci} is the cut-in wind speed, V_{co} is the cut-out wind speed, V_r is the rated wind speed, and y_r is the corresponding rated power output. In the right panel where real power production data are shown, the power outputs are normalized by the rated power output, to protect the identity of the turbine manufacturer. The same treatment is applied to all power curve plots throughout the paper.

The curve shown in the left panel of Figure 3 is an ideal power curve, also known

as the nominal power curve, typically provided by a turbine manufacturer. The right panel of Figure 3 shows the actual power output and wind measurement data associated with a turbine, which presents a more complicated picture. Even though the general trend shown in the data tends to agree with the nominal power curve, there appears to be a considerable amount of information that cannot be accounted for by a simple v-versus-y curve. Between 5 meters per second (m/s) and 15 m/s, there are large amounts of power data for any given wind speed. This implies is that if v is used as the sole explanatory variable, the prediction of wind power suffers from a high degree of uncertainty. We investigate the possibility and the means of including more explanatory variables to present a better fit to the power data.

In fact, the meteorological mast on each wind farm included a wide array of sensors that measure more than just wind speed. Other environmental variables measured include wind direction, d, temperature, t_m, air pressure, a_p, and humidity, h_m. Based on the wind speed measurements, it is also possible to calculate turbulence intensity, t_b (equal to the standard deviation of short-duration wind speeds divided by the average wind speed of the same duration) and wind shear, w_s (using wind speeds measured at different heights). If we expand our input variable set to include these environmental factors, we could have $\boldsymbol{x} := (v, d, t_m, a_p, h_m, t_b, w_s)$. Then, our technical objective would be to estimate the conditional density, $p(y|\boldsymbol{x})$, or the conditional expectation, $E(y|\boldsymbol{x})$. Technically, $E(y|\boldsymbol{x})$ is no longer a power curve when $\boldsymbol{x}$ includes multiple elements; it becomes a power response surface. For the sake of being consistent with industrial convention, we use the term "power curve" in its broad definition, covering the cases of both one-dimensional power curves and multi-dimensional power response surfaces.

The wind industry makes use of power curves for at least two important purposes. The first is to forecast wind power, which requires two steps (Giebel *et al.*, 2011,

Monteiro *et al.*, 2009). First, wind speeds are forecast and then this forecast is converted to a power forecast using a power curve. We note that wind forecasting is beyond the focus of this study, given that it is a subfield of its own. For details on wind forecasting, interested readers should refer to Gneiting *et al.* (2006), Monteiro *et al.* (2009), Hering and Genton (2010), and Zhu and Genton (2012). The second purpose of power curves is for turbine performance assessment and turbine health monitoring (Albers *et al.*, 1999, Stephen *et al.*, 2011, Uluyol *et al.*, 2011), in which a power curve is used to characterize a turbine's power production efficiency by noting the changes in the position and slope of the turbine's power curve. For an illustration, see Figure 3, right panel. Under the assumption that future wind speeds are given, the procedure of wind power forecasting is very simple and straightforward. On the other hand, for the latter application, we should address an additional issue related to test procedures.

For performance assessment, the general approach is to use the so-called endogeneous power output excluding environmental effects, also known as the *power residual* and denoted by r. It can be expressed as in the following:

$$r = y - \hat{y}, \tag{1.4}$$

where y is the observed power output, and $\hat{y}$ is the power output estimated by a power curve model. Given this difference, the next step is to distinguish between the normal and abnormal behavior of a turbine by making use of the power residual. Here, abnormal behaviors can be caused by a variety of physical changes in wind turbines including intentional retrofit, natural degradation, or system faults. It is important to study effective methods that can detect the occurrence of anomalies.

I.3.   Research objective and outline

Due to uncertainty and stochastic issues, the wind industry has been greatly interested in explaining the relationship between environmental factors $\boldsymbol{x}$ and maximum load or power output y. However, it is generally difficult to precisely characterize the relationship between environmental factors and structural loads or power outputs by using physical law-based dynamic equations. For example, a power curve provided by manufacturers is often different from the observed curves. Therefore, we are interested in understanding and explaining these relationships by using a data-driven statistical approach with field data.

For addressing uncertainty in observed data, the International Electro-technical Commission (IEC) recommends the assessment of structural loads (extreme load levels) or power outputs with reference to site-specific conditions (IEC, 2005a,b). The basic idea of the recommended method, called the *binning* method, is to discretize the domain of a few environmental factors and to estimate within each bin. The detailed explanation can be found in Chapter II. The current industrial practices follow the IEC recommendation because of the simplicity of its idea and procedure. However, the current method has several serious shortcomings. It often results in low accuracy and high uncertainty because most bins have limited data, or sometimes no data at all. Also, it is not easy to extend the binning method to include more environmental factors. Thus, the objective of this dissertation research is to develop and apply new conditional models for estimating extreme load levels and power curves under stochastic weather conditions.

I.3.1. Bayesian spline method for assessing extreme loads on wind turbines

In this study, we present a new extreme load analysis model estimating the site-specific structural design load (extreme load levels). This model allows us to calculate the structural design load for the same kind of wind turbine that will be built at a different site. Our new procedure consists of two sub-models, which are the short-term distribution, $p(y|\boldsymbol{x})$ and the distribution of environmental variables, $p(\boldsymbol{x})$, respectively. The novelty of the new procedure primarily concerns the ways of modelling the short-term distribution $p(y|\boldsymbol{x})$. Particularly, we establish a parametric load distribution for $y|\boldsymbol{x}$ using spline models. As such, we label the resulting method *a Bayesian spline method for extreme loads.*

I.3.2. Power curve estimation and turbine performance assessment with multivariate environmental factors

In this study, we estimate the power curve associated with individual turbines at both inland and offshore wind farms using turbine-specific power output data and environmental data measured from a meteorological mast on the corresponding farm. Our research shows that inclusion of the extra environmental factors in a power curve model can indeed improve wind power predictions. A power curve model that incorporates multiple environmental factors also provides a useful tool for studying the relative importance of these environmental variables on wind power generation. To fulfill the objective of developing a power curve method with multivariate dependencies, we expanded the kernel-based methodology developed by Jeon and Taylor (2012). The main challenge to overcome was to find a novel way of handling the extra variables in $\boldsymbol{x}$, because a typical kernel density estimation method easily runs

into computational problems when the number of explanatory variables exceeds two or three. Our approach is to devise an additive multivariate kernel model that can handle the multi-dimensional power curve estimation problem in a reasonable computational time.

To demonstrate how the resulting new power curve model can facilitate the mission of turbine performance assessment, we present three case studies in which wind turbines are either upgraded or had experienced faults. To detect a turbine change resulting from either upgrades or faults, we incorporate the power curve models and suggest a new procedure for performance testing.

I.4. Organization of this dissertation

This dissertation is organized as follows. Chapter II surveys the methods available in the literature on wind turbine design and on operations. In addition, this chapter includes an explanation of the current models used in the wind industry. Then, the limitation of the existing approaches for wind turbine design and operations is discussed and arguments for the necessity of developing new models are presented.

Chapter III presents a Bayesian spline method for estimating the extreme load level l_T, using load data collected in a period much shorter than a turbine's service life. The spline method is applied to three sets of a turbine's load response data to estimate the corresponding extreme loads at the roots of the turbine blades. Compared to the current industry practice or the unconditional existing method, the spline method appears to provide better extreme load assessment.

Chapter IV presents an additive multivariate kernel method that can include a variety of environmental factors as a new power curve model. Our model provides, conditional on a given environmental condition, both the point estimation and density

estimation of the power output. It is able to capture the nonlinear relationships between environmental factors and wind power output, as well as the high-order interaction effects among some of the environmental factors. Using operational data associated with four turbines on an inland wind farm and two turbines on an offshore wind farm, we demonstrate the improvement achieved by our kernel method.

Chapter V presents a new procedure for performance testing or monitoring by using the new power curve model presented in Chapter IV. Then, we compare the performance assessment results provided by both the additive multivariate kernel method and the binning method. The new procedure is illustrated by its application to three operational datasets of wind turbines that underwent different physical changes.

Chapter VI summarizes the contribution of this dissertation with its broad impact on wind turbine design and operations. The chapter also discusses future research directions and possible extensions to this dissertation study.

CHAPTER II

LITERATURE REVIEW

This chapter reviews the current practice and existing methods related to extreme load analysis and power generation analysis. The first section reviews deterministic and statistical approaches for extreme load assessment, and discusses the limitations of existing methods. The second section explains the current industrial practices for power generation forecasting or turbine assessment.

II.1. Studies on extreme load analysis

II.1.1. Deterministic approaches for extreme load analysis

The previous edition of the international standard, IEC 61400-1:1999, offers a set of design load cases with *deterministic* wind conditions such as annual average wind speeds, higher and lower turbulence intensities, and extreme wind speeds (IEC, 1999). In other words, the loads in the IEC 61400-1:1999 are specified as discrete events based on design experiences and empirical models (Moriarty *et al.*, 2002). Veers and Butter-field (2001) point out that these deterministic models do not represent the stochastic nature of structure responses, and suggest using statistical modeling to improve design load estimates. Moriarty *et al.* (2002) examine the effect of varying turbulence levels on the statistical behavior of a wind turbine's extreme load. They conclude that the loading on a turbine is stochastic at high turbulence levels, significantly influencing the tail of the load distribution.

In response to these developments, the new edition of IEC 61400-1 standard (IEC 61400-1:2005), issued in 2005, replaces the deterministic load cases with *stochastic* models, and recommends the use of *statistical* approaches for determining the ex-

treme load level in the design stage. Freudenreich and Argyriadis (2008) compare the deterministic load cases in the IEC 61400-1:1999 with the stochastic cases in IEC 61400-1:2005, and observe that when statistical approaches are applied, higher extreme load estimates are obtained in some structural responses, such as the blade tip deflection and flap-wise bending moment.

II.1.2. Statistical approaches for extreme load analysis

After IEC 61400-1:2005 was issued, many studies reported focusing on devising and recommending statistical approaches for extreme load analysis (Agarwal and Manuel, 2008, Fogle *et al.*, 2008, Freudenreich and Argyriadis, 2008, Moriarty, 2008, Natarajan and Holley, 2008, Peeringa, 2009, Regan and Manuel, 2008). These studies adopt a common framework, which we refer to as the binning method. The basic idea of the binning method is to discretize the domain of a wind profile vector $\boldsymbol{x}$ into a finite number of bins. For example, one can divide the range of wind speeds, from the *cut-in* speed to the *cut-out* speed, into multiple bins and set the width of each bin to, say, 2 m/s. Then, in each bin, the conditional short-term distribution of $y|\boldsymbol{x}$ is approximated by a stationary distribution, with the parameters of the distribution estimated by the method of moments or the maximum likelihood method. Then, the contribution from each bin is summed over all possible bins to determine the final long-term extreme load. In other words, integration in (1.3) for calculating the long-term distribution is approximated by the summation of finite elements as follows:

$$1 - P_T \quad \approx \quad \int_{-\infty}^{l_T} \sum_{\boldsymbol{x}} p(y|\boldsymbol{x})p(\boldsymbol{x})\Delta\boldsymbol{x}dy. \tag{2.1}$$

According to the classical extreme value theory (Coles, 2001, Smith, 1990), the short-term distribution of $y|\boldsymbol{x}$ can be approximated by a generalized extreme value

(GEV) distribution. The probability density function of the GEV is

$$
\begin{aligned}
p(y) &= \frac{1}{\sigma}\exp\left[-\left(1+\xi\left(\frac{y-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right]\left(1+\xi\left(\frac{y-\mu}{\sigma}\right)\right)^{-1-\frac{1}{\xi}} \quad \text{if } \xi \neq 0, \\
&= \frac{1}{\sigma}\exp\left[-\frac{y-\mu}{\sigma}-\exp\left(-\frac{y-\mu}{\sigma}\right)\right] \qquad\qquad \text{if } \xi = 0. \quad (2.2)
\end{aligned}
$$

for $\{y : 1 + \xi(y-\mu)/\sigma > 0\}$, where, $\mu \in \Re$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\xi \in \Re$ is the shape parameter that determines the weight of the tail of $p(y)$. $\xi > 0$ corresponds to the Fréchet distribution with a heavy upper tail, $\xi < 0$ to the Weibull distribution with a short upper tail and light lower tail, and $\xi = 0$ (or, $\xi \to 0$) to the Gumbel distribution with a light upper tail. The main focus of interest in extreme value theory is in deriving the quantile value (which, in our study, is defined as the extreme load level l_T), given the target probability P_T. The quantile value can be expressed in terms of the function of parameters as follows:

$$
\begin{aligned}
l_T &= \mu - \frac{\sigma}{\xi}\left[1-(-\log(1-P_T))^{-\xi}\right] \quad \text{if} \quad \xi \neq 0 \\
&= \mu - \sigma\log\left[-\log(1-P_T)\right] \qquad\qquad \text{if} \quad \xi = 0. \quad (2.3)
\end{aligned}
$$

The virtue of the binning method is that by modeling the short-term distribution with a homogeneous GEV distribution (i.e. keep the parameters therein constant), it provides a simple way to handle the overall non-stationary load response across different wind speeds. The binning method is perhaps the most common method used in the wind industry and is also recommended by IEC (2005a). For example, Agarwal and Manuel (2008) use the binning method to estimate the extreme loads for a 2MW offshore wind turbine. In each weather bin, they use the Gumbel distribution to explain the probabilistic behavior of the mudline bending moments of the turbine tower. The data were collected for a period of 16 months. However, most bins have a small number of data, or sometimes, no data at all. For the bins without data,

the authors estimate the short-term distribution parameters by using a weighted average of all non-empty bins with the weight related to the inverse squared distance between bins. They quantify the uncertainty of the estimated extreme loads using a bootstrapping technique and report 95% confidence intervals for the short-term extreme load given specific weather conditions (weather bin). Because bootstrapping resamples the existing data for a given weather bin, it cannot precisely capture the uncertainty for those bins with limited data or without data.

II.1.3.   Limitations of the binning method

Despite its popularity, the binning method has obvious shortcomings in estimating extreme loads. A major limitation lies in that the short-term load distribution in one bin is constructed separately from the short-term distributions in other bins. This approach requires an enormous amount of data to define the tail of each short-term distribution. In reality, the field data can only be collected in a short duration (e.g., one year out of the 50-year service period), and consequently, some bins do not have enough data. Then, the binning method may end up with inaccuracies or big uncertainties in the estimate of extreme loads. In practice, how many bins to use is also under debate, and there is not yet a consensus. The answer to the action of binning appears to depend on the amount of data – if one has more data, more bins can be used; otherwise, fewer bins.

II.2. Studies on power curve estimation and turbine performance assessment

II.2.1. Current industrial method

As in the extreme load analysis, the current industrial practice of estimating the power curve relies again on binning (IEC, 2005b). For power curve estimation, the binning method is to discretize the domain of wind speed into a greater number of bins, say, using a bin width of 0.5 m/s. Then, the value to be used for representing the power output for a given bin is simply the sample average of all the data points falling within that specific bin, namely:

$$y_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}, \qquad (2.4)$$

where $y_{i,j}$ is the power output of the jth data point in bin i, and N_i is the number of data points in bin i. In the binning method, almost all other environmental variables are ignored, except for the so-called air density adjustment, for which we will present a detailed expression later.

II.2.2. Studies on power curve estimation

Many existing methods of fitting a power curve are similar to the binning method in the sense that only wind speed is used as the sole explanatory variable, although the specific techniques used for curve fitting were quite different (Hayes *et al.*, 2011, Kusiak *et al.*, 2009, Osadciw *et al.*, 2010, Uluyol *et al.*, 2011, Yan *et al.*, 2009). For instance, Yan *et al.* (2009) and Osadciw *et al.* (2010) use a polynomial fitting, a symmetric sigmoid function and a Gaussian cumulative density function (CDF) for curve fitting, and Kusiak *et al.* (2009) use a logistic function. These methods are of a parametric flavor. Kusiak *et al.* (2009) also suggest a nonparametric approach, which

is to use the k-nearest neighborhood (k-NN) method to make a power prediction.

Wan *et al.* (2010) extend to the binning method. In one aspect, they study the wind direction effect, but their approach is simply to divide wind direction into a few disjoint sub-directions; doing this is, in fact, an action of binning. Another extension is that they try a neural network model that took both wind speed and air density as inputs. However, their study concluded that doing so does not appear beneficial. When comparing a few different options, including curve fitting (they do not specify which curve fitting method they used) and binning, Wan *et al.* (2010) conclude that the binning method with air density correction produced the best power curve fitting outcome.

A handful of studies do explicitly include both wind speed and wind direction in their models (Jeon and Taylor, 2012, Nielsen *et al.*, 2002, Pinson *et al.*, 2008, Sánchez, 2006). The inclusion of wind direction is not surprising because of the physical intuition that the turbine's wind power production is directly influenced by how the wind blows. The specific approaches employ in these studies are different: Nielsen *et al.* (2002) use a local polynomial regression; Sánchez (2006) present a dynamic combination of several prediction models based on time-varying coefficients and a recursive solution procedure; Pinson *et al.* (2008) use a total least squares criterion (i.e., orthogonal distance least squares), together with a Huber M-estimator, to achieve a certain degree of robustness. Jeon and Taylor (2012) present inarguably the most sophisticated approach; they employed a conditional kernel density method to estimate $p(y|\boldsymbol{x})$. Not only does their model consider both wind speed and wind direction, but it also produces a density estimation that can be used to account for uncertainty in wind power prediction. Unlike Jeon and Taylor (2012), nearly all the other studies produce only a point estimation.

As a side note, we commented earlier on the two primary utilities of power curves,

namely for wind power estimation and for turbine performance assessment. Among the literature discussed, the works where power curves are used principally for power prediction include Nielsen *et al.* (2002), Sánchez (2006), Pinson *et al.* (2008), Wan *et al.* (2010), Hayes *et al.* (2011) and Jeon and Taylor (2012), and the works where power curves are used principally for assessing turbine performance include Yan *et al.* (2009), Kusiak *et al.* (2009), Uluyol *et al.* (2011), Osadciw *et al.* (2010), and Stephen *et al.* (2011).

II.2.3. Studies on turbine performance assessment

In this section, we summarize the works on turbine performance assessment under the assumption that the power curve is already estimated. Keep in mind that most studies regarding turbine performance tests only consider wind speed as the sole environmental variable in x affecting power output y.

Yan *et al.* (2009) and Osadciw *et al.* (2010) compare the measured data to the fitted nominal power curve for anomaly detection. In the studies concerning anomaly detection, a general solution is to specify a common threshold of deviation over the whole domain of wind speed. However, Yan *et al.* (2009) show that power residuals calculated by power curve estimation indicate a heavy-belly pattern. For this reason, using a common threshold is not applicable to the power residuals. Thus, instead of using power residuals, Yan *et al.* (2009) propose to use the wind residuals and then specify a threshold over the whole domain of wind speed. Here, the wind speed residuals are calculated by using the inverse functions of different power curve estimation models such as a polynomial fitting, a sigmoid function, or a Gaussian CDF fitting function. Osadciw *et al.* (2010) point out the limitations of the inverse function method using a polynomial fitting or a sigmoid function fitting. They suggest employing the Inverse Diagnostic Curve Detector (IDCD) method , which is based

on the Gaussian CDF function. They demonstrate the superiority of their method by using data with a major component failure as well as with a faulty anemometer.

Uluyol *et al.* (2011) calculate their power residuals by using the nominal power curve provided by the turbine manufacturer. In the absence of a manufacturer-provided power curve, they use a polynomial fitting to estimate the power curves. Then, instead of using a common threshold, they recommend a condition indicator (CI) method, which is to generate various statistics of the power residuals such as mean, skewness, and kurtosis. Collectively, all these statistics are called the CI. To calculate these CIs, they use the same idea as the binning method, which is to divide the power residuals into a number of bins and then to compute the statistics for each bin. When compared to a universal threshold method, their method has the advantage of being robust.

For the purpose of on-line monitoring, Kusiak *et al.* (2009) use the statistics of power residuals such as mean and standard deviation from a bin. They subsequently calculate the upper and lower control limits in each bin, and then use the control limits for anomaly detection.

II.2.4. Limitations of existing studies

Based on the findings from the literature review, it seems evident that despite the availability of other environmental measurements and their potential impact on power curve estimation, the current methods predominately made use of wind speed only, and wind speed and direction with less frequency. Consequently, their models have high uncertainty caused by other environmental factors. Moreover, the need to develop power curve methods with multivariate dependency has been recently noted, both directly by Stephen *et al.* (2011) and indirectly in the studies by Tindal *et al.* (2008) and Albers *et al.* (1999).

Particularly, for turbine performance assessment, the power curve plays a critical role. The power residuals calculated by the current methods with speed only reveal a non-random pattern and high variance due to the effects of other environmental factors. These characteristics of the residuals hinder us from detecting small changes in turbine performance. Although such changes can be found, they make it difficult to identify where the changes come from.

CHAPTER III

BAYESIAN SPLINE METHOD FOR ASSESSING EXTREME LOADS ON WIND TURBINES

In this chapter, we first provide some background information regarding wind turbine load responses and the datasets used in this study. Then, we proceed to present the details of our spline method in Section III.2. In Section III.3, we compare the spline method with the current industry practice or the unconditional existing method, arguing that the spline method produces better estimates. Finally, we end this study with some concluding remarks in Section III.4.

III.1. Background and datasets

Figure 4 shows examples of mechanical loads at different components in a turbine system. The flap-wise bending moments measure the loads at the blade roots that are perpendicular to the rotor plane, while the edge-wise bending moments measure the loads that are parallel to the plane. Shaft- and tower- bending moments measure, in two directions, the stresses on the main shaft connected to the rotor and on the tower supporting the wind power generation system (i.e., blades, rotor, generator etc.), respectively.

We only study inland turbines (ILTs) in this work, and use the datasets from three ILTs located at different sites. These datasets were collected by Risø-DTU (Technical University of Denmark) (WindData, 2010). Table 1 summarizes the specifications of the datasets.

We would like to first explain a few terms used in the table as well as in the rest of the paper:
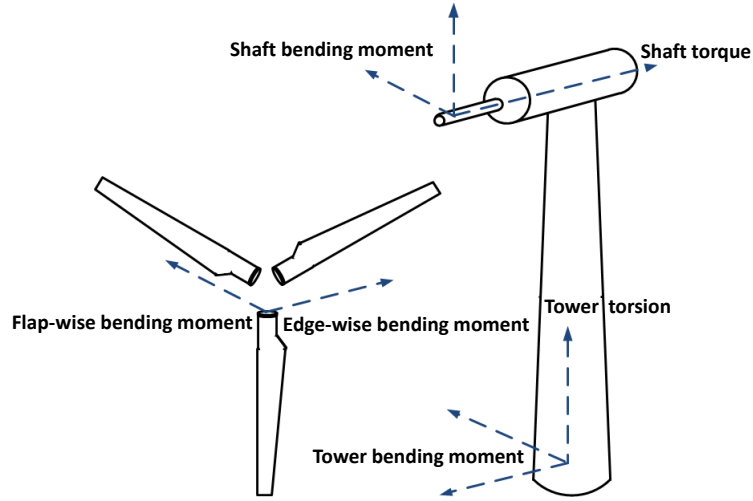
Fig. 4. Illustration of structural loads at different components [Source: The illustration is modified based on a figure originally available at WindData (2010)]

- *Pitch control*: To avoid production of excessive electricity, turbines hold its rotor at an approximately constant speed in high wind speeds. A pitch controlled turbine turns its blades to regulate its rotor speed.

- *Stall control*: This serves the same purpose as in pitch control. But the blade angles do not adjust during operation. Instead the blades are designed and shaped to increasingly stall the blade's angle of attack with the wind to protect the turbine from excessive wind speeds.

- *Cut-in wind speed*: The lowest wind speed at a hub height at which a wind turbine starts to produce power.

- *Cut-out wind speed*: The speed beyond which a wind turbine shuts itself down to protect the turbine.

- *Rated wind speed*: The speed beyond which the turbine's output power needs to be limited, and consequently, the rotor speeds are regulated, by using, for example, a pitch control mechanism.

Among the structural load responses, we consider only the flap-wise bending moments measured at the root of blades. In other words, y in this study is the maximum blade-root flap-wise bending moment. But please note that our method applies to other load responses as well. Regarding weather characteristics, since we consider only the ILTs, we include in $\boldsymbol{x}$ the average wind speed v and the standard deviation of wind speed s; namely $\boldsymbol{x} := (v, s)$.

The data are recorded at different frequencies on the ILTs, as follows:

- *ILT1*: 25Hz (measuremets/sec) = 15,000 measurements/10-min

- *ILT2*: 32Hz (measurements/sec) = 19,200 measurements/10-min

- *ILT3*: 35.7Hz (measurements/sec) = 21,420 measurements/10-min

The raw measured variables are v_{ij} and y_{ij}, where $i = 1, \cdots, n$ represents a 10-minute block of data and $j = 1, \cdots, N$ is the index of the measurements. We use N to represent the number of measurements in a 10 minute block, equal to 15,000, 19,200, and 21,420 for ILT1, ILT2, and ILT3, respectively, and use n to represent the total number of the 10-minute intervals in each dataset, taking the value of 1,154, 595, and 5,688, respectively, for ILT1, ILT2, and ILT3. For these variables, the statistics of the observations in each 10-minute block are calculated as follows:

$$v_i = \frac{1}{N} \sum_{j=1}^{N} v_{ij}, \qquad (3.1)$$

$$s_i = \sqrt{\frac{1}{N-1} \sum_{j=1}^{N} (v_{ij} - v_i)^2} \text{ and} \qquad (3.2)$$

$$y_i = \max \{y_{i1}, y_{i2}, \cdots, y_{iN}\}. \qquad (3.3)$$

Table 1. Specifications of wind turbines in three datasets

Wind turbine model	NEG-Micon/2750	Vestas V39	Nordtank 500
(Name of dataset)	(ILT1)	(ILT2)	(ILT3)
Hub height (m)	80	40	35
Rotor diameter (m)	92	39	41
Cut-in wind speed (m/s)	4	4.5	3.5
Cut-out wind speed (m/s)	25	25	25
Rated wind speed (m/s)	14	16	12
Nominal power (kW)	2,750	500	500
Control system	Pitch	Pitch	Stall
Location	Alborg, Denmark	Tehachapi Pass, California	Roskilde, Denmark
Terrain	Coastal	Bushes	Coastal

III.2.   Bayesian spline method for extreme loads

In this section, we present two sub-models used in our new procedure of estimating the extreme load. The first sub-model (in Section III.2.1) is the conditional maximum load model $p(y|\boldsymbol{x})$, and the second sub-model (in Section III.2.3) is the distribution of wind characteristics $p(\boldsymbol{x})$. Our major undertaking in this study is on the first sub-model, where we present an alternative to the current binning method.

We begin by presenting some scatter plots for the three datasets. Figure 5 shows the scatter plots between the 10-minute maximum loads and 10-minute average wind speeds. We observe nonlinear patterns between the loads and the average wind speeds in all three scatter plots, while individual turbines exhibit different response patterns. ILT1 and ILT2 are two pitch controlled turbines, so when the wind speed reaches or exceeds the rated speed, the blades are adjusted to reduce the absorption of wind energy. As a result, we observe that the loads show a downward trend after the rated wind speed. But different from that of ITL1, the load response of ILT2 has a large variation beyond the rated wind speed. This large variation can be attributed to the less capable control system since ILT2 is one of the early turbine models using a pitch control system. ILT3 is a stall controlled turbine, and its load pattern in Figure 5(c) does not have an obvious downward trend beyond the rated speed.

Figure 6 presents the scatter plots between the 10-minute maximum load and the standard deviation of wind speed during the 10-minute interval. We also observe nonlinear relationships between them, especially for the new pitch-controlled ITL1. Figure 7 shows scatter plots of 10-minute standard deviation versus 10-minute average wind speed. Some previous studies (Fitzwater *et al.*, 2003, Moriarty *et al.*, 2002) suggest that the standard deviation of wind speed varies with the average wind speed, which appears consistent with what we observe in Figure 7.

(a) ILT1   (b) ILT2   (c) ILT3
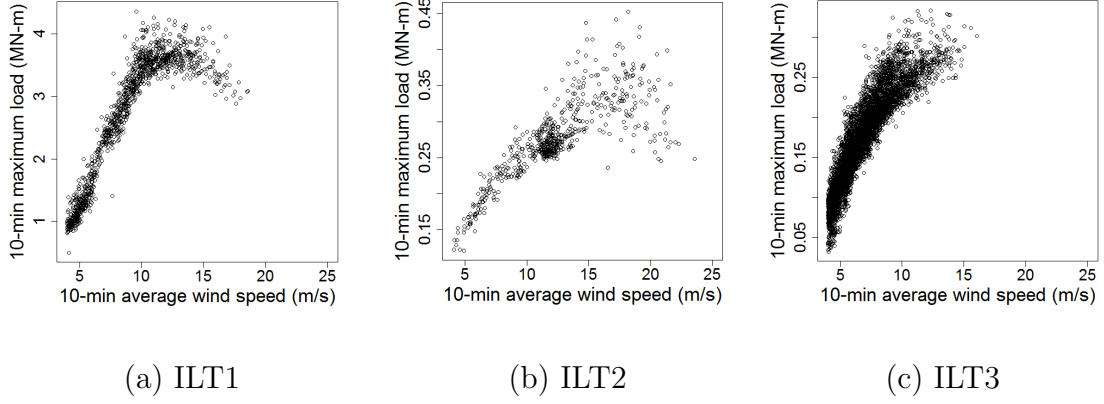
Fig. 5. Scatter plots of 10-minute maximum load versus 10-minute average wind speed

(a) ILT1   (b) ILT2   (c) ILT3
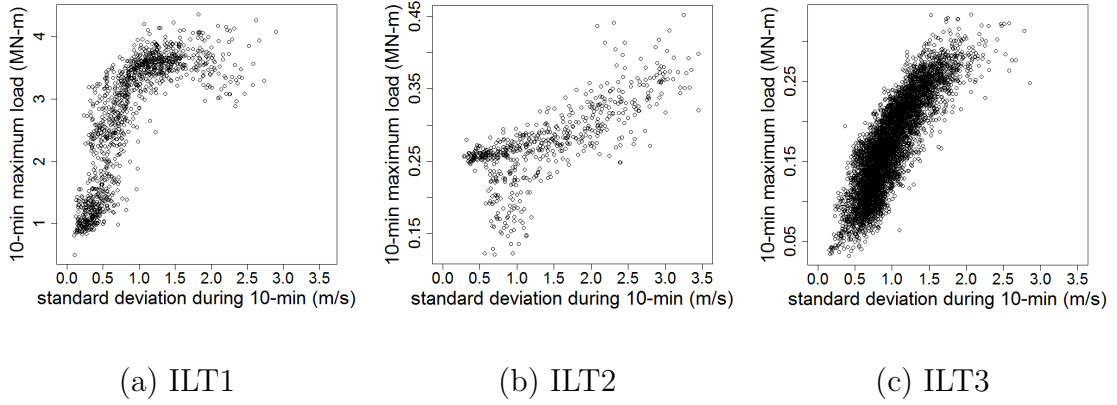
Fig. 6. Scatter plots of 10-minute maximum load versus 10-minute standard deviation of wind speed

(a) ILT1   (b) ILT2   (c) ILT3
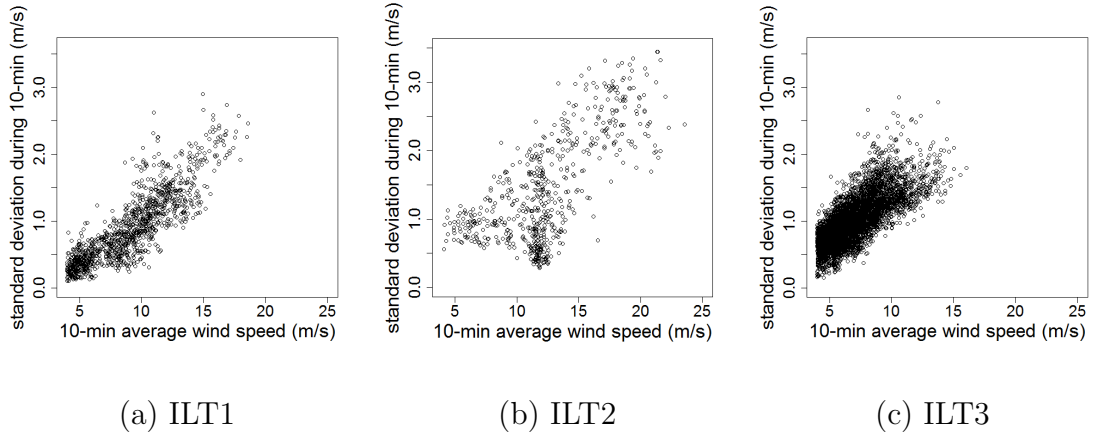
Fig. 7. Scatter plots of 10-minute average wind speed versus 10-minute standard deviation of wind speed

III.2.1. Sub-model 1: Bayesian spline model for conditional maximum load

Recall that in the binning method, a homogeneous GEV distribution is used to model the short-term load distribution, for it appears reasonable to assume stationarity if the chosen weather bin is narrow enough. A finite number of the homogeneous GEV distributions are then stitched together to represent the non-stationary nature of the loads across the entire wind profile. What we propose here is to get rid of the bins and instead use a non-homogeneous GEV distribution whose parameters are not constant but depend on weather conditions.

Our research starts with trying simple approaches based on polynomial models. It turns out that polynomial-based approaches lack the flexibility of adapting to the datasets from different types of turbines. Moreover, due to the nonlinearity around the rated wind speed and the lack of data under high wind speeds, polynomial-based approaches performed poorly in those regions that are generally important for capturing the maximum load. Spline models, on the other hand, appear to work better than a global polynomial model, because they have more supporting points spreading over the input domains. In the sequel, we present two flexible Bayesian spline models for the purpose of establishing the desired non-homogeneous GEV distribution.

Suppose we observe 10-minute maximum loads $y_1, \cdots, y_n$ with corresponding covariate variables $\boldsymbol{x}_1 = (v_1, s_1), \cdots, \boldsymbol{x}_n = (v_n, s_n)$. We choose to model y_i with a GEV distribution:

$$y_i | \boldsymbol{x}_i \;\sim\; GEV(\mu(\boldsymbol{x}_i), \sigma(\boldsymbol{x}_i), \xi), \quad \sigma(\cdot) > 0, \tag{3.4}$$

where the location parameter μ and scale parameter σ in this GEV distribution are a nonlinear function of wind characteristics $\boldsymbol{x}$. The shape parameter ξ is fixed across the

wind profile, while its value will still be estimated using the data from a specific wind turbine. The reason that we keep ξ fixed is to keep the final model from becoming overly complicated. Let us denote $\mu(\boldsymbol{x}_i)$ and $\sigma(\boldsymbol{x}_i)$ by

$$\mu(\boldsymbol{x}_i) = f(\boldsymbol{x}_i), \tag{3.5}$$

$$\sigma(\boldsymbol{x}_i) = \exp(g(\boldsymbol{x}_i)), \tag{3.6}$$

where in the second expression above, an exponential function is used to ensure the positivity of the scale parameter.

Our strategy of modeling $f(\cdot)$ and $g(\cdot)$ is to use a Bayesian MARS (multivariate adaptive regression splines) model (Denison *et al.*, 2002, 1998) for capturing the non-linearity between the load response and the wind-related covariates. The Bayesian MARS model has a high flexibility. It includes the number and locations of knots as part of its model parameters and decides them from observed data. In addition, interaction effects among input factors can be modeled if choosing appropriate basis functions.

More specifically, the Bayesian MARS models $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ for the location parameter μ and the scale parameter σ are represented as a linear combination of the basis functions $B_k^\mu(\boldsymbol{x})$ or $B_k^\sigma(\boldsymbol{x})$:

$$f(\boldsymbol{x}) = \sum_{k=1}^{K_\mu} \beta_k B_k^\mu(\boldsymbol{x}), \tag{3.7}$$

$$g(\boldsymbol{x}) = \sum_{k=1}^{K_\sigma} \theta_k B_k^\sigma(\boldsymbol{x}), \tag{3.8}$$

where $\beta_k, k = 1, \cdots, K_\mu$ and $\theta_k, k = 1, \cdots, K_\sigma$ are the coefficients of the basis functions $B_k^\mu(\cdot)$ and $B_k^\sigma(\cdot)$, respectively, and K_μ and K_σ are the number of basis functions for the location parameter μ and the scale parameter σ, respectively. According to the study by Denison *et al.* (1998), which proposed the Bayesian MARS, its basis

function is specified as follows:

$$
B_k(\boldsymbol{x}) = \begin{cases} 1, & k = 1 \\ \prod_{j=1}^{J_k} \left[h_{jk}(x_{r(j,k)} - t_{jk}) \right]_+, & k = 2, 3, \cdots, K. \end{cases} \tag{3.9}
$$

Here, $[\cdot]_+ = \max(0, \cdot)$, J_k is the degree of interactions modeled by the basis function $B_k(\boldsymbol{x})$, h_{jk} is the sign indicators, taking the value of either -1 or $+1$, $r(j, k)$ produces the index of the predictor variable which is being split on t_{jk}, commonly referred to as the knot points.

In this study, because we only consider two predictors for inland turbines, namely v and s, J_k takes the value of either 1 or 2. When $J_k = 1$, two types of basis functions are used and defined in each of the predictors, represented by $T_k = 1$ and $T_k = 2$, respectively, and no interaction effects between the predictors are modeled. When $J_k = 2$, three types of basis functions are used, represented by $T_k = 1$, $T_k = 2$, and $T_k = 3$, respectively. The three types of basis functions are used to model each predictor as well as their interaction. In our model, we set $J_k = 2$ in the model of the location parameter μ for ILT1 and ILT3 data to allow the interaction to be modeled. For ILT2, however, due to its relatively smaller data amount, a model setting $J_k = 2$ produces unstable and unreasonably wide credible intervals. So for ILT2, only two types of basis functions are used to model its location parameter μ. For the scale parameter σ, we always set $J_k = 1$.

Let $\boldsymbol{\Psi}_a = (\boldsymbol{\Psi}_\mu, \boldsymbol{\Psi}_\sigma, \xi)$ denotes all the parameters used in model (3.4), where $\boldsymbol{\Psi}_\mu$ and $\boldsymbol{\Psi}_\sigma$ include the parameters in function $f(\cdot)$ and $g(\cdot)$, respectively. Those parameters are grouped into two sets: (1) the coefficients of the basis functions in $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{K_\mu})$ or $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_{K_\sigma})$, and (2) the number and locations of the

knots, and the types of basis function in ϕ_μ or ϕ_σ, as follows:

$$\phi_\mu = \left(K_\mu, \Lambda_2^\mu, \cdots, \Lambda_{K_\mu}^\mu \right),$$

$$\text{where } \Lambda_k^\mu = \begin{cases} (T_k^\mu, h_{1k}^\mu, t_{1k}^\mu), & \text{when } T_k^\mu = 1, 2; \\ (T_k^\mu, h_{1k}^\mu, h_{2k}^\mu, t_{1k}^\mu, t_{2k}^\mu), & \text{when } T_k^\mu = 3, \end{cases} \tag{3.10}$$

and

$$\phi_\sigma = \left(K_\sigma, \Lambda_2^\sigma, \cdots, \Lambda_{K_\sigma}^\sigma \right),$$

$$\text{where } \Lambda_k^\sigma = (T_k^\sigma, h_{1k}^\sigma, t_{1k}^\sigma), \quad \text{when } T_k^\sigma = 1, 2. \tag{3.11}$$

Using the above notation, we have $\Psi_\mu = (\beta, \phi_\mu)$ and $\Psi_\sigma = (\theta, \phi_\sigma)$.

To complete the Bayesian formulation for the model in (3.4), priors of the parameters involved should be specified. In this paper, we use uniform priors on ϕ_μ and ϕ_σ as follows;

$$\begin{aligned} p(K) &= \frac{1}{n}, & K = \{1, \cdots, n\} \\ p(T_k) &= \begin{cases} 1, & T_k = \{1\} \\ \frac{1}{2}, & T_k = \{1, 2\} \\ \frac{1}{3}, & T_k = \{1, 2, 3\} \end{cases} \\ p(h_{\cdot k}) &= \frac{1}{2}, & h_{\cdot k} = \{+1, -1\} \\ p(t_{\cdot k}) &= \frac{1}{n}, & t_{\cdot k} = \{v_1, \ldots, v_n\} \text{ or } \{s_1, \ldots, s_n\}. \end{aligned} \tag{3.12}$$

Given ϕ_μ and ϕ_σ, we specify the prior distribution for the parameters (β, θ, ξ) as the unit-information prior, i.e., UIP (Kass and Wasserman, 1995). The UIP is a multivariate normal prior distribution with its mean at the maximum likelihood estimate and its covariance matrix equal to the Fisher information of one observation.

III.2.2. Sub-model 1: Posterior distribution of parameters

The Bayesian MARS model treats the number and locations of the knots as random quantities in a posterior sampling procedure. When the number of knots changes, the dimension of the parameter space changes with it. To handle a varying dimensionality in the probability distributions in a random sampling procedure, we use a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm developed by Green (1995). To allow for dimensional changes, there are three actions in a RJMCMC algorithm: BIRTH, DEATH and MOVE, which adds, deletes, or alters a basis function, respectively. Accordingly, the number of knots as well as the locations of some knots are changed. The detailed definitions of the three actions are given in Denison *et al.* (2002, p. 53), so we need not repeat them here. They suggest the following: use the equal probability (i.e., $\frac{1}{3}$) to propose any of the three moves, and then, use the following acceptance probability α for a proposed move from a model having k basis functions to a model having k^c basis functions:

$$\alpha = \min\left\{1, \textit{the ratio of marginal likelihood} \times R\right\}, \tag{3.13}$$

where R is a ratio of probabilities defined as:

- For a BIRTH action, $R = \dfrac{\text{probability of DEATH in model } k^c}{\text{probability of BIRTH in model } k}$;

- For a DEATH action, $R = \dfrac{\text{probability of BIRTH in model } k^c}{\text{probability of DEATH in model } k}$;

- For a MOVE action, $R = \dfrac{\text{probability of MOVE in model } k^c}{\text{probability of MOVE in model } k}$.

$R = 1$ for most cases, because those probabilities are equal, except when k reaches either the upper or the lower bound.

In our study, the marginal likelihood can be expressed as follows:

$$p\left(\mathcal{D}_y | \boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma\right) =$$
$$\int p\left(\mathcal{D}_y | \boldsymbol{\beta}, \boldsymbol{\theta}, \xi, \boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma\right) p\left(\boldsymbol{\beta}, \boldsymbol{\theta}, \xi | \boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma\right) d\boldsymbol{\beta} d\boldsymbol{\theta} d\xi. \quad (3.14)$$

where $\mathcal{D}_y = (y_1, \cdots, y_n)$ represents a set of observed load data. Kass and Wasserman (1995) and Raftery (1995) showed that when the UIP priors are used, the above marginal log-likelihood can be reasonably approximated by the Schwarz information criterion (SIC) (Schwartz, 1978). The SIC is expressed as:

$$\mathrm{SIC}_{\boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma} = \log\left(p(\mathcal{D}_y | \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\xi}, \boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma)\right) - \frac{1}{2} d_k \log(n).$$

where $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\xi}$ are the maximum likelihood estimators (MLEs) of the corresponding parameters obtained conditional on $\boldsymbol{\phi}_\mu$ and $\boldsymbol{\phi}_\sigma$, and d_k is the total number of parameters to be estimated. In this case, $d_k = K_\mu + K_\sigma + 1$.

Recall that we have two dimension-varying states $\boldsymbol{\phi}_\mu$ and $\boldsymbol{\phi}_\sigma$ in the RJMCMC algorithm. Depending on which state vector is changing, two marginal log-likelihood ratios are needed, and they are approximated by the corresponding SICs, such as:

$$\log \frac{p\left(\mathcal{D}_y | \boldsymbol{\phi}_\mu^c, \boldsymbol{\phi}_\sigma\right)}{p\left(\mathcal{D}_y | \boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma\right)} \simeq \mathrm{SIC}_{\boldsymbol{\phi}_\mu^c, \boldsymbol{\phi}_\sigma} - \mathrm{SIC}_{\boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma}, \text{ and} \quad (3.15)$$

$$\log \frac{p\left(\mathcal{D}_y | \boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma^c\right)}{p\left(\mathcal{D}_y | \boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma\right)} \simeq \mathrm{SIC}_{\boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma^c} - \mathrm{SIC}_{\boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma}. \quad (3.16)$$

Then, we use two acceptance probabilities α_μ and α_σ, respectively, for accepting or rejecting a new state in $\boldsymbol{\phi}_\mu$ and $\boldsymbol{\phi}_\sigma$. Using the SICs, α_μ and α_σ are expressed as:

$$\alpha_\mu = \min\left\{1, \exp\left(\mathrm{SIC}_{\boldsymbol{\phi}_\mu^c, \boldsymbol{\phi}_\sigma} - \mathrm{SIC}_{\boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma}\right) \times R\right\}, \text{ and} \quad (3.17)$$

$$\alpha_\sigma = \min\left\{1, \exp\left(\mathrm{SIC}_{\boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma^c} - \mathrm{SIC}_{\boldsymbol{\phi}_\mu, \boldsymbol{\phi}_\sigma}\right) \times R\right\}. \quad (3.18)$$

As such, we sequentially draw samples for $\boldsymbol{\phi}_\mu$ and $\boldsymbol{\phi}_\sigma$ by using the two acceptance probabilities, while marginalizing out $(\boldsymbol{\beta}, \boldsymbol{\theta}, \xi)$; and then, conditional on the sampled $\boldsymbol{\phi}_\mu$ and $\boldsymbol{\phi}_\sigma$, draw samples for $(\boldsymbol{\beta}, \boldsymbol{\theta}, \xi)$ using a Normal approximation based on the maximum likelihood estimates and the observed information matrix. This RJMCMC algorithm produces the samples from the posterior distribution of parameters in $\boldsymbol{\Psi}_a$. The detailed simulation procedure can be found in Step I of Subsection III.2.5.

III.2.3. Sub-model 2: Distribution of wind characteristics

To find a site-specific load distribution, the distribution of wind characteristics $p(\boldsymbol{x})$ in (1.3) needs to be specified. Since a statistical correlation is noticed between the 10-minute average wind speed v and the standard deviation of wind speed s in Figure 7, the distribution of wind characteristics $p(\boldsymbol{x})$ can be written as a product of the average wind speed distribution $p(v)$ and the conditional wind standard deviation distribution $p(s|v)$. In this section, we separately discuss how to specify each model.

For modeling the 10-minute average wind speed v, the IEC standard suggests using a 2-parameter Weibull distribution (W2) or Rayleigh distribution (RAY) (IEC, 2005a). These two distributions are arguably the most widely used ones for this purpose. Carta *et al.* (2008) and Li and Shi (2010) note that under different wind regimes other distributions may fit wind speed data better, including the 3-parameter Weibull distribution (W3), 3-parameter log-Normal distribution (LN3), 3-parameter Gamma distribution (G3), and 3-parameter inverse-Gaussian distribution (IG3). We take a total of six candidate distribution models for average wind speed (W2, W3, RAY, LN3, G3, IG3) from Li and Shi (2010), and conduct a Bayesian model selection to choose the best distribution fitting a given average wind speed dataset.

We assume UIP priors for the parameters involved in the aforementioned models, and our approach is again based on maximizing the SIC. Once the best wind speed

model is chosen, we denote it by $\mathcal{M}_v$. Then, the 10-minute average wind speed v is expressed as:

$$v_i \quad \sim \quad \mathcal{M}_v(\boldsymbol{\nu}), \tag{3.19}$$

where $\boldsymbol{\nu}$ is the set of parameters specifying $\mathcal{M}_v$. For instance, if $\mathcal{M}_v$ is W3, then $\boldsymbol{\nu} = (\nu_1, \nu_2, \nu_3)$, where ν_1, ν_2 and ν_3 represent the shape, scale, and shift parameter, respectively, of a 3-parameter Weibull distribution.

For modeling the standard deviation of wind speed s, given the average wind speed v, the IEC standard recommends using a 2-parameter Truncated Normal distribution (TN2) (IEC, 2005a), which appears to be what people have commonly used; see, for example, Fitzwater *et al.* (2003). The distribution is characterized by a location parameter η and a scale parameter δ. In the literature, both η and δ are each treated as a constant. But we observe that datasets measured at different sites have different relationship between the average wind speed v and the standard deviation s. Some of v-versus-s scatter plots show nonlinear pattens.

Motivated by this observation, we decide to employ a Bayesian MARS model for modeling η and δ, similar to what we did in Sub-model 1. The standard deviation of wind speed s, conditional on the average wind speed v, can then be expressed as:

$$s_i | v_i \quad \sim \quad TN2(\eta(v_i), \delta(v_i)), \tag{3.20}$$
$$where \quad \eta(v_i) = f_\eta(v_i) \quad and \quad \delta(v_i) = \exp(g_\delta(v_i)),$$

where f_η and g_δ, like their counterpart in (3.7) and (3.8), are a linear combination of the basis functions taking the general form of (3.9). Notice that both of the functions have only one input variable, which is the average wind speed.

Let $\boldsymbol{\Psi}_\eta = (\boldsymbol{\beta}_\eta, \boldsymbol{\phi}_\eta)$ and $\boldsymbol{\Psi}_\delta = (\boldsymbol{\theta}_\delta, \boldsymbol{\phi}_\delta)$ denote the parameters in function $f_\eta(\cdot)$

and $g_\delta(\cdot)$. Since the basis functions f_η and g_δ in (3.20) have only one input variable, only one type of basis function (i.e., $T_k = 1$) is needed. Hence, ϕ_η and ϕ_δ are much simpler than ϕ_μ and ϕ_σ, their counterparts in (3.10) and (3.11), and are expressed as follows:

$$\phi_\eta = \left(K_\eta, \mathbf{\Lambda}_2^\eta, \cdots, \mathbf{\Lambda}_{K_\eta}^\eta \right),$$
$$where \ \mathbf{\Lambda}_k^\eta = (T_k^\eta, h_{1k}^\eta, t_{1k}^\eta) \quad and \quad T_k^\eta = 1; \tag{3.21}$$

and

$$\phi_\delta = \left(K_\delta, \mathbf{\Lambda}_2^\delta, \cdots, \mathbf{\Lambda}_{K_\delta}^\delta \right), \tag{3.22}$$
$$where \ \mathbf{\Lambda}_k^\delta = (T_k^\delta, h_{1k}^\delta, t_{1k}^\delta) \quad and \quad T_k^\delta = 1.$$

We choose the prior distribution for $(\boldsymbol{\beta}_\eta, \boldsymbol{\theta}_\delta)$ as UIP and the prior for $(\boldsymbol{\phi}_\eta, \boldsymbol{\phi}_\delta)$ as uniform distribution in (3.12), and solve this Bayesian MARS model by using a RJMCMC algorithm, same as in the preceding two subsections. The predictive distributions of the average wind speed $\tilde{v}$ and the standard deviation $\tilde{s}$ are

$$p(\tilde{v}|\mathcal{D}_v) = \int p(\tilde{v}|\boldsymbol{\nu})p(\boldsymbol{\nu}|\mathcal{D}_v)d\boldsymbol{\nu} \tag{3.23}$$
$$p(\tilde{s}|\tilde{v}, \mathcal{D}_v, \mathcal{D}_s) = \int \int p(\tilde{s}|\tilde{v}, \mathbf{\Psi}_\eta, \mathbf{\Psi}_\delta)p(\mathbf{\Psi}_\eta, \mathbf{\Psi}_\delta|\mathcal{D}_v, \mathcal{D}_s)d\mathbf{\Psi}_\eta d\mathbf{\Psi}_\delta \tag{3.24}$$

where $\mathcal{D}_v$ and $\mathcal{D}_s$ are the datasets of the observed average wind speeds and the standard deviations. The detailed simulation procedure is included in Step II in Subsection III.2.5.

III.2.4. Empirical predictive distribution of the extreme load level l_T

We are interested in getting the empirical predictive distribution of the quantile value l_T, based on the observed load and wind data $\mathcal{D} := (\mathcal{D}_y, \mathcal{D}_v, \mathcal{D}_s)$. In order to do so,

we need to draw samples, $\tilde{y}$'s, from the predictive distribution of the maximum load given parameters, $p[\tilde{y}|\mathcal{D}, \mathbf{\Psi}_a]$, which is

$$p[\tilde{y}|\mathcal{D}, \mathbf{\Psi}_a] = \int \int p[\tilde{y}|\tilde{v}, \tilde{s}, \mathbf{\Psi}_a, \mathcal{D}]p[\tilde{v}, \tilde{s}|\mathcal{D}_v, \mathcal{D}_s]d\tilde{v}d\tilde{s}, \qquad (3.25)$$

where $p[\tilde{v}, \tilde{s}|\mathcal{D}_v, \mathcal{D}_s]$ can be expressed as the product of (3.23) and (3.24).

To calculate a quantile value of the load for a given P_T (as in (1.2)), we go through the following steps:

- Draw samples from the joint posterior predictive distribution $p[\tilde{v}, \tilde{s}|\mathcal{D}_v, \mathcal{D}_s]$ of wind characteristics $(\tilde{v}, \tilde{s})$ (Step II in in Section III.2.5);

- Draw a set of samples from the posterior distribution of model parameters $\mathbf{\Psi}_a = (\mathbf{\Psi}_\mu, \mathbf{\Psi}_\sigma, \xi)$; this is realized by employing the RJMCMC algorithm in Section III.2.2 (or Step I in Section III.2.5);

- Given the above samples of wind characteristics and model parameters, one can calculate (μ, σ, ξ) that are needed in a GEV distribution; this yields a short-term distribution $p[\tilde{y}|\tilde{v}, \tilde{s}, \mathbf{\Psi}_a]$;

- Integrating out the wind characteristics $(\tilde{v}, \tilde{s})$, one obtains the long-term distribution $p[\tilde{y}|\mathcal{D}, \mathbf{\Psi}_a]$.

- Draw samples from $p[\tilde{y}|\mathcal{D}, \mathbf{\Psi}_a]$, so that one can compute a quantile value $l_T[\mathbf{\Psi}_a]$ corresponding to P_T.

In fact, the predictive mean and Bayesian credible interval of the extreme load level l_T is realized when running the RJMCMC algorithm. The RJMCMC runs through M_l iterations, and at each iteration, we obtain a set of samples of the model parameters $\mathbf{\Psi}_a$ and calculate a $l_T[\mathbf{\Psi}_a]$. Once M_l values of $l_T[\mathbf{\Psi}_a]$ are obtained, its mean and credible intervals can then be numerically computed.

III.2.5. Implementation details

In this section, we provide the detailed implementation procedure for the spline method. The procedure consists of two major steps: (1) construct the empirical predictive distribution of the extreme load level l_T and (2) obtain the joint posterior predictive distribution of wind characteristics (v, s).

1. *Step I*: Construct the empirical predictive distribution of the extreme load level using the Bayesian spline models.

    (a) Set $t = 0$ and the initial $\phi_\mu^{(t)}$ and $\phi_\sigma^{(t)}$ both to be a constant scalar.

    (b) At iteration t, K_μ and K_σ are equal to the number of basis functions specified in $\phi_\mu^{(t)}$ and $\phi_\sigma^{(t)}$. Find the MLEs of $\boldsymbol{\beta}^{(t)}, \boldsymbol{\theta}^{(t)}, \xi^{(t)}$ and the inverse of the negative of Hessian matrix, given $\phi_\mu^{(t)}$ and $\phi_\sigma^{(t)}$.

    (c) Generate u_μ^1 uniformly on $[0, 1]$ and choose a move in the RJMCMC procedure. In the following, $b_{K_\mu}, r_{K_\mu}, m_{K_\mu}$ are the proposal probabilities associated with a move type, and they are all set as $\frac{1}{3}$.

        • If $(u_\mu^1 \leq b_{K_\mu})$ then go to BIRTH step, denoted by $\phi_\mu^* = $ BIRTH-proposal$(\phi_\mu^{(t)})$, which is to augment $\phi_\mu^{(t)}$ with a

        $\boldsymbol{\Lambda}_{K_\mu+1}^\mu$ that is selected uniformly at random;

        • Else if $(b_{K_\mu} \leq u_\mu^1 \leq b_{K_\mu} + r_{K_\mu})$

        then goto DEATH step, denoted by

        $\phi_\mu^* = $ DEATH-proposal$(\phi_\mu^{(t)})$, which is to remove from $\phi_\mu^{(t)}$ a $\boldsymbol{\Lambda}_k^\mu$, where $1 \leq k \leq K_\mu$ is selected uniformly at random;

        • Else go to MOVE step, denoted by $\phi_\mu^* = $MOVE-proposal$(\phi_\mu^{(t)})$, which first do $\phi_\mu^\dagger = $ DEATH-proposal$(\phi_\mu^{(t)})$ and then do $\phi_\mu^* = $ BIRTH-proposal$(\phi_\mu^\dagger)$.

(d) Find the MLEs $(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*, \xi^*)$ and the inverse of the negative of Hessian matrix, given $\boldsymbol{\phi}_\mu^*$ and $\boldsymbol{\phi}_\sigma$.

(e) Generate u_μ^2 uniformly on $[0,1]$ and compute the acceptance ratio α_μ in (3.17), using the results from (b) and (d).

(f) Accept $\boldsymbol{\phi}_\mu^*$ as $\boldsymbol{\phi}_\mu^{(t+1)}$ with probability $\min(\alpha_\mu, 1)$. If $\boldsymbol{\phi}_\mu^*$ is not accepted, let $\boldsymbol{\phi}_\mu^{(t+1)} = \boldsymbol{\phi}_\mu^{(t)}$.

(g) Generate u_σ^1 uniformly on $[0,1]$ and choose a move in the RJMCMC procedure. In the following, $b_{K_\sigma}, r_{K_\sigma}, m_{K_\sigma}$ are the proposal probabilities associated with a move type, and they are all set as $\frac{1}{3}$.

- If $(u_\sigma^1 \leq b_{K_\sigma})$ then go to BIRTH step, denoted by $\boldsymbol{\phi}_\sigma^* = $ BIRTH-proposal$(\boldsymbol{\phi}_\sigma^{(t)})$, which is to augment $\boldsymbol{\phi}_\sigma^{(t)}$ with a $\boldsymbol{\Lambda}_{K_\sigma+1}^\sigma$ that is selected uniformly at random;

- Else if $(b_{K_\sigma} \leq u_\sigma^1 \leq b_{K_\sigma} + r_{K_\sigma})$ then goto DEATH step, denoted by $\boldsymbol{\phi}_\sigma^* = $ DEATH-proposal$(\boldsymbol{\phi}_\sigma^{(t)})$, which is to remove from $\boldsymbol{\phi}^{(t)}$ a $\boldsymbol{\Lambda}_k^\sigma$ where $1 \leq k \leq K_\sigma$ is selected uniformly at random;

- Else go to MOVE step, denoted by $\boldsymbol{\phi}_\sigma^* = $ MOVE-proposal$(\boldsymbol{\phi}_\sigma^{(t)})$, which first do $\boldsymbol{\phi}_\sigma^\dagger = $ DEATH-proposal$(\boldsymbol{\phi}^{(t)})$ and then do $\boldsymbol{\phi}_\sigma^* = $ BIRTH-proposal$(\boldsymbol{\phi}_\sigma^\dagger)$.

(h) Find the MLEs $(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*, \xi^*)$ and the inverse of the negative of Hessian matrix, given $\boldsymbol{\phi}_\mu^{t+1}$ and $\boldsymbol{\phi}_\sigma^*$.

(i) Generate u_σ^2 uniformly on $[0,1]$ and compute the acceptance ratio α_σ in (3.18), using the results from (d) and (h).

(j) Accept $\boldsymbol{\phi}_\sigma^*$ as $\boldsymbol{\phi}_\sigma^{(t+1)}$ with probability $\min(\alpha_\sigma, 1)$. If $\boldsymbol{\phi}_\sigma^*$ is not accepted, let $\boldsymbol{\phi}_\sigma^{(t+1)} = \boldsymbol{\phi}_\sigma^{(t)}$.

(k) After initial burn-ins (in our implementation, initial burn-in is 1,000), draw a posterior sample of $(\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\theta}^{(t+1)}, \xi^{(t+1)})$ from the approximated multivariate normal distribution at the maximum likelihood estimates and the inverse of the negative of Hessian matrix. Depending on the acceptance or rejection that happened in (f) and (j), the MLEs to be used are obtained from either (b), (d), or (h).

(l) Take the posterior sample of $\boldsymbol{\Psi}_a$, obtained in (f),(j), and (k), and one pair of the $N_w \times N_{sw}$ samples of (v, s), obtained in Step II, calculate a sample of μ and σ using (3.7) and (3.8), respectively. Repeat this for all $N_w \times N_{sw}$ samples of (v, s) to get $N_w \times N_{sw}$ samples of μ and σ.

(m) Draw N_l samples for the 10-minute maximum load $\tilde{y}$ from each GEV distribution with μ_i, σ_i, and ξ_i, $i = 1, \cdots, N_w \times N_{sw}$, where μ_i and σ_i are $N_w \times N_{sw}$ samples obtained in (l), and ξ_i is always set as $\xi^{(t+1)}$.

(n) Get the quantile value (that is, the extreme load level $l_T[\boldsymbol{\Psi}_a]$) corresponding to $1 - P_T$ from the $N_w \times N_{sw} \times N_l$ samples of $\tilde{y}$.

(o) To obtain a credible interval for l_T, repeat (b) through (n) M_l times.

2. *Step II*: Obtain the joint posterior predictive distribution of wind characteristics (v, s)

(a) Find the MLEs of $\boldsymbol{\nu}$ for all candidate distributions listed in Section III.2.3.

(b) Use the SIC to select the "best" distribution model for the average wind speed v. The chosen distribution is used in the subsequent steps to draw posterior samples.

(c) Draw a posterior sample of $\boldsymbol{\nu}$ from the approximated multivariate normal

distribution at the maximum likelihood estimates and the inverse of the negative of Hessian matrix.

(d) Draw N_w samples of $\tilde{v}$ using the distribution chosen in (b) with the parameter sampled in (c).

(e) Implement the RJMCMC algorithm again, namely (a) through (k) in Step I, to get one posterior sample of $\boldsymbol{\Psi}_\eta = (\boldsymbol{\beta}_\eta, \boldsymbol{\phi}_\eta)$ and $\boldsymbol{\Psi}_\delta = (\boldsymbol{\theta}_\delta, \boldsymbol{\phi}_\delta)$.

(f) Take the posterior sample of $\boldsymbol{\Psi}_\eta$ and $\boldsymbol{\Psi}_\delta$, obtained in (e), and one sample of $\tilde{v}$, calculate a sample of η and δ using (3.20). Repeat this step for all N_w samples of $\tilde{v}$ to get N_w samples of η and δ.

(g) Draw a sample for the standard deviation of wind speed $\tilde{s}$ from each truncated normal distribution with $\eta_i, \delta_i, i = 1, \cdots, N_w$. Using the N_w samples of η and δ obtained in (f), we obtain N_w samples of $\tilde{s}$.

(h) To get $M_w \times N_w$ samples of $\tilde{v}$ and $\tilde{s}$, repeat (c) through (g) M_w times.

In our implementation, we use $M_w = 1,000$, $M_l = 10,000$, $N_w = 100$, and $N_l = 100$.

III.3. Results

III.3.1. Model selection

Table 2 presents the SIC values of the six candidate average wind speed models using different ILT datasets. In Table 2, the boldfaced values indicate the largest SIC for a given dataset, and consequently, the corresponding models are chosen for that dataset.

Regarding the average wind speed model, all candidate distributions except RAY provide generally a good model fit for ILT1, with a similar level of fitting quality, but W3 edges out slightly. For the ILT2 data, W2, W3, LN3 and G3 produce similar

SIC values. In the ILT3 data, W3, LN3, G3 and IG3 perform similarly. Still W3 is slightly better. So we choose W3 as our average wind speed model.

Table 2. SIC for the average wind speed model

Datasets	W2	W3	RAY	LN3	G3	IG3
ILT1	-2984.472	**-2941.11**	-3120.00	-2989.03	-2974.11	-2986.38
ILT2	-1666.826	**-1663.26**	-1778.61	-1665.73	-1665.56	-2312.63
ILT3	-12287.11	**-11242.02**	-13395.62	-11443.59	-11289.79	-11410.18

III.3.2. Point-wise credible intervals

As a form of checking the conditional maximum load model, we present in Figures 8 and 9 the 95% point-wise credible intervals under different wind speeds and standard deviations. To generate these figures, we take a dataset and fix v or s at one specific speed or standard deviation at a time and then draw the posterior samples for $\tilde{y}$ from the posterior predictive distribution of conditional maximum load, $p(\tilde{y}|\boldsymbol{x})$. Suppose that we want to generate the credible intervals at wind speed v_* or standard deviation s_*. The posterior predictive distributions are computed as follows:

$$p(\tilde{y}|(v,s) \in \mathcal{D}_{v_*}, \mathcal{D}_y) = \int p(\tilde{y}|(v,s) \in \mathcal{D}_{v_*}, \boldsymbol{\Psi}_a)p(\boldsymbol{\Psi}_a|\mathcal{D}_y)d\boldsymbol{\Psi}_a,$$

$$p(\tilde{y}|(v,s) \in \mathcal{D}_{s_*}, \mathcal{D}_y) = \int p(\tilde{y}|(v,s) \in \mathcal{D}_{s_*}, \boldsymbol{\Psi}_a)p(\boldsymbol{\Psi}_a|\mathcal{D}_y)d\boldsymbol{\Psi}_a,$$

where $\mathcal{D}_{v_*}$ and $\mathcal{D}_{s_*}$ are subsets of the observed data such that $\mathcal{D}_{v_*} = \{(v_i, s_i) : v_* - 0.5 < v_i < v_* + 0.5, \text{ and, } (v_i, s_i) \in \mathcal{D}_{v,s}\}$ and $\mathcal{D}_{s_*} = \{(v_i, s_i) : s_* - 0.05 < s_i <$

43

$s_* + 0.05$, and, $(v_i, s_i) \in \mathcal{D}_{v,s}$}. Given these distributions, samples for $\tilde{y}$ are drawn to construct the 95% credible intervals at v_* or s_*. The result is shown as one vertical bar in either a v-plot (Figure 8) or a s-plot (Figure 9). To complete those figures, the process is repeated in the v-domain with 1 m/s increment and in the s-domain with 0.2 m/s increment. These figures show that the variability in data are reasonably captured by the spline method.

| (a) ILT1 | (b) ILT2 | (c) ILT3 |
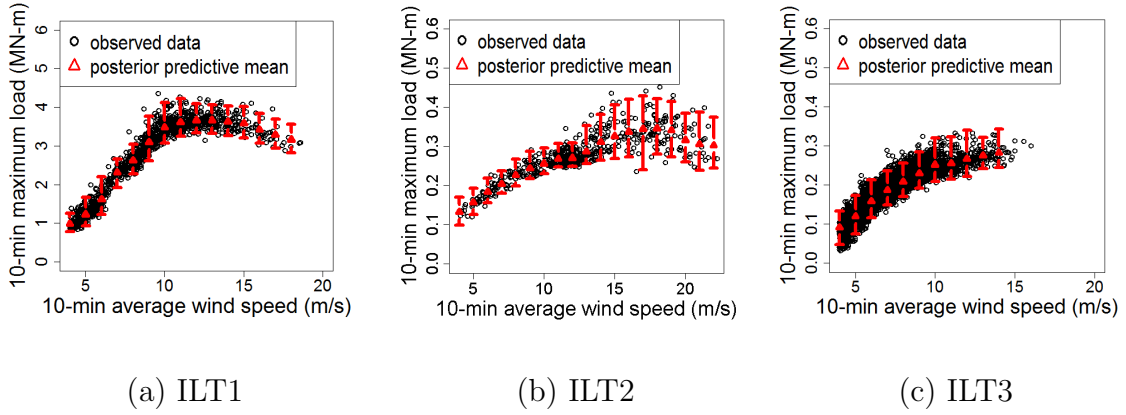
Fig. 8. 95% point-wise credible intervals for different 10-min average wind speeds

III.3.3. Comparison between the binning method and spline method for conditional maximum load

In our procedure for estimating extreme load level, two different distributions are considered for maximum loads y: one is the conditional maximum load distribution $p(y|\boldsymbol{x})$, In our procedure for estimating the extreme load level, two different distributions of maximum load y are involved: one is the conditional maximum load distribution $p(y|\boldsymbol{x})$, a.k.a. the short-term distribution, and the other is the unconditional maximum load distribution $p(y)$, a.k.a. the long-term distribution. Using the observed field data, it is difficult to assess the estimation accuracy of the extreme load

(a) ILT1   (b) ILT2   (c) ILT3
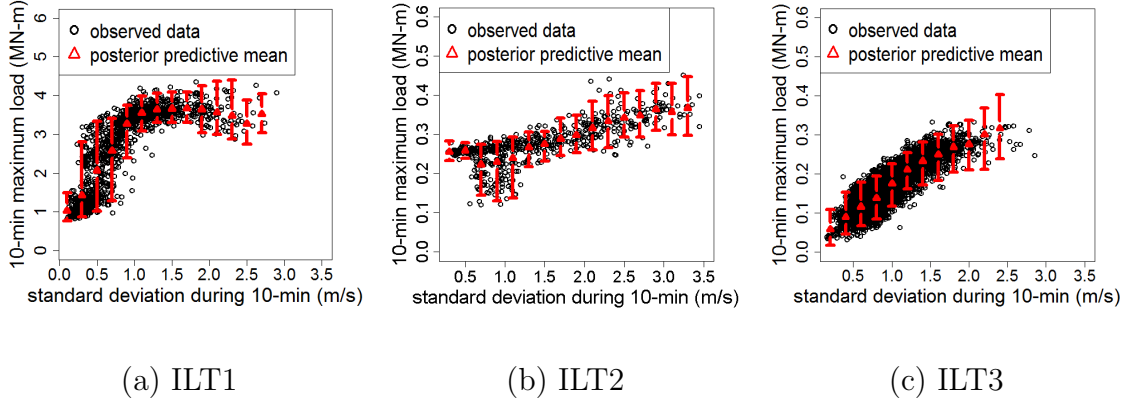
Fig. 9. 95% point-wise credible intervals for different standard deviations

levels based on the long-term distribution, because of the relatively small amount of observation records. What we undertake in this subsection is to evaluate a method's performance of estimating the tail of the short-term distribution $p(y|\boldsymbol{x})$. We argued before that the short-term distribution underlies the difference between the proposed Bayesian spline method and the binning method. The comparison is intended to show the advantage of the Bayesian spline method. In the next subsection, we employ a simulation study that generates a much larger dataset, allowing us to compare the different methods in terms of their performances in estimating the extreme load level based on the long-term distribution.

To evaluate the tail part of a conditional maximum load distribution, we need to compute a set of upper quantile estimators and assess their estimation qualities using the generalized piecewise linear (GPL) loss function (Gneiting, 2011). A GPL

is defined as follows:

$$S_{\tau,b}(\hat{l}(\boldsymbol{x}_i), y(\boldsymbol{x}_i))$$

$$= \begin{cases} \left(\mathbb{1}(\hat{l}(\boldsymbol{x}_i) \geq y(\boldsymbol{x}_i)) - \tau\right) \frac{1}{|b|}([\hat{l}(\boldsymbol{x}_i)]^b - [y(\boldsymbol{x}_i)]^b) & if \quad b \neq 0, \\ \left(\mathbb{1}(\hat{l}(\boldsymbol{x}_i) \geq y(\boldsymbol{x}_i)) - \tau\right) \log\left(\frac{\hat{l}(\boldsymbol{x}_i)}{y(\boldsymbol{x}_i)}\right) & if \quad b = 0, \end{cases} \quad (3.26)$$

where $\hat{l}(\boldsymbol{x}_i)$ is the τ-quantile estimation of $p(y|\boldsymbol{x}_i)$ for a given $\boldsymbol{x}_i$, $y(\boldsymbol{x}_i)$ is the observed maximum load in the test dataset, given the same $\boldsymbol{x}_i$, b is a power parameter, and $\mathbb{1}$ is an indicator function. The power parameter b usually ranges between 0 and 2.5. When $b = 1$, the GPL loss function is the same as the piecewise linear (PL) loss function.

For the above empirical evaluation, we randomly divide a dataset into a partition of 80% for training and 20% for testing. We use the training set to establish a short term distribution $p(y|\boldsymbol{x})$. For any $\boldsymbol{x}_i$ in the test set, the τ-quantile estimation $\hat{l}(\boldsymbol{x}_i)$ can be computed using $p(y|\boldsymbol{x})$. And then, the GPL loss function value for a method is taken as the average of all $S_{\tau,b}$ values over the test set, as follows:

$$\overline{S}_{\tau,b} = \frac{1}{n_t} \sum_{i=1}^{n_t} S_{\tau,b}(\hat{l}_i(\boldsymbol{x}_i), y_i), \quad (3.27)$$

where n_t is the number of data points in a test set, and y_i is the same as $y(\boldsymbol{x}_i)$. We call $\overline{S}_{\tau,b}$ the mean score. We repeat the training/test procedure 10 times, and the final mean score is the average of the ten mean scores. For notational simplicity, we still call the final score the mean score and use $\overline{S}_{\tau,b}$ to represent it, as long as its meaning is clear in the context.

In this comparison, we use two methods to establish the short term distribution: the binning method and the proposed Bayesian spline method. In our RJMCMC algorithm in Section III.2.5, we draw $N_l = 100$ samples from the short term distribution. Accordingly, we evaluate the quality of quantile estimations of the short term

distribution for a τ to be from 0.9 to 0.99.

We first take a look at the comparisons in Figure 10, which compares the PL loss (i.e., $b = 1$) of both methods as τ varies in the above-mentioned range. The left vertical axis shows the values of the mean score of the PL loss, while the right axis is the percentage values of the reduction in mean scores when the spline method is compared with the binning method. For all three datasets, the spline method maintains lower mean scores than the binning method.

When τ is approaching 0.99 in Figure 10, it looks like that the PL losses of the spline and binning methods are getting closer to each other. This is largely due to the fact that the PL loss values are smaller at a higher τ, so that their differences are compressed in the figure. If one looks at the dotted line in a plot, which represents the percentage of reduction in the mean score, the spline method's advantage over the binning method is more evident when τ gets larger in the cases of ILT1 and ILT3 datasets; the spline method produces a PL loss $33\% \sim 50\%$ lower than the binning method. The trend is different when using the ILT2 dataset. But still, the spline method can reduce the mean scores of the PL loss from $8\% \sim 20\%$. Please note that ILT2 dataset is the smallest set, having slightly fewer than 600 data records. We believe that the difference observed over the ILT2 case is attributable to the scarcity of data.

Furthermore, we compute the mean scores of the GPL loss under three different power parameters $b = 0, 1, 2$ when using the two methods to estimate the conditional maximum load. Tables 3 presents the results under $\tau = 0.9$, while Tables 4 is for $\tau = 0.99$. In Table 3, the spline method has a mean score 20% to 42% lower than the binning method. In Table 4, the reductions in mean scores are in a similar range. Overall, the above sets of results show clearly the improvement achieved by employing the Bayesian spline method.
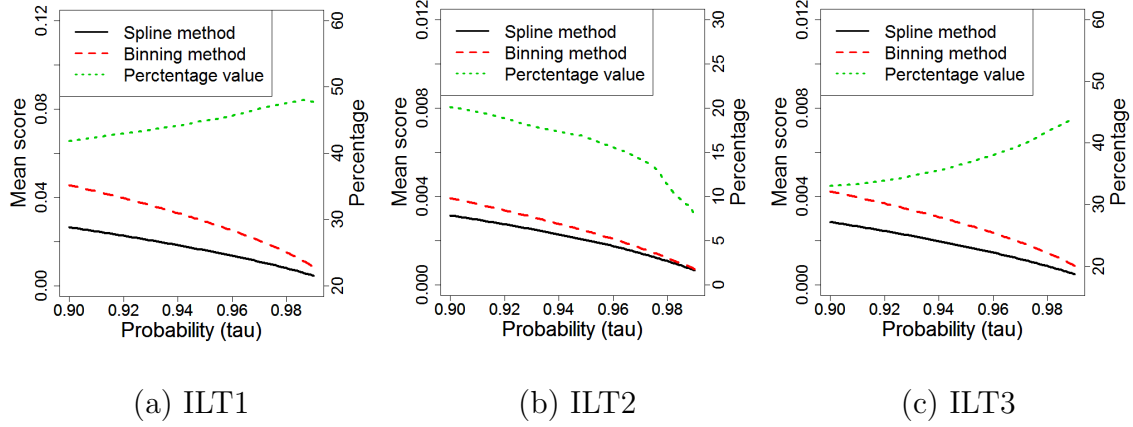
Fig. 10. Comparison of PL function: left and right Y-axis represents mean score values and percentage values, which are the reduction in terms of mean scores compared with those of the binning method, respectively.

In order to understand the difference between the spline method and binning method, we compare the 0.99 quantiles of the 10-minute maximum load conditional on a specific wind condition. This is done by computing the difference in the quantile values of maximum load from the two methods for different weather bins. The wind condition of each bin is approximated by the median values of v and s in that bin. Figure 11 shows the standardized difference of the two 0.99 quantile values in each bin. The darker the color is, the bigger the difference. Note that we exclude comparisons in the weather bins with very low likelihood; namely low wind speed and high standard deviation or high wind speed and low standard deviation.

We can observe that the two methods produce similar results at the bins having a sufficient number of data points (mostly weather bins in the central area), and the results are different when the data are scarce – this tends to happen at the two ends of the average wind speed and standard deviation. This echoes the point we made earlier that without binning the data, the spline method is able to make better use

Table 3. Mean scores of GPL/PL for the 0.9-quantile estimators

Power parameter	ILT1		ILT2		ILT3	
	Binning	Spline	Binning	Spline	Binning	Spline
$b = 0$	0.0185	0.0108	0.0129	0.0103	0.0256	0.0171
$b = 1$	0.0455	0.0265	0.0040	0.0031	0.0042	0.0028
$b = 2$	0.1318	0.0782	0.0013	0.0010	0.0008	0.0005

Table 4. Mean scores of GPL/PL for the 0.99-quantile estimators

Power parameter	ILT1		ILT2		ILT3	
	Binning	Spline	Binning	Spline	Binning	Spline
$b = 0$	0.0031	0.0018	0.0022	0.0020	0.0045	0.0027
$b = 1$	0.0086	0.0045	0.0007	0.0006	0.0008	0.0005
$b = 2$	0.0270	0.0135	0.0003	0.0002	0.0002	0.0001

of the available data and overcome the limited data problem for rare weather events.

III.3.4. Simulation of extreme load

In this subsection, a simulation study is undertaken to assess a method's estimation accuracy of extreme load level based on the long-term distribution. The simulations use one single covariate x, mimicking the wind speed, and a dependent variable y, corresponding to the maximum load. We use the following procedure to generate the simulated data:

(a) ILT1 dataset   (b) ILT2 dataset   (c) ILT3 dataset
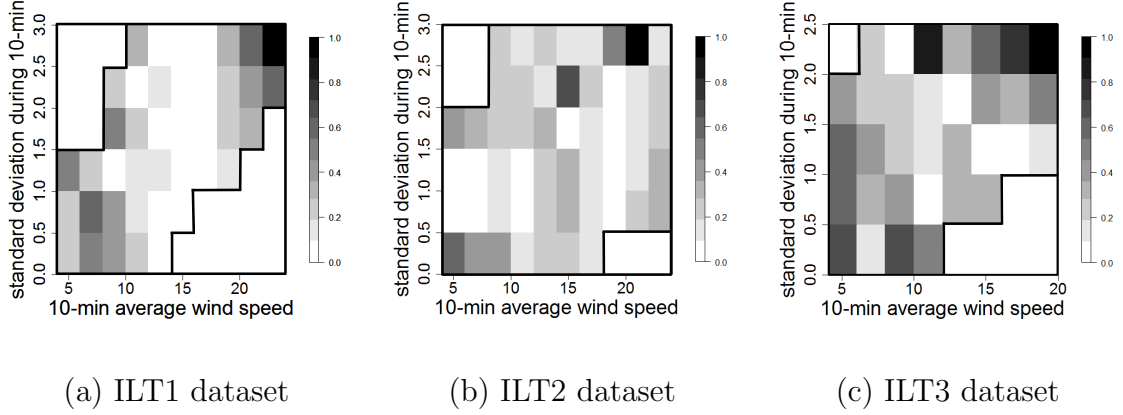
Fig. 11. Comparison of the 0.99-quantiles between the binning method and spline method

1. Generate a sample x_i from a 3-parameter Weibull distribution. Here the 3-parameter Weibull distribution is chosen since x is supposed to be the wind speed. Then sample x_{ij}, $j = 1, \cdots, 1,000$, from a normal distribution with its mean as x_i and having a unit variance. The set of x_{ij}'s is supposed to be the different wind speeds within a bin.

2. Draw the samples y_{ij} from a normal distribution with its mean as μ_{ij}^s and its standard deviation as σ_{ij}^s, which are expressed as follows:

$$
\begin{aligned}
\mu_{ij}^s &= \begin{cases} \frac{1.5}{[1+48\times\exp(-0.3\times x_{ij})]} & if \quad x_i < 17 \\[2mm] \frac{1.5}{[1+48\times\exp(-0.3\times x_{ij})]} + [0.5 - 0.0016 \times (x_i + x_i^2)] & if \quad x_i \geq 17 \end{cases} \\
\sigma_{ij}^s &= 0.1 \times \log(x_{ij}).
\end{aligned}
$$

The above set of equations is used to create a y response resembling the load data we observe. The parameters used in the equations are chosen through trials-and-errors so that the simulated y looks like the actual mechanical load response. While many of the parameters used above do not have any physical

50

meaning, some of them do; for instance, the "17" in "$x_i < 17$" bears the same meaning of the rated wind speed.

3. Find the maximum value $y_i = \max\{y_{i,1}, \cdots, y_{i,1000}\}$, corresponding to x_i. According to the classical extreme value theory (Coles, 2001, Smith, 1990), y_i produced in such a way can be modeled by a GEV distribution.

4. Repeat (a) through (c) for $i = 1, ..., 1000$ to produce the training dataset with $n = 1,000$ data pairs, and denote this dataset by

   $\mathcal{D}_{TR} = \{(x_1, y_1), \cdots, (x_{1000}, y_{1000})\}$.

Once the training dataset $\mathcal{D}_{TR}$ is simulated, both the binning method and spline method are used to estimate the extreme load levels l_T corresponding to two probabilities: 0.0001 and 0.00001. This estimation is based on drawing samples from the long-term distribution of y, as described in Section III.2.4, which produces the empirical predictive distribution of l_T. To compare the estimation accuracy of the extreme quantile values, we also generate 100 simulated datasets; each dataset consists of 100,000 data points, which are obtained by repeating the above (a) through (c). For each dataset, we find the observed quantile values $l_{0.0001}$ and $l_{0.00001}$. Using the 100 simulated datasets, we end up with 100 different quantile values $l_{0.0001}$ and $l_{0.00001}$ and then use them to calculate the 95% empirical confidence intervals.

The methods used for estimating extreme load levels are the binning method, the spline method, and and a Peak Over Threshold (POT) method based on a Generalized Pareto (GP) distribution. The GP-based POT method is another popular approach to estimate extreme quantiles. It uses the large-valued observations which exceed a high threshold after fitting a GP distribution. The POT method assumes that the observations above the threshold are independent and identically distributed, so that the parameters in the GP distribution do not depend on wind covariates. For this

reason, POT can be applied to the observations of y directly, and is an unconditional extreme value method. Please see Coles (2001) and Smith (1990) for more details of the POT method.

Conditional modeling is necessary for extreme load assessment of wind turbines. But POT is an unconditional method. In this study, however, since we use the wind data from a given site and then estimate the extreme load for the same site, POT as an unconditional method is applicable, and can be compared to the spline method and binning method, both of which use conditional models. We hope that including POT in our comparison can offer additional insights.

Figure 12(a) shows a scatter plot of the simulated x's and y's in $\mathcal{D}_{TR}$, which resembles the load responses we see previously. Figures 12(b) and (c) present the extreme load levels estimated by the three methods as well as the observed extreme quantile values under the two selected probabilities. We observe that the binning method tends to overestimate the extreme quantile values and yield wider confidence intervals than the spline method. On the other hand, the POT method tends to underestimate the extreme quantile values. Furthermore, the degree of overestimation and underestimation appear to be more pronounced as the probability corresponding to an extreme quantile value goes higher. This observation will be confirmed by what will be observed in the next subsection using the field data. This simulation result suggests that using the binning method for extreme load estimation is not a good practice.

III.3.5. Estimation of extreme load

Finally, Tables 5 and 6 show the estimates of the extreme load level l_T, corresponding to $T = 20$ and $T = 50$ years, respectively. The values in parenthesis are the 95% credible (or confidence) intervals. Figure 13 presents a graphic illustration of the

(a) Simulation dataset   (b) $l_{0.0001}$ (c) $l_{0.00001}$
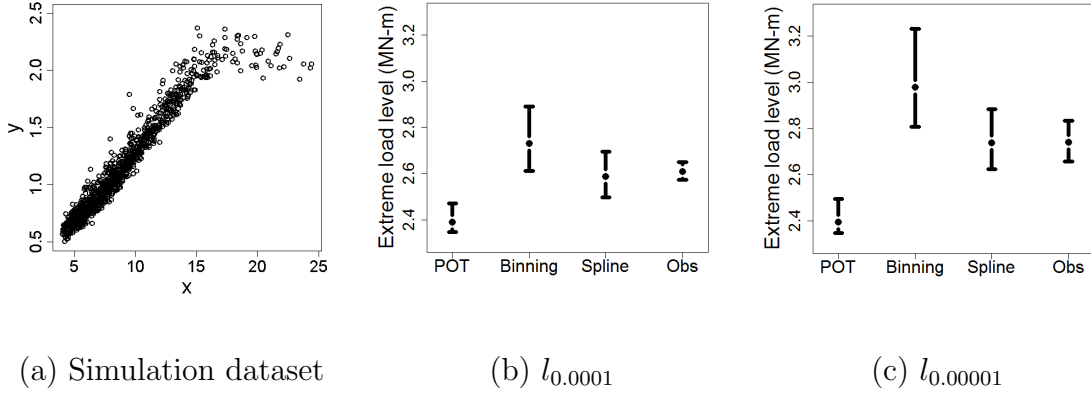
Fig. 12. Simulation dataset, estimated and observed extreme quantile values: For the binning method and spline method, the middle point in (b) and (c) represents the mean of the extreme quantile estimate in the long-term distribution, while the two ending points correspond to the 95% credible (or confidence) intervals. For the observed extreme quantile values, the middle point and the two ending points represent, respectively, the average and the 95% empirical confidence intervals of the 100 observed extreme quantile values.

results in Table 5, facilitating the comparison.

We observe that the extreme load level l_T obtained by the binning method is generally higher than those obtained by the other two methods. This appears consistent with what we see in Figures 12(b) and (c). This should not come as a surprise. As we push for a high quantile, more data would be needed in each weather bin but the amounts in reality are limited due to the binning method's compartmentalization of data. The binning method also produces a wider confidence interval than the spline method, as a result of the same rigidity in data handling. The POT method on ILT1 and ILT3 still seems to underestimate the extreme load levels and this observation is consistent to our simulation result. However, the extreme loads estimated by the

53

spline and POT methods have overlapping credible (or confidence) intervals in general, although less significant in the case of ILT3 data. The spline method produces, on average, an estimation closer to that of POT than that of the binning method.

Based on the above observations, one can conclude a significant difference in the estimations obtained by the spline method and binning method but not so significant between the results from the spline method and POT method. That, however, does not mean that the POT method can be used to substitute for the spline method in the extreme load assessment. Recall that the unconditional POT method is not practical because it requires the load and wind data to be obtained for every new turbine site, while the spline method (or the binning method) does not need this costly proposition and can make use of the conditional load model obtained elsewhere. The detailed procedures for computing confidence intervals in the POT method and the binning method are included in Appendix A and B, respectively.

Table 5. Estimates of extreme load level ($l_T, T = 20$ years), unit: MN-m

Datasets	POT method	Binning method	Spline method
ILT1	4.485 (4.285, 4.748)	6.455 (6.063, 7.092)	4.750 (4.579, 4.955)
ILT2	0.581 (0.495, 0.722)	0.752 (0.658, 0.903)	0.576 (0.538, 0.627)
ILT3	0.369 (0.343, 0.400)	0.505 (0.465, 0.584)	0.428 (0.398, 0.463)

III.4. Discussion

This study presents a Bayesian spline method for estimating the extreme load on wind turbines. The spline method essentially supports a non-homogeneous GEV

Table 6. Estimates of extreme load level ($l_T, T = 50$ years), unit: MN-m

Datasets	POT method	Binning method	Spline method
ILT1	4.492 (4.287, 4.767)	6.710 (6.240, 7.485)	4.800 (4.611, 5.019)
ILT2	0.589 (0.498, 0.742)	0.786 (0.682, 0.957)	0.589 (0.547, 0.646)
ILT3	0.370 (0.344, 0.404)	0.527 (0.480, 0.621)	0.438 (0.405, 0.476)

distribution to capture the nonlinear relationship between the load response and the wind-related covariates. Such treatment avoids binning the data. The underlying spline models instead connect all the bins across the whole wind profile, so that load and wind data are pooled together to produce better estimates. This is demonstrated by applying the spline method to three sets of inland wind turbine load response data and making comparisons with the binning method.

Popularity of the binning method in industrial practice is due to the simplicity of its idea and procedure. But engineers in the wind industry sometimes mistake the simplicity of a procedure as simplicity of a model. Suppose that one uses a 6×10 grid to bin the two-dimensional wind covariates (as we did in this study) and fixes the shape parameter ξ across the bins (a common practice in the industry). The binning method yields 60 local GEV distributions, each of which has two parameters, translating to a total of 121 parameters for the overall model (counting the fixed ξ as well). By contrast, the spline method, although conceptually and procedurally more involved, produces an overall model with fewer parameters. To see this, consider the following: for the three ILT datasets, the average $(K_\mu + K_\sigma)$ from the RJMCMC algorithm is between 12 and 18. The number of model parameters d_k in (3.15) is

(a) ILT1 dataset   (b) ILT2 dataset   (c) ILT3 dataset
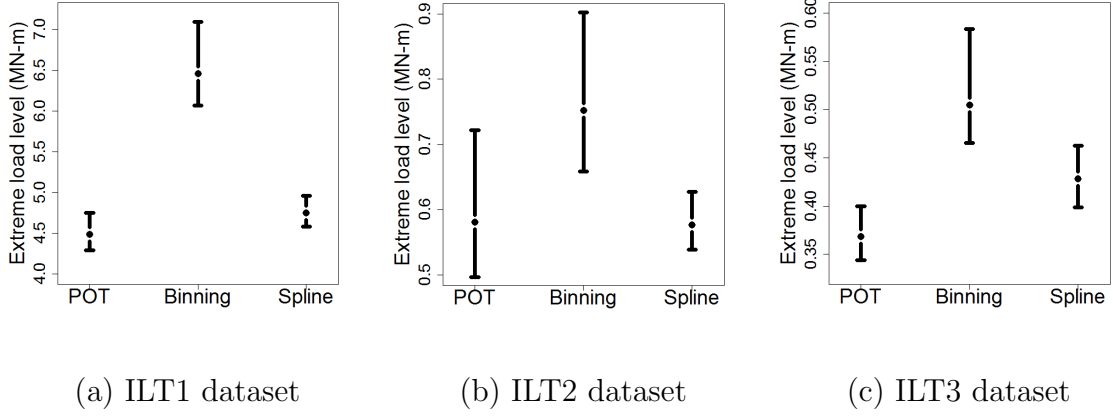
Fig. 13. Graphic illustration of the results in Table 5. The middle point represents the mean of the extreme load estimate, while the two extreme points correspond to the 95% credible (or confidence) intervals.

generally less than 20, a number far smaller than the number of parameters in the binning method. In the end, the spline method uses a sophisticated procedure to find a simpler model that is more capable.

CHAPTER IV

POWER CURVE ESTIMATION WITH MULTIVARIATE ENVIRONMENTAL FACTORS

In this chapter, we first describe the datasets used in this study. We proceed in Section IV.2 to present the details of our additive multivariate kernel model. In Section IV.3, we compare our method with some alternative methods reviewed here, arguing that the resulting kernel method produces better estimates. Finally, we end this study with some discussion in Section IV.4. Note that different from Chapter III, y in this chapter means the power output from a turbine.

IV.1. Datasets

We study both inland wind turbines (ILTs) and offshore turbines (OSTs) in this study, and have in possession two datasets corresponding to an inland wind farm (ILWF) and an offshore wind farm (OSWF), respectively. The datasets are denoted generally by $\mathcal{D}$ or specifically by $\mathcal{D}_{\mathrm{ILWF}}$ or $\mathcal{D}_{\mathrm{OSWF}}$, respectively. Table 7 summarizes the specifications of the datasets.

We choose four wind turbines and two meteorological masts from $\mathcal{D}_{\mathrm{ILWF}}$, and two wind turbines and the single meteorological mast from $\mathcal{D}_{\mathrm{OSWF}}$. The six turbines are denoted as WT1 to WT6, respectively, where the first four are inland turbines, while the last two are offshore ones. The environmental data in $\boldsymbol{x}$ were collected by sensors on a meteorological mast, while the power output y was measured at a wind turbine. Each meteorological mast has two wind turbines associated with it, meaning that the $\boldsymbol{x}$'s measured at this mast are paired with the y's of those associating turbines. For the turbines/masts layout and turbine-to-mast distances, please refer to Figure 14.

Table 7. Specifications of the two wind farms

Wind farm	$\mathcal{D}_{\mathrm{ILWF}}$	$\mathcal{D}_{\mathrm{OSWF}}$
Number of meteorological masts	Multiple	Single
Number of wind turbines	200+	30+
Hub height (m)	80	70
Rotor diameter (m)	about 80	about 90
Cut-in wind speed (m/s)	3.5	3.5
Cut-out wind speed (m/s)	20	25
Rated wind speed (m/s)	around 13	around 15
Rated power (MW)	1.5 - 2.0	around 3
Location	Inland, U.S.	Offshore, Europe

For WT1 and WT2 of the ILWF, the data were collected from July 30, 2010 through July 31, 2011; for WT3 and WT4 (still of the ILWF), the data were from April 29, 2010 through April 30, 2011, and for WT5 and WT6 of the OSWF, the data were from January 01, 2009 through December 31, 2009.

In the current practice, data collected at wind farms are arranged in 10-minute blocks because wind speeds were considered stationary, and other environmental factors nearly constant, over a 10-minute duration. As a result, the power output y as well as the environmental factors v, d, t_m, a_p, h_m are the averages of the recording in a 10-minute duration. Moreover, a few other variables can be computed as follows:

- Wind velocities under the Cartesian coordinates: Wind direction d is a circular variable so that extra cares are needed when calculating the difference between
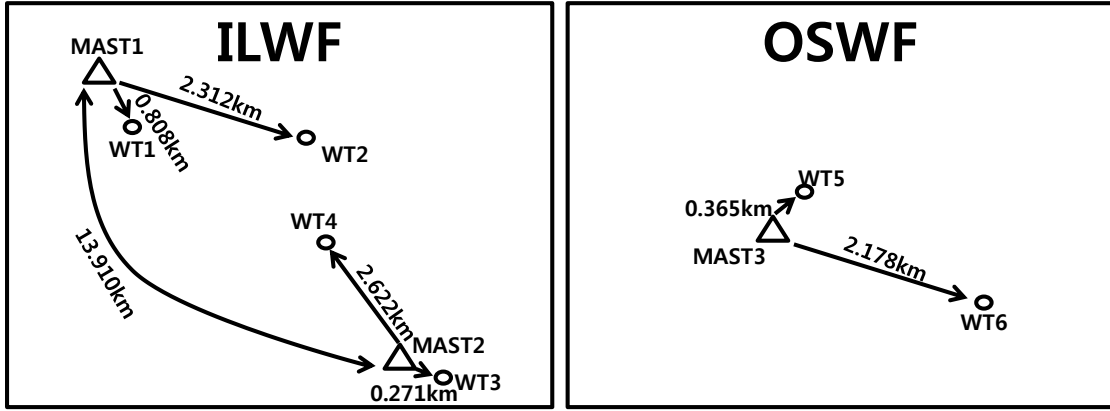
Fig. 14. Layout of the turbines and masts and turbine-to-mast distances: ILWF and OSWF

two wind direction values. An alternative approach that can avoid potential wind direction miscalculation is to convert wind speed and direction into two wind velocities under the Cartesian coordinates, namely $vd_1 = v\cos(d)$ and $vd_2 = v\sin(d)$.

- Turbulence intensity t_b: First compute the standard deviation of the wind speed in a 10-minute duration and denote it by t_b. Then $t_b = \frac{s}{v}$, where v is the average wind speed of the same 10-minute duration.

- Wind shear w_s: Given wind speeds v_1 and v_2, measured at heights g_1 and g_2, respectively. Then, $w_s = \frac{\ln(v_2/v_1)}{\ln(g_2/g_1)}$ (Rehman and Al-Abbadi, 2005). For $\mathcal{D}_{\text{ILWF}}$, wind speeds are measured at two heights of 80 m and 50 m, where 80 m is the hub height. Given this instrumentation capability, one wind shear value is calculated, which is a below-hub wind shear. For $\mathcal{D}_{\text{OSWF}}$, wind speeds are measured at the heights of 116 m, 70 m, 21 m, respectively, where 70 m is the hub height. So two wind shear values can be calculated: using the 116 m/70

m pair produces an above-hub wind shear, while using the 70 m/21 m pair produces a below-hub wind shear. We denote the above-hub wind shear by w_s^a and the below-hub wind shear by w_s^b.

- Air density ρ (kg/m^3): Given air temperature T expressed in Kelvin and air pressure a_p expressed in Newtons/m^2, $\rho = \frac{a_p}{R \cdot t_m}$, where $R = 287$ (Joule)(kg)$^{-1}$ (Kelvin)$^{-1}$ is the gas constant (Uluyol et $al.$, 2011). In the subsequent analysis, the air density ρ, instead of t_m and a_p, is included as an explanatory variable in $\boldsymbol{x}$. The reason will be explained shortly in the next section.

Considering the descriptions presented above, one can see that for $\mathcal{D}_{\mathrm{OSWF}}$, there are seven explanatory variables, i.e., $\boldsymbol{x} = (vd_1, vd_2, \rho, h_m, t_b, w_s^a, w_s^b)$. In $\mathcal{D}_{\mathrm{ILWF}}$, humidity measurements are not available, and the dataset has only the below-hub wind shear. Consequently, $\mathcal{D}_{\mathrm{ILWF}}$ has five explanatory variables, namely $\boldsymbol{x} = (vd_1, vd_2, \rho, t_b, w_s^b)$, two fewer than what $\mathcal{D}_{\mathrm{OSWF}}$ has. Throughout the paper, by "a data point" we refer to a pair of $(\boldsymbol{x}, y)$, and we denote the total number of data points associated with a turbine by N.

IV.2.   An additive multivariate kernel method for power curve estimation

In this section, we first provide some background information on the physical understanding of wind power generation. This physical understanding helps motivate our modeling approach undertaken subsequently.

IV.2.1.   The physics behind wind power generation

The physical law of wind power generation (Ackermann, 2005, Belghazi and Cherkaoui, 2012) states that:

$$y = \frac{1}{2} \cdot C_p(\beta, \lambda) \cdot \rho \cdot \pi R^2 \cdot v^3, \tag{4.1}$$

where R is the radius of the rotor and C_p is the so-called power coefficient, which is believed to be a function of (at least) the blade pitch angle, β, and the turbine's tip speed ratio, λ. What else might affect C_p is still a matter under investigation. Currently no formula exists to express C_p analytically in terms of its influencing factors. C_p is therefore empirically estimated and turbine manufacturers usually provide for a specific turbine its nominal power curve with the corresponding C_p values under different combinations of wind speed, v, and air density, ρ. The above expression also provides the rationale why temperature, t_m, and air pressure, a_p, are converted into air density, ρ, rather than used individually, to explain wind power.

Even though the expression in (4.1) on the surface suggests that the electrical power that a wind turbine extracts from the wind is proportional to v^3, an actual power curve may exhibit a different nonlinear relationship. This happens because of the tip speed ratio, $\lambda = \frac{\omega \cdot R}{v}$, where ω is the rotor speed. Consequently, C_p is also a function of wind speed, v.

The power law in (4.1) governs the wind power generation before the rated wind speed, v_r. The use of the pitch control mechanism levels off, and ultimately caps, the power output when it reaches the rated power output, y_r. Recall the shape of power curve shown in Figure 3. The power curve has an inflection point somewhere nearby the rated wind speed, so that the whole curve consists of a convex segment, between v_{ci} and the inflection point and a concave segment, between the inflection point and

v_{co}.

Given the physical relation expressed in (4.1), the wind industry recognizes the need to include air density as a factor in calculating the power output, and does so through a formula known as the air density correction. If v is the raw average wind speed measured in a 10-minute duration, the air density correction is to adjust the wind speed based on the measured average air density, ρ, in the same 10-minute duration, namely

$$v' = v\left(\frac{\rho}{\rho_0}\right)^{\frac{1}{3}},\tag{4.2}$$

where ρ_0 is the sea-level dry air density ($=1.225\ kg/m^3$) per ISO atmosphere standard. The binning method with air density correction uses this corrected wind speed, v', and the power output, y, to establish a power curve. In the subsequent analysis, as well as in Section IV.3 where we conduct comparisons of methods, by "binning method" we refer to this air density corrected version, unless otherwise noted.

IV.2.2. Additive multivariate kernel density estimation

The underlying physics of wind power generation expressed above provides some clues concerning a preferable power curve model. The following summarizes our observations:

- There appear at least three important factors that affect wind power generation: wind speed v, wind direction d, and air density ρ. This does not exclude the possibility that other environmental factors may also influence the power output.

- The functional relationships between the environmental factors and the power response are generally nonlinear. The complexity partially comes from the lack

of understanding of C_p, which is affected by many environmental factors (v, d, and ρ included). We also stated above that there is no analytical expression linking C_p to any of the influencing factors. As a result, the functional form of a potential power curve is not known either.

- The environmental factors appear in a multiplicative relationship in the power law equation (4.1), indicating interactions among the factors.

The last observation suggests that purely additive models or generalized additive models (GAM) are unlikely to work well in modeling a power curve. The advantage of GAM is that they can be easily expanded to include extra environmental factors, if any, but the interactions among the factors cannot be adequately captured.

To illustrate the need to capture the interaction among the factors, we present the following plots in Figures 15 and 16. Figure 15 uses the data from $\mathcal{D}_{\text{ILWF}}$ and shows the scatter plots between the power output y and air density ρ, turbulence intensity t_b and wind shear w_s^b, respectively. There seems to be no apparent meaningful relationship between the power output and these factors. So unconditional on wind speed v and wind direction d, these environmental factors have no obvious effect on the power output. Including them in a GAM in addition to v and d does not make a power prediction better than the GAM having only v and d as its inputs. On the other hand, Figure 16 presents the scatter plots between the power output and environmental factors under different wind speed v and wind direction d. We do observe nonlinear relationships in these plots, and the relationships appear to be different depending on the wind conditions. This implies that interaction effects exist among wind speed v, wind direction d, and other environmental factors. A power curve model should characterize not only the nonlinear effects of wind speed and wind direction, but also the nonlinear interaction effects among the environmental factors.

(a) Air density   (b) Turbulence intensity   (c) Wind shear
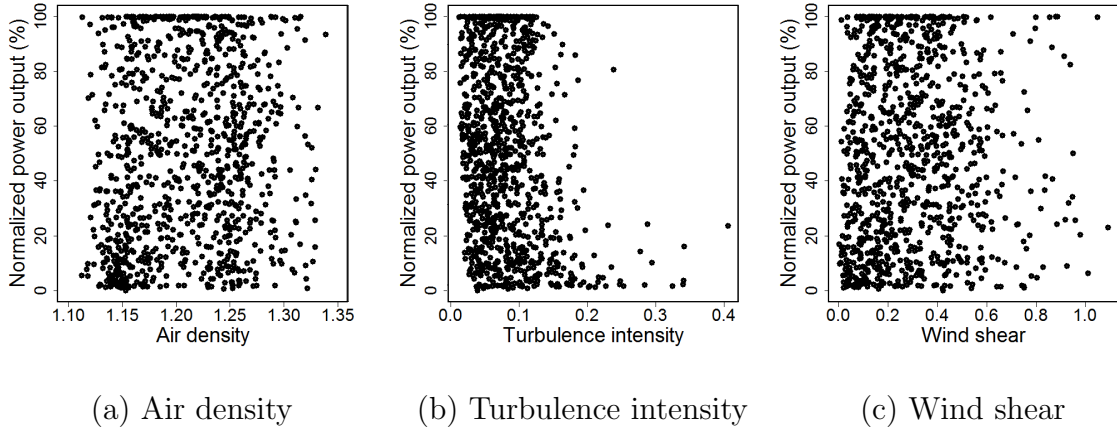
Fig. 15. Scatter plots of the power output versus environmental factors for $3.5 < v < 20, 0 < d < 360$

Because of the complexity of the functional relationships between environmental factors and the power output, nonparametric approaches have been popular in modeling the power curves. Among them, the multivariate conditional kernel density estimation (CKD), introduced by Rosenblatt (1969) and Hyndman *et al.* (1996), or the multivariate kernel regression described by Nadaraya (1964) and Watson (1964), appear to be a capable statistical modeling tool, not only capturing the complicate higher order interaction effects but also avoiding the need to specify a functional form of the power curve relationship. Using the CKD, one can also model the distribution aspect of the power curve for the purpose of quantifying the uncertainty in wind power prediction. A bivariate CKD including wind speed and direction is indeed the choice of Jeon and Taylor (2012) when they modeled the power curve.

Specifically, using the Rosenblatt's CKD (Rosenblatt, 1969), the density of y conditional on $\boldsymbol{x}$ can be expressed as

$$\hat{f}(y|\boldsymbol{x}) = \sum_{i=1}^{N} w_i(\boldsymbol{x})\mathcal{K}_{h_y}(y - y_i), \tag{4.3}$$

where

$$w_i(\boldsymbol{x}) = \frac{\mathcal{K}_{\boldsymbol{h}_x}(||\boldsymbol{x} - \boldsymbol{x}_i||)}{\sum_{i=1}^{N} \mathcal{K}_{\boldsymbol{h}_x}(||\boldsymbol{x} - \boldsymbol{x}_i||)}, \tag{4.4}$$

$\boldsymbol{h}_x = (h_1, \ldots, h_q)$ and h_y are bandwidth parameters controlling the smoothness in, respectively, the environmental factors $\boldsymbol{x}$ and the power output y, and q is the number of explanatory variables in $\boldsymbol{x}$. In our study, $q = 7$ for $\mathcal{D}_{\text{OSWF}}$ and $q = 5$ for $\mathcal{D}_{\text{ILWF}}$.

The above formulation contains kernel functions of two different dimensions $\mathcal{K}_{h_y}(l)$ and $\mathcal{K}_{\boldsymbol{h}_x}(||\boldsymbol{l}||)$. $\mathcal{K}_{h_y}$ is a scaled kernel function and takes the form of $h_y^{-1} K(\frac{l}{h_y})$, where $K(\cdot)$ is assumed to be a real valued, integrable and non negative even function. In our study, K is chosen to be a univariate Gaussian kernel function. $\mathcal{K}_{\boldsymbol{h}_x}(||\boldsymbol{l}||)$ is a multivariate kernel function and is composed of a product kernel that is a multiplication of univariate kernel functions such as

$$\mathcal{K}_{\boldsymbol{h}_x}(||\boldsymbol{l}||) := \mathcal{K}_{h_1}(l_1)\mathcal{K}_{h_2}(l_2)\cdots\mathcal{K}_{h_q}(l_q), \tag{4.5}$$

where $\mathcal{K}_{h_j}(l_j) = h_j^{-1} K(\frac{l_j}{h_j})$ for $j = 1, ..., q$.

In addition, the mean of the conditional density estimator in (4.3) provides an estimator of the conditional mean function $m(\boldsymbol{x}) := E(y|\boldsymbol{x})$ as

$$\hat{m}(\boldsymbol{x}) = \int y \hat{f}(y|\boldsymbol{x}) dy. \tag{4.6}$$

Hyndman *et al.* (1996) note that the estimator in (4.6) is equivalent to the Nadaraya-Watson (NW) regression estimator and only depends on $\boldsymbol{h}_x$, the smoothing parameter related to $\boldsymbol{x}$. The Nadaraya-Watson estimator is

$$\hat{m}(\boldsymbol{x}) = \sum_{i=1}^{N} w_i(\boldsymbol{x}) y_i. \tag{4.7}$$

In the rest of the paper, we will use the expression in (4.7) as our mean function estimator.

65

(a) Air density   (b) Turbulence intensity   (c) Wind shear

(d) Air density   (e) Turbulence intensity   (f) Wind shear

(g) Air density   (h) Turbulence intensity   (i) Wind shear
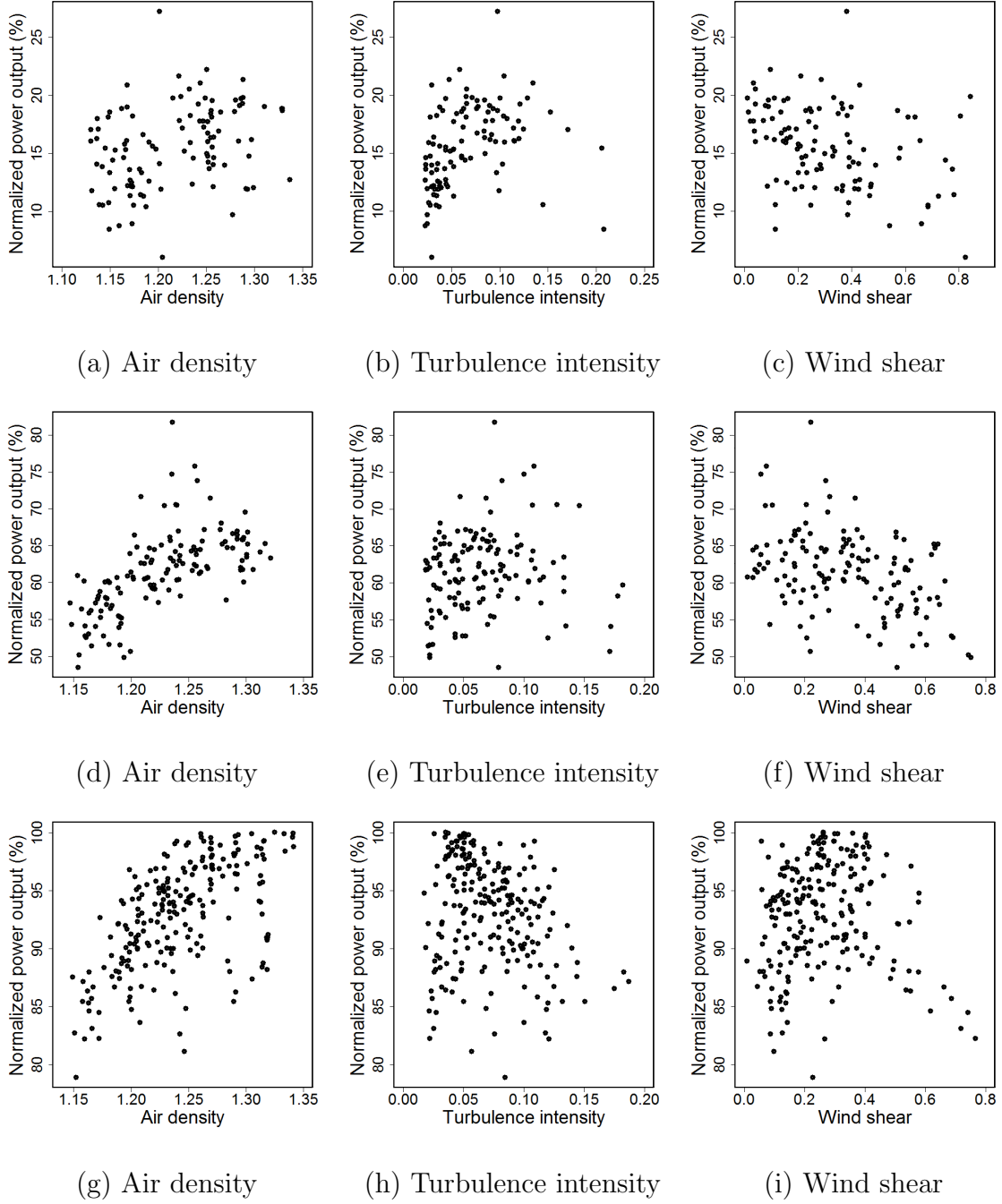
Fig. 16. Scatter plots of the power output versus environmental factors under specific wind speeds and wind directions. Top Panel: $6.1 < v < 6.2, 270 < d < 300$; Middle Panel: $9.1 < v < 9.2, 270 < d < 300$; and Bottom Panel: $11.1 < v < 11.2, 270 < d < 300$.

If all the environmental factors ($q = 7$ for $\mathcal{D}_{\text{OSWF}}$ and $q = 5$ for $\mathcal{D}_{\text{ILWF}}$) are included in the Rosenblastt's CKD or the NW estimator through a product kernel, the resulting model will be very expensive to use computationally. Not only that, the method's performance will not be good, either, because data will easily become sparse in a high-dimensional input space. This is to say that even though we have multiple years of environmental and power data, once they are dispersed into a seven dimensional space, certain combinations of environmental conditions could have very little data or even no data at all. If a technology innovation makes additional measurements available so that the model could entertain more than seven explanatory variables, the current CKD approach will clearly run into scalability problems. In fact, Jeon and Taylor (2012) included only two variables in their CKD. Generally, we believe that a CKD having three inputs can still be practically handled but anything more than that will hardly be practical.

Our solution of addressing the multivariate kernel density estimation problem is to use an additive multivariate kernel. Let us present the mathematical expression of the kernel first and then elaborate its merit. Recall that wind direction d is a circular variable that may cause troubles in numerical computation. For this reason, we convert v and d into two wind velocities vd_1 and vd_2 and will use vd_1 and vd_2 in our model as a way to incorporate the information embedded in v and d. We want to note that Jeon and Taylor (2012) took the same approach.

For notation simplicity, we designate the first two elements of $\boldsymbol{x}$ of both $\mathcal{D}_{\text{ILWF}}$ and $\mathcal{D}_{\text{OSWF}}$, namely x_1 and x_2, as vd_1 and vd_2, respectively. We introduce a new symbol $\boldsymbol{x}^j$, $j = 3, ..., q$, such that $\boldsymbol{x}^j := (x_1, x_2, x_j)$. With this notation, we propose to estimate the density of y, conditional on an $\boldsymbol{x}$, by using

$$\hat{f}(y|\boldsymbol{x}) \;\; = \;\; \sum_{i=1}^{N} \frac{1}{(q-2)} \left[w_i(\boldsymbol{x}^3) + \ldots + w_i(\boldsymbol{x}^q) \right] \mathcal{K}_{h_y}(y - y_i), \qquad (4.8)$$

and the conditional mean function by

$$\hat{m}(\boldsymbol{x}) \;\; = \;\; \frac{1}{(q-2)}\left[\hat{m}(\boldsymbol{x}^3) + \ldots + \hat{m}(\boldsymbol{x}^q)\right]. \tag{4.9}$$

As in the above expression, our resulting model keeps the multivariate kernels but we limit them to be product kernels of three inputs. Based on the observations from Figures 15 and 16, we decide to include vd_1 and vd_2, incorporating wind speed and direction information, in every multivariate kernel so that the three-variable kernel can capture the interaction effect between the third environmental factor with wind speed and wind direction. Then, all the multivariate kernels constitute an additive model such that the resulting model has good scalability. The resulting model can be used for high-dimensional data without causing computational or data sparsity problems. When additional explanatory variables become available, we would envision to add extra additive terms, each of which has the same structure as the current terms, namely a three-variable multivariate kernel having inputs of vd_1, vd_2, and a third explanatory variable.

IV.2.3. Bandwidth selection

The key parameters in our kernel model are the bandwidths h_y and $\boldsymbol{h}_x$. In this study, we employ a data-driven selection criterion proposed by Hall *et al.* (2004) and Fan and Yim (2004), known as the integrated squared error (ISE) criterion, as follows:

$$\begin{aligned}
\text{ISE}(\boldsymbol{h}_x, h_y) \;\; &= \;\; \int\int\left(f(y|\boldsymbol{x}) - \hat{f}(y|\boldsymbol{x})\right)^2 f(\boldsymbol{x})dyd\boldsymbol{x} \tag{4.10}\\
&= \;\; \int\int \hat{f}(y|\boldsymbol{x})^2 f(\boldsymbol{x})dyd\boldsymbol{x} - 2\int\int \hat{f}(y|\boldsymbol{x})f(y|\boldsymbol{x})f(\boldsymbol{x})dyd\boldsymbol{x}\\
&\quad + \int\int f(y|\boldsymbol{x})^2 f(\boldsymbol{x})dyd\boldsymbol{x}\\
&= \;\; I_1 - 2I_2 + I_3.
\end{aligned}$$

With this criterion, one would choose the bandwidths that minimize the ISE. Because I_3 in the ISE expression does not depend on the bandwidth selection, it can be omitted during the minimization of ISE.

Fan and Yim (2004) suggested a cross-validation estimators of I_1 and I_2 as

$$
\begin{aligned}
\hat{I}_1 &= \frac{1}{N} \sum_{i=1}^{N} \int \left(\hat{f}_{-i}(y_i|\boldsymbol{x}_i) \right)^2, \text{ and} \\
\hat{I}_2 &= \frac{1}{N} \sum_{i=1}^{N} \hat{f}_{-i}(y_i|\boldsymbol{x}_i),
\end{aligned}
\tag{4.11}
$$

where $\hat{f}_{-i}(y|\boldsymbol{x}_i)$ is the estimator $\hat{f}(y|\boldsymbol{x}_i)$ with observation i omitted. Practically, the data-driven bandwidth selection is simply to choose the bandwidths $\boldsymbol{h}_x$ and h_y that minimize $\hat{I}_1 - 2\hat{I}_2$.

The above data-driven procedure works well with low-dimensional multivariate kernels (one or two dimensions, for example). Our kernel model is a little more complex because of the mixed additive and product kernels where the product kernels have three variables. Using the above-expressed leave-one-out cross-validation algorithm runs into a serious computational challenge. In this study, we devise two heuristic procedures for selecting the bandwidths, which are described in the following.

1. *Algorithm I*: Simple kernel bandwidth selection

   (a) Only consider a simple univariate kernel regression corresponding to individual environmental variables;

   (b) Calculate the bandwidth for each univariate kernel following the direct plug-in (DPI) approach by Ruppert *et al.* (1995). This DPI approach provides an optimal bandwidth formula, expressed below, that is supposed

to minimize the asymptotically weighted integrated squared error.

$$\hat{h}_{\text{DPI}} = \left(\frac{1}{2\sqrt{\pi}}\right)^{1/5} \left[\frac{\hat{\sigma}^2(b-a)}{N\hat{\theta}_{22}}\right]^{1/5},\tag{4.12}$$

where $[a, b]$ is the range of each environmental variable, and $\hat{\sigma}^2$ and $\hat{\theta}_{22}$ are estimated from the data using the DPI algorithm; for details, please refer to Ruppert *et al.* (1995);

(c) Denote the resulting bandwidths as $(h_1^{SIM}, h_2^{SIM}, \ldots, h_q^{SIM})$;

(d) Use the most basic power curve model that includes only the two wind velocities vd_1 and vd_2 as inputs, and fix the bandwidths for the two univariate kernels corresponding to vd_1 and vd_2 as h_1^{SIM} and h_2^{SIM}, respectively. Then, find the bandwidth h_y^{SIM} that minimizes $\hat{I}_1 - 2\hat{I}_2$.

2. *Algorithm II*: Forward stepwise bandwidth selection

(a) Start with a bivariate kernel model including only wind velocities (vd_1, vd_2);

(b) Find the bandwidths (h_1, h_2, h_y), corresponding to vd_1, vd_2 and y, respectively, by using the leave-one-out cross validation algorithm in (4.12);

(c) Fix the bandwidths (h_1, h_2, h_y) obtained in (b) and let $Q := \{\hat{m}(\boldsymbol{x}^3), \ldots, \hat{m}(\boldsymbol{x}^q)\}$ be the set of additive terms;

(d) Choose from Q one additive term $\hat{m}(\boldsymbol{x}^*)$ and find the corresponding bandwidth h_* that minimizes $\hat{I}_1 - 2\hat{I}_2$;

(e) Let $Q = Q \setminus \hat{m}(\boldsymbol{x}^*)$, and fix all the bandwidths so far determined;

(f) Repeat (d) and (e) until all bandwidths are decided.

Being heuristic, these algorithms cannot guarantee the optimality of the chosen bandwidths. But, as we will show in the subsequent section, our kernel model with these

heuristically chosen bandwidths is able to produce remarkable error reduction as compared with the current power curve methods.

IV.3. Results

In this section, we evaluate the performance of our kernel method using the wind farm measurements in $\mathcal{D}_{\mathrm{ILWF}}$ and $\mathcal{D}_{\mathrm{OSWF}}$ and compare its performance with the existing methods.

IV.3.1. Performance criteria

We evaluate the performance of our method in terms of point estimation as well as density estimation. We therefore use two criteria: for point estimation, we use the root mean square error (RMSE), and for density estimation, we use the mean continuous ranked probability score (CRPS) (Gneiting and Raftery, 2007). We randomly divide each dataset into a partition of 80% for training and 20% for testing, and use the test dataset to empirically evaluate the above two criteria. Specifically, RMSE is computed as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{TS}} \sum_{i=1}^{N_{TS}} (\hat{m}(\boldsymbol{x}_i) - y_i)^2}, \tag{4.13}$$

where N_{TS} is the number of data points in a test dataset. The CRPS compares the estimated cumulative distribution function (CDF) with the observed value. It is computed as

$$\mathrm{CRPS} = \frac{1}{N_{TS}} \sum_{i=1}^{N_{TS}} \int \left(\hat{F}(y|\boldsymbol{x}_i) - y_i \right)^2 dy, \tag{4.14}$$

where $\hat{F}(y|\boldsymbol{x}_i)$ is the estimated CDF, given a setting of the environmental variable $\boldsymbol{x}_i$.

We test both Algorithm I and II for bandwidth selection. It turns out that for the point estimation, Algorithm I gives slightly better selections than those selected by Algorithm II, while for the density estimation, Algorithm II is slightly better. However, the results obtained from the two algorithms are not much different. In the following, we report RMSE values based on bandwidth selection using Algorithm I and CRPS values based on bandwidth selection using Algorithm II. A final note is that to determine the bandwidths, we use only 20% of a training dataset because otherwise the computation time is too long.

IV.3.2. Important environmental factors affecting power output

From the physical understanding presented in Section IV.2.1, we believe that wind speed, direction, and air density should be important factors to be included in a power curve model. The question is what else may also need to be included. This section sets out to find what set of environmental factors makes the best prediction for a given dataset. In the following, we present the results of using point estimates and RMSE values.

Our first set of results is to show the RMSE values when the additive multivariate kernel model include a single additive term from $\boldsymbol{x}^3$ to $\boldsymbol{x}^q$. Recall that each additive term is a three-variable multivariate kernel with two of the variables always being the two wind velocities vd_1 and vd_2.

We choose the baseline model for comparison as the kernel model that has only the two wind velocities (vd_1, vd_2) in a product kernel. In fact, this bivariate kernel (BVK) model is the same as the one used by Jeon and Taylor (2012). The results are shown in Table 8.

Table 8. Impact on RMSE when including different environmental factors. The notation of $(\cdot, \cdot, \rho)$ means that the additive term included in the model has the two wind velocities vd_1, vd_2 and air density ρ as its inputs, where the two wind velocities are shorthanded as two dots. Other notations follow the same convention. The percentages in the parentheses are the reduction in terms of RMSE when the corresponding model's point estimation is compared with that of BVK.

WT	BVK	$(\cdot, \cdot, \rho)$	$(\cdot, \cdot, t_b)$	$(\cdot, \cdot, w_s^b)$	$(\cdot, \cdot, w_s^a)$	$(\cdot, \cdot, h_m)$
WT1	148.2	124.7	144.3	142.2	·	·
		(15.8%)	(2.6%)	(4.0%)	·	·
WT2	146.1	130.4	155.9	169.9	·	·
		(10.8%)	(-6.7%)	(-16.2%)	·	·
WT3	149.8	122.2	135.4	134.7	·	·
		(18.4%)	(9.6%)	(10.1%)	·	·
WT4	199.0	175.3	199.5	191.2	·	·
		(11.9%)	(-0.3%)	(3.9%)	·	·
WT5	279.7	252.6	265.2	266.3	265.0	261.6
		(9.7%)	(5.2%)	(4.8%)	(5.2%)	(6.5%)
WT6	293.7	261.6	309.2	292.3	293.9	276.5
		(10.9%)	(-5.3%)	(0.5%)	(-0.1%)	(5.8%)

Based on the above results, we make the following observations:

- In both the inland wind farm and offshore wind farm, air density ρ is indeed, after the two wind velocities, the most significant factor on wind power generation. Including ρ in the model delivers a reduction in RMSE from nearly 10% to 18% across the board. This outcome is consistent with the physical understanding expressed earlier.

- For the offshore wind turbines, humidity h_m appears to be another important factor for explaining the variation of power outputs. Unfortunately, we will not be able to know for sure whether humidity is also a significant factor in the inland wind farm because its measurements were not available in our dataset. Given its significance in the offshore farm, this should provide strong enough motivation for practitioners to make measurements at some inland wind farms and test the hypothesis.

- For the remaining three factors, namely turbulence intensity and the two wind shears, which all represent some other aspects of wind dynamics, they show mixed outcomes, oftentimes depending on the distance between a turbine and its associated mast. For the turbines next to a mast (WT1, WT3, and WT5), the other factors all play some positive roles in explaining the variation in power outputs, although their impact is much less significant than air density, while for the turbines that have some distance from their mast (WT2, WT4, and WT6), these factors appear to have much less appreciable effect on power outputs. This may not necessarily mean that the other wind dynamics measure does not impact a turbine's power output. What it could mean is that when a turbine is more than 1,000 meters away from a mast, the wind dynamics measured by the mast does not represent the actual wind dynamics in front of

that specific turbine, and consequently, the mast measurements do not have a sufficient explanatory power.

The next step we undertake is to determine which other factors may impact the power output when we include more than one additive term in our model, conditional on the factors that have already been included. Based on the observations expressed above, for both inland and offshore turbines, the first additive term included is always (vd_1, vd_2, ρ). For the inland turbines, in addition to this first term, there are two more terms that have either turbulence intensity, t_b, or the below-hub wind shear, S_b. For the offshore turbines, we also always include a second additive term (vd_1, vd_2, h_m). Then, in addition to the first two terms, there are three more terms that have either the two wind shears, w_s^a, w_s^b, or turbulence intensity, t_b. The two wind shears are always included or excluded together in the numerical analysis to keep the total number of model comparisons manageable. Tables 9 and 10 present the model comparison results.

For some of the inland turbines, the best additive multivariate kernel model explaining their power output includes the input factors of the two wind velocities (vd_1 and vd_2), air density (ρ), and turbulence intensity (t_b), while some others include the two wind velocities (vd_1 and vd_2), air density (ρ), and wind shear (w_s^b), or some others include all the environmental factors. These versions differ very marginally. In the next subsection where the additive multiplicative kernel model is compared with other methods, we choose the model with four factors, vd_1, vd_2, ρ, and t_b, as the "best model" for the inland turbines, because when it does not produce the smallest RMSE, the difference between its RMSE and the smallest RMSE is less than one percentage point.

For the offshore turbines, it is rather clear that the model with the two wind

Table 9. Model comparisons using data in $\mathcal{D}_{\mathrm{ILWF}}$. RMSE values are reported in the table. Bold-face values represent the smallest RSME for its row.

WT	$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,t_b)$	$(\cdot,\cdot,\rho,w_s^b)$	$(\cdot,\cdot,\rho,t_b,w_s^b)$
WT1	124.7	**124.2**	124.3	125.5
WT2	130.4	**128.0**	128.1	128.4
WT3	122.2	118.9	**118.7**	119.0
WT4	175.3	171.3	171.2	**170.4**

Table 10. Model comparisons using data in $\mathcal{D}_{\mathrm{OSWF}}$. RMSE values are reported in the table. Boldface values represent the smallest RSME for its row.

WT	$(\cdot,\cdot,\rho,h_m)$	$(\cdot,\cdot,\rho,h_m,t_b)$	$(\cdot,\cdot,\rho,h_m,w_s^a,w_s^b)$	$(\cdot,\cdot,\rho,h_m,t_b,w_s^a,w_s^b)$
WT5	**248.2**	252.4	254.1	257.2
WT6	**256.7**	264.4	263.8	270.9

velocities (vd_1 and vd_2), air density (ρ), and humidity (h_m) produces the lowest RMSE. Including other environmental factors in the model could increase the RMSE. The increase in RMSE is noticeable and can be as much as 5% for turbine WT6. In the next section, we choose the model with vd_1, vd_2, ρ, and h_m as the "best model" for the offshore turbines.

If we repeat the above analysis using the CRPS measure, the insights remain the

same. We therefore omit the presentation of the detailed CRPS results.

IV.3.3.   Comparison of estimation accuracy of different models

In this subsection, we compare the "best" additive multivariate kernel model, selected in the preceding subsection, with two existing methods: the binning method (BIN), popular in the wind industry and arguably the most widely used method in practice, and the BVK by Jeon and Taylor (2012). Recall that the binning method we use here is the version having incorporated the air density adjustment. To make this explicit, we use the notation BIN_a.

It is not difficult to notice that BIN_a conceptually makes use of two pieces of information to fit the power data: wind speed and air density, while BVK makes also use of two pieces of information: wind speed and wind direction. To the best of our knowledge, no power curve model before ours has made use of all three pieces of information. But one straightforward extension of BVK could possibly accomplish that. This extension is to let the BVK model use the air density adjusted wind speed, instead of the original wind speed, as an input. That way, the air density information is utilized at least in the same fashion as in the binning method. The resulting air density-adjusted extension is labeled as BVK_a.

The comparison results using the inland turbines are included in Table 11. Note that the binning method can produce only point estimations, while BVK and BVK_a produce both point and density estimations. In the RMSE-based comparisons, the baseline model used in the table is BIN_a, and in the CRPS-based comparison, the baseline model is BVK. The percentage values in the parentheses are the reduction in terms of a performance measure when a given method is compared with the corresponding baseline model.

We notice that the BVK model produces a significant improvement over the

Table 11. Comparing RMSE and CRPS using data from $\mathcal{D}_{\text{ILWF}}$

Measure	Turbine	BIN_a	BVK	BVK_a	Additive multivariate kernel	
					$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,t_b)$
RMSE	WT1	218.7	148.2	148.6	124.7	124.2
			(32.3%)	(32.0%)	(43.0%)	(43.2%)
	WT2	194.3	152.2	148.8	130.4	128.0
			(24.8%)	(26.2%)	(32.9%)	(34.1%)
	WT3	209.2	149.8	144.7	122.2	118.9
			(28.4%)	(30.9%)	(41.6%)	(43.2%)
	WT4	267.9	199.0	195.8	175.3	171.3
			(25.7%)	(26.9%)	(34.6%)	(36.1%)
CRPS	WT1	·	99.7	98.6	93.4	92.6
				(1.1%)	(6.3%)	(7.1%)
	WT2	·	100.6	98.9	93.7	92.9
				(1.7%)	(6.9%)	(7.7%)
	WT3	·	104.4	102.5	93.4	92.3
				(1.9%)	(10.5%)	(11.6%)
	WT4	·	134.4	133.5	120.9	119.1
				(0.7%)	(10.1%)	(11.4%)

industry standard binning method, with a reduction of RMSE ranging from 25% to 32%. Our additive multivariate kernel method improves further from BVK another

9% to 15%. In other words, the additive multivariate kernel method can reduce RMSE from the binning method by 34% to 43%. We believe that the improvement is the result of including the additional environmental factors in the model.

Comparing BVK and BVK_a, we find that the straightforward extension to include the air density adjustment does not produce much benefit to reduce errors in wind power prediction. This does not, however, imply that the air density is not important. If comparing BVK and BVK_a with our method with only the $(\cdot, \cdot, \rho)$ term, we see an improvement made by our method, already accounting for the majority portion of the 9% to 15% error reduction as mentioned above. What this suggests is that the additive multivariate kernel method captures the relationship and interactions among the key factors in a better away.

As we have seen from the preceding subsection, there are some benefits from including t_b and w_s^b but the benefit appears to be marginal, at least for the inland wind farm data we have at hand.

Table 12 presents the comparison results for the offshore turbines. The results show a level of improvement consistent with what we have seen in the inland turbines.

Comparing BVK, BVK_a, and our method using the CRPS values, we also see good improvements. Figure 17 presents an illustration of density estimations using BVK and our method. To produce the result in Figure 17, we used WT5's data and the additive multivariate kernel model with $(\cdot, \cdot, \rho, h_m)$. The left panel of Figure 17 shows the predictive distributions of the power output from the two models, when their CRPS values are not much different. The two distributions are similar and either model produces a good estimate. The right panel presents the predictive distributions of the two models, when their CRPS values differ considerably. We can see that the distribution from the BVK model is centered incorrectly. On average, the additive multivariate kernel model produces a CRPS that is 7% to 12% better than the BVK

Table 12. Comparing RMSE and CRPS using data from the $\mathcal{D}_{\mathrm{OSWF}}$

Measure	Turbine	BIN_a	BVK	BVK_a	Additive multivariate kernel	
					$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,h_m)$
RMSE	WT5	332.1	279.7	278.1	252.6	248.2
			(15.8%)	(16.2%)	(23.9%)	(25.6%)
	WT6	376.0	293.7	292.2	261.6	256.7
			(21.9%)	(22.3%)	(30.4%)	(31.7%)
CRPS	WT5	·	154.3	153.6	135.9	135.4
				(0.4%)	(4.3%)	(9.7%)
	WT6	·	160.3	159.7	153.3	144.7
				(0.5%)	(11.9%)	(12.3%)

model.

IV.4.   Discussion

This study presents an additive multivariate kernel method for modeling power curves with a variety of environmental factors. It is an appealing approach because this new power curve model can capture the nonlinear relationships between environmental factors and the wind power output, as well as the high-order interaction effects among some of the environmental factors. As compared with existing approaches that consider only wind speed and direction, the new power curve model can make further reductions in terms of prediction errors for both mean estimation and density estimation. The resulting method has good scalability and can easily include
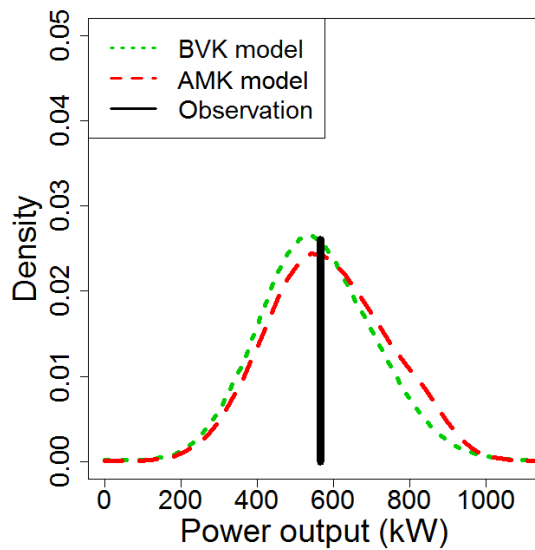
(a) Two models produce similar results
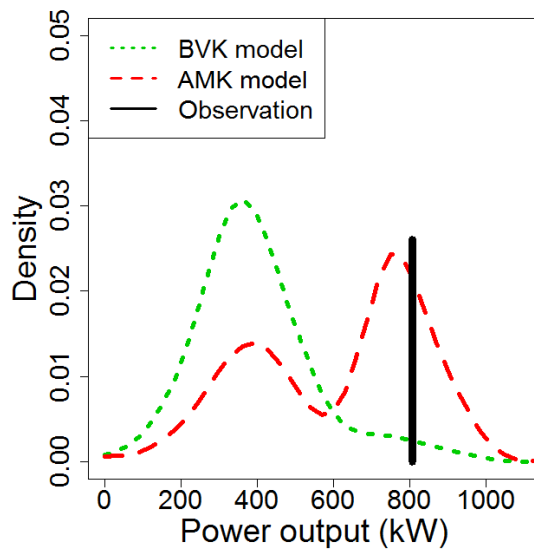
(b) Two models produce different results

Fig. 17. Comparison of the predictive distributions of power output when the two models produce similar CRPS values versus when the two models produce different CRPS values. BVK: Bivariate kernel, AMK: Additive multivariate kernel.

extra explanatory variables should technology innovation make their measurements available.

CHAPTER V

TURBINE PERFORMANCE ASSESSMENT BY USING THE POWER CURVE

This chapter presents our work on turbine performance assessment, which is an application of the power curve estimation with multivariate environmental factors presented in Chapter IV. Turbine performance assessment helps wind farm operators or manufacturers make informed decisions. For example, in the wind industry, people sometimes retrospectively add an additional device or adjust certain operation logic to improve the power production efficiency of the turbines; examples of such would include adjusting the pitch angle, installing vortex generators, and increasing the rated power. However, such upgrades are generally quite expensive, and this motivates the wind farm operators to decide whether or not there are enough benefits to justify such action. Another line of potential applications is on turbine fault diagnosis. To reduce O&M cost caused by unexpected failures, the wind farm operators are interested in preventive maintenance or conditioned-based maintenance (Chattopadhyay, 2004, Zhang *et al.*, 2009). In doing so, information about the health of wind turbines is needed for fault diagnosis. For that purpose, the power curve can serve as an indicator of a wind turbine's health.

In the remainder of this chapter, we first describe some background information regarding wind turbine upgrades, faults, and the datasets used in this study. We modify the additive multivariate model for turbine performance assessment in Section V.2. In Section V.3, the details of our testing procedure is presented. In Section V.4, our procedure is verified and tested by using three different cases, which include a case of vortex generator installation, a case of pitch angle adjustment, and a

case of fault occurrence. Finally, we end this chapter with a discussion in Section V.5.

V.1. Background and datasets

We only study inland wind turbines (ILTs) in this work, and use the datasets obtained from the same inland wind farm (ILWF) described in Chapter IV. For the characteristics of wind turbines used in this study, please refer to Table 7. For turbine performance assessment, we choose three sets of wind turbines, each set consisting of two wind turbine pairs; one is an experimental wind turbine (EWT) and the other is a control wind turbine (CWT) in the vicinity of the EWT.
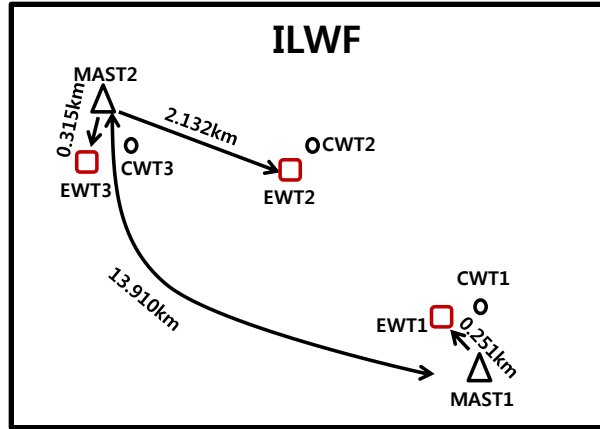

Fig. 18. Layout of the turbines and masts and turbine-to-mast distances: EWTs and CWTs

Figure 18 shows the turbines/masts layout and EWT-to-mast distance. For EWT1 and CWT1, the data were collected from April 29, 2010 through October 30, 2011; for all the other wind turbines, data were collected from June 10, 2010 through

October 31, 2011. The wind farm operator installed vortex generators on the blades of EWT1 on June 19, 2011. Since data are not available from those with pitch angle adjustment and faults, we generate synthetic data based on data measured on EWT2 and EWT3, and physical knowledge regarding pitch angle adjustment and faults. The detailed explanation on generating synthetic data can be found in Section V.4.

Listed below are explanations of the background information regarding upgrades and faults employed in this study:

- Vortex generator installation: Figure 19 shows an example of a blade equipped with vortex generators. Vortex generators are often used in the wind energy industry in order to improve the performance of blades. They enhance the mixing of the outer flow at the blade's boundary layer. However, fitting a wind turbine blade with the right set of vortex generators is not a simple task since each wind turbine blade design requires different vortex generator designs and positioning in order to achieve optimal performance. Figure 20 presents an empirical study comparing power curves of 1000kW wind turbine before and after installing vortex generators.

- Pitch angle adjustment: One of the design parameters C_p in (4.1) depends on the blade pitch angle β. Understandably, the pitch angle is an important variable in maximizing power outputs captured from the wind. Wind farm operators modify pitch angles. Figure 21 depicts the rotor power performances corresponding to different pitch angels. Here, the power performances are calculated by using a GHBladed aerodynamic model (Bossanyi, 2003).

- Faults: Many studies already show that degradation in power generation performance is a good indication of problems such as blade faults, yaw system faults, pitch system faults, and gear or bearing faults (Gill *et al.*, 2012, Kusiak

85

et al., 2009, Uluyol *et al.*, 2011). Figure 22 shows the abnormal status of a wind turbine detected by a reference power curve. The reference power curve is built with the *k*-nearest neighbor (*k*-NN) algorithm and the residual mean control chart.
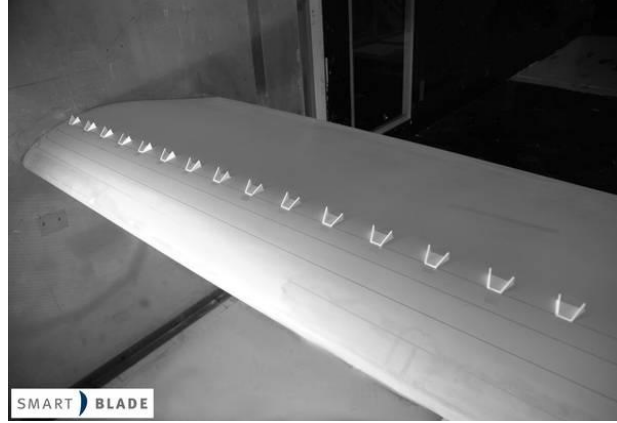
Fig. 19. Blade equipped with vortex generators [Source: Smart Blade (2013)]

On real wind farms, it is difficult to quantify performance changes caused by the aforementioned modifications or by faults on a single wind turbine because their effect is dependent on the type of turbine and the site. Plus, the power outputs have many variations resulting from environmental factors in addition to the wind speed. But the power curve method described in Chapter IV can help the effort of turbine performance assessment here.

V.2. Kernel plus method

For assessing turbine performance, we consider the power residual r in (1.4), which is the difference between the observed power output y and the predicted power out-
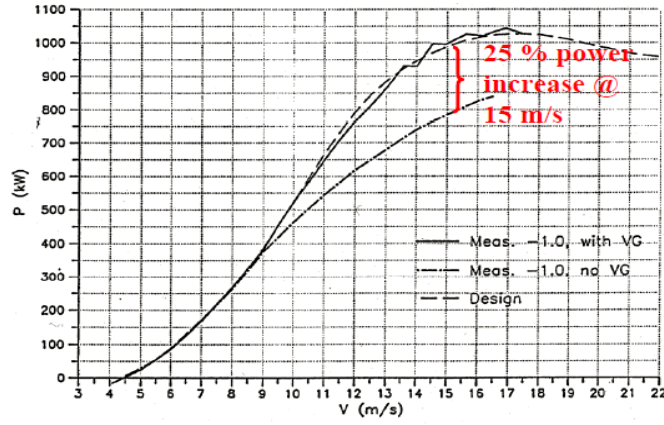
Fig. 20. Example of the application of vortex generators on wind turbine blades: ELKRAFT 1000kW wind turbine at Averdore [Source: Øye (1995)]
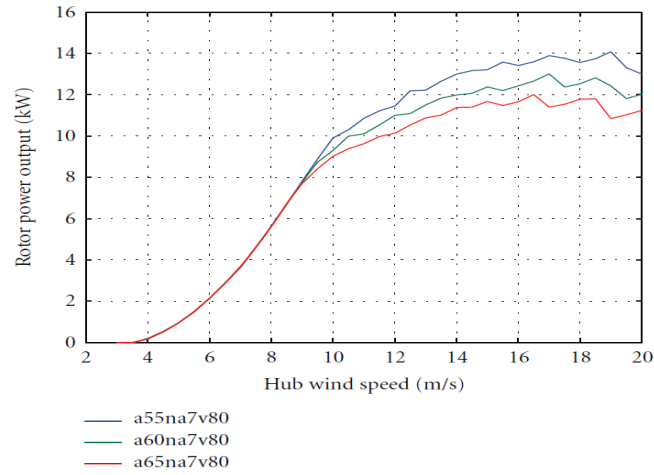
Fig. 21. Power output simulation based on different pitch angles [Source: Wang *et al.* (2012)]: for the expression 'a**na7v80', 'a' represents attack angle, a55 means attach angle of 5.5^o; na7 means design tip speed ratio of 7; v80 means design wind speed of 8.0 m/s.
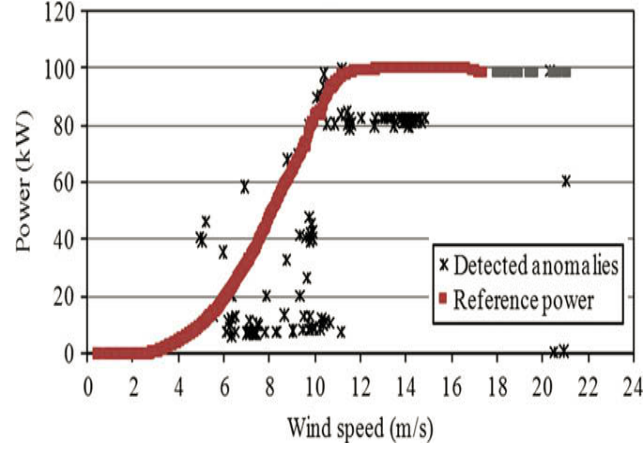
Fig. 22. Anomalies detected by the reference power curve [Source: Kusiak *et al.* (2009)]

put $\hat{y}$. In this study, for calculating the power residual, the binning method and a modified additive multivariate kernel model are both employed. Let y^{BIN} and y^{AMK} denote the binning method estimator and the additive multivariate kernel estimator, respectively. Furthermore, the power residuals calculated by y^{BIN} and y^{AMK} are referred to as r^{BIN} and r^{AMK}. More specifically, as mentioned in Chapter II, the binning method is to discretize the domain of wind speed with a bin width of 0.5 m/s. In the additive multivariate kernel model, we consider four explanatory variables, namely, $\boldsymbol{x} = (vd_1, vd_2, \rho, t_b)$ based on the results in Table 9. As a result, our additive multivariate kernel estimator can be expressed as follows:

$$\hat{y}^{AMK}(vd_1, vd_2, \rho, t_b) = \frac{1}{2}\left[\hat{m}(vd_1, vd_2, \rho) + \hat{m}((vd_1, vd_2, t_b)\right], \qquad (5.1)$$

where the bandwidth $\boldsymbol{h_x}$ is selected by employing Algorithm I presented in Section IV.2.3.

Figure 23 illustrates the variations in the two power residuals r^{BIN} and r^{AMK}

with respect to wind speed. For this figure, we use the data obtained from CWT1 and CWT2; for training sets and testing sets, data were collected from August 2010 through July 2011 and from August 2010 through October 2011, respectively. We observe three things from this figure:

- Notice that all residual plots are divided into a below and above the rated wind speed region. The above regions in all plots show a small variation and random pattern, while in the below regions, we observe that all residual plots have large variations. The variation in Figure 23(b) is a little smaller than those of the other three plots. We believe that the below-rated region residual plots contain more useful information.

- Please note that CWT1 is located close to MAST1 and CWT2 is relatively distant from MAST2. In Figures 23(c) and (d), the long distance between CWT2 and MAST2 appears to lead a high variance due to wake effects. Please also note that the IEC recommends that the distance between the wind turbine and the meteorological mast shall be $2.5 \times$ rotor diameter (IEC, 2005b). The distance between CWT1 and MAST1 is therefore more preferred.

- Figures 23(a) and (b) show that there are apparent differences between the patterns of r^{BIN} and r^{AMK}. r^{AMK} appears random, while r^{BIN} forms a non-random bird-like shape. The non-random pattern indicates that the sole explanatory variable in the binning, namely the wind speed, is not capable of capturing all the systematic information in the power output data.

- In Figures 23(b) and (d), the horizontal dotted line indicates a residual value of 0. r^{AMK} is not symmetric with respect to the horizontal dotted line. This shows that the additive multivariate kernel estimator based on the Nadaraya-Waston

kernel method has bias.

The occurrence of bias is problematic because it can negatively affect the results of the turbine performance test. To address the bias issue of the Nadaraya-Watson kernel estimator, we propose a self-calibration procedure, which we call the *kernel plus method.* Before the procedure is presented, a few terminologies must be explained.

To quantify changes in a turbine's performance, three datasets are used: The training dataset, denoted by $\mathcal{D}_{\mathrm{TR}}$, is used to fit a power curve. Notice that a whole year's worth of historical data are preferred for training the power curve model. The two testing datasets, denoted by $\mathcal{D}_{\mathrm{BF}}$ and $\mathcal{D}_{\mathrm{AF}}$, were collected for the same length of time before and after an upgrade or a fault occurrence, respectively. They are used to detect and quantify changes in a turbine's performance caused by an upgrade or a fault occurrence. In addition, let r_{BF} and r_{AF} denote power residuals calculated from these two groups, respectively.

The self calibration procedure is done primarily by using subsets of the training data in $\mathcal{D}_{\mathrm{TR}}$. To select a calibration set of data that has similar weather conditions to those in $\mathcal{D}_{\mathrm{BF}}$ and $\mathcal{D}_{\mathrm{AF}}$, we define a distance measure, which is in spirit similar to the Mahalanobis distance (Mahalanobis, 1936), a measure weighted distance between two multidimensional points based on the corresponding covariance matrix. Our defined distance measure is still a weighted distance but instead of using the covariance matrix, we employ a diagonal matrix whose diagonal elements are from the bandwidth vector $\boldsymbol{h_x}$. Specifically, we denote this diagonal matrix as $\boldsymbol{H}$, so that $\boldsymbol{H}_{i,i} = h_i$ and $\boldsymbol{H}_{i,j} = 0 \quad \forall i \neq j$. The distance of a training data point $\boldsymbol{x}^i \in \mathcal{D}_{\mathrm{TR}}$ and a testing data point $\boldsymbol{x}^j$ in either $\mathcal{D}_{\mathrm{BF}}$ or $\mathcal{D}_{\mathrm{AF}}$ is

$$D(\boldsymbol{x}^i, \boldsymbol{x}^j) = \sqrt{(\boldsymbol{x}^i - \boldsymbol{x}^j)^T \boldsymbol{H}^{-1}(\boldsymbol{x}^i - \boldsymbol{x}^j)}. \tag{5.2}$$
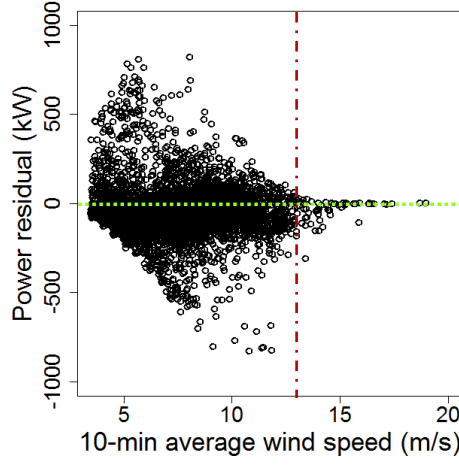
(a) CWT1: BIN
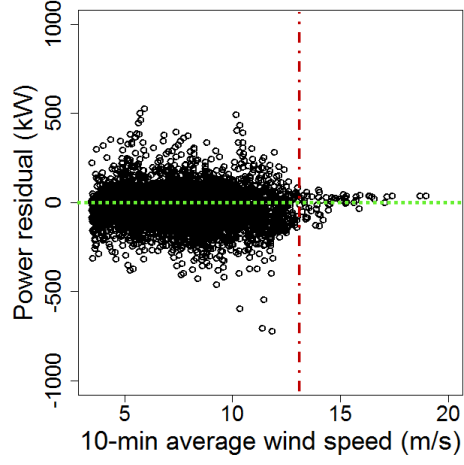
(b) CWT1: AMK
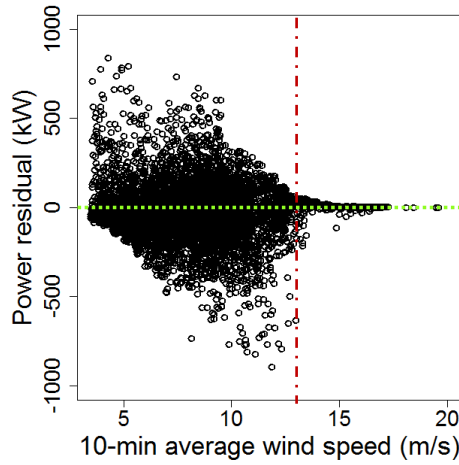
(c) CWT2: BIN
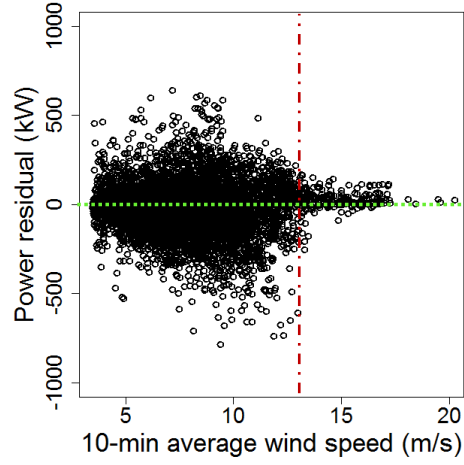
(d) CWT2: AMK

Fig. 23. Scatter plots of the power residual (kW) versus 10-min average wind speed (m/s) for CWT1 and CWT2. Left Panel: BIN is the residual plot calculated by the binning method; Right Panel: AMK is the residual plot calculated by the additive multivariate kernel method. Vertical dashed lines and horizontal dotted lines indicate the rated wind speed (13 m/s) and a residual value of 0, respectively.

The reason we choose this distance measure to define the similarity between $\boldsymbol{x}$'s is as follows. A simple Euclidean distance does not reflect the similarity between $\boldsymbol{x}$'s well because different elements in $\boldsymbol{x}$ have different physical units, leading to different value ranges. To define a sensible similarity measure, a key issue is to weight different elements in $\boldsymbol{x}$ consistently with their relative importance pertinent to the power output. The original Mahalanobis distance does not serve this purpose because the squared distance associated with an input variable is weighted by the inverse of its variance. In a power curve model, wind speed is arguably the most important variable, yet it has a large variance in the same time. Because of this large variance, using the Mahalanobis distance will in fact diminish the importance of wind speed in a relative sense to other variables which have a smaller variance. Our choice of using the kernel bandwidth parameter as the weighting coefficients in $\boldsymbol{H}$ is consistent with our goal of weighting each element according to their relative importance. Recall that the bandwidth is selected based on how sensitive the power output changes in a unit change in the corresponding input variable. If an input variable has a small bandwidth, it means that the power output could produce an appreciable difference with a small change in the corresponding input, suggesting that this variable is relatively important, while on the other hand, a large bandwidth indicates a less important input variable.

For any testing data point $\boldsymbol{x}^j$, we can choose a calibration point $\boldsymbol{x}_{\mathrm{cal}}^j$ from $\mathcal{D}_{\mathrm{TR}}$, which has the minimum $D(\boldsymbol{x}^i, \boldsymbol{x}^j)$ distance. Then, we employ the self calibration procedure as follows:

- For $\boldsymbol{x}_{\mathrm{cal}}^j \in \mathcal{D}_{\mathrm{TR}}$, compute $\hat{y}^{AMK}(\boldsymbol{x}_{\mathrm{cal}}^j)$

- Compute the calibration value $r_{\mathrm{cal}}^j = y(\boldsymbol{x}_{\mathrm{cal}}^j) - \hat{y}^{AMK}(\boldsymbol{x}_{\mathrm{cal}}^j)$

- For $\boldsymbol{x}^j$, the final power estimate from the kernel plus method is $\hat{y}^{KP}(\boldsymbol{x}^j) =$

$\hat{y}^{AMK}(\boldsymbol{x}^j) + r^j_{\text{cal}}$; the power residual is $r^{KP}(\boldsymbol{x}^j) = y(\boldsymbol{x}^j) - \hat{y}^{KP}(\boldsymbol{x}^j)$.

After the self calibration, the bias resulting from a kernel estimator is greatly alleviated but not completely eliminated.

V.3. Performance test procedure

Figure 24 outlines the procedure for detecting and quantifying changes caused by an upgrade or a fault occurrence using the kernel plus method and the three sets of data as described in the previous section.
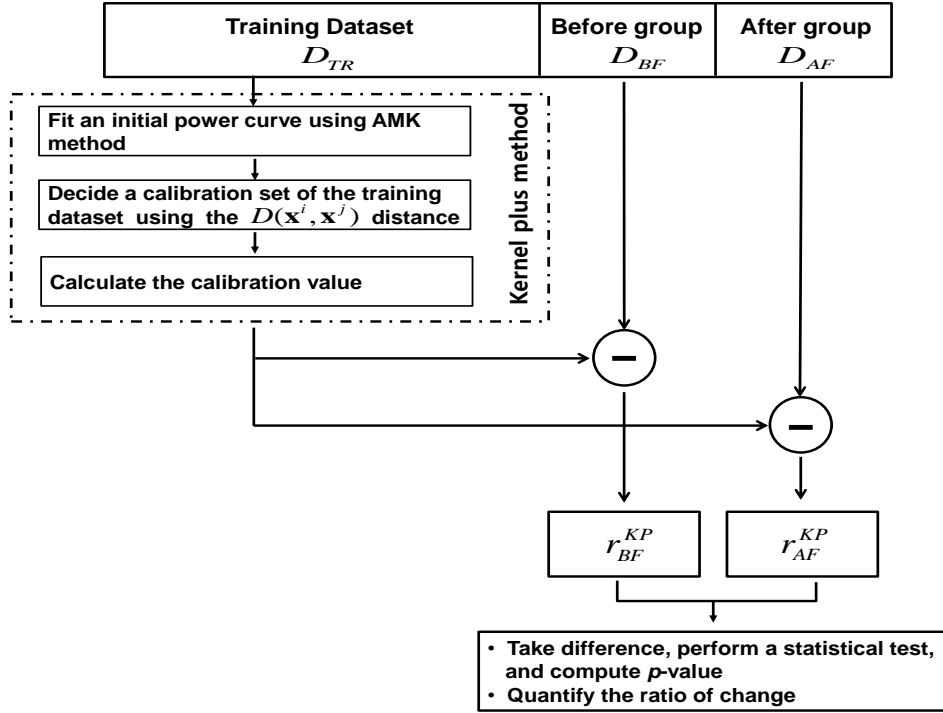

Fig. 24. Overview of the proposed turbine performance testing procedure

We first establish a power curve using the kernel plus method. This step includes

93

the use of both the additive multivariate method and the self calibration. Then, this power curve model is used to make a prediction/estimation of power output under a new weather profile $\boldsymbol{x}$ in either $\mathcal{D}_{\mathrm{BF}}$ or $\mathcal{D}_{\mathrm{AF}}$. Consequently, the corresponding power residuals can be computed, and the result is denoted as r_{BF}^{KP} and r_{AF}^{KP} for $\mathcal{D}_{\mathrm{BF}}$ and $\mathcal{D}_{\mathrm{AF}}$, respectively.

Had a turbine undergone an upgrade or a fault, we would expect the power residuals r_{BF} and r_{AF} to be different. To detect a potential difference in the power residuals, it is necessary to invoke a statistical test because the power output data are noisy; specifically a Student's t-test (Snedecor and Cochran, 1989) is used. Suppose that N_{BF} and N_{AF} are the number of data points in $\mathcal{D}_{\mathrm{BF}}$ and $\mathcal{D}_{\mathrm{AF}}$, respectively. The statistical test procedure is as follows:

- Compute the power residuals before and after an upgrade r_{BF} and r_{AF};

- Compute the two sample means and the corresponding standard deviations by using the following formula:

$$
\begin{aligned}
\bar{r}_{\mathrm{BF}} &= \frac{\sum_{j=1}^{N_{\mathrm{BF}}} r(\boldsymbol{x}^j)}{N_{\mathrm{BF}}} \text{ and } S_{\mathrm{BF}} = \sqrt{\frac{\sum_{j=1}^{N_{\mathrm{BF}}} (r(\boldsymbol{x}^j) - \bar{r}_{\mathrm{BF}})^2}{N_{\mathrm{BF}} - 1}} \quad \boldsymbol{x}^j \in \mathcal{D}_{\mathrm{BF}} \\
\bar{r}_{\mathrm{AF}} &= \frac{\sum_{j=1}^{N_{\mathrm{AF}}} r(\boldsymbol{x}^j)}{N_{\mathrm{AF}}} \text{ and } S_{\mathrm{AF}} = \sqrt{\frac{\sum_{j=1}^{N_{\mathrm{AF}}} (r(\boldsymbol{x}^j) - \bar{r}_{\mathrm{AF}})^2}{N_{\mathrm{AF}} - 1}} \quad \boldsymbol{x}^j \in \mathcal{D}_{\mathrm{AF}}
\end{aligned}
$$

$$(5.3)$$

- Then, the pooled estimated of standard deviation σ_r is calculated:

$$
\sigma_r = \sqrt{\frac{(N_{\mathrm{BF}} - 1)S_{\mathrm{BF}}^2 + (N_{\mathrm{AF}} - 1)S_{\mathrm{AF}}^2}{N_{\mathrm{BF}} + N_{\mathrm{AF}} - 2}}
$$

- The t statistic is calculated:

$$
t = \frac{\bar{r}_{\mathrm{BF}} - \bar{r}_{\mathrm{AF}}}{\sigma_r \cdot \sqrt{\frac{1}{N_{\mathrm{BF}}} + \frac{1}{N_{\mathrm{AF}}}}}
$$

- Finally, calculate the p-value of the t statistic. A p-value is a probability, taking values between 0 and 1. The smaller a p-value is, the more significant the difference is.

The above procedure is devised to confirm a detectable difference resulting from an upgrade or a fault. The output is binary: either the upgrade produces a statistically significant difference in a turbine's performance or it does not. For a EWT, a method is supposed to produce a small p-value to indicate a significant difference when comparing the residuals, while for a CWT, a large p-value is expected. In the statistical inference, the threshold of a p-value to declare significance is sometimes 0.01; we use the 0.01 threshold in this analysis.

Another practical question is that if the t-test above does declare a significant difference, how much a difference in terms of power generation an upgrade or a fault produces? To find an increment rate, we define a quantifier as follows:

$$\text{DIFF}(\boldsymbol{x}) = \frac{\sum_{\boldsymbol{x}^j \in \mathcal{D}_{test}} \left(y(\boldsymbol{x}^j) - \hat{y}(\boldsymbol{x}^j) \right)}{\sum_{\boldsymbol{x}^j \in \mathcal{D}_{test}} y(\boldsymbol{x}^j)} \times 100\%, \tag{5.4}$$

where $\mathcal{D}_{test}$ is a test dataset and can be either $\mathcal{D}_{\text{BF}}$ or $\mathcal{D}_{\text{AF}}$, so that $\text{DIFF}(\boldsymbol{x})$ can be either DIFF_{BF} or DIFF_{AF} accordingly. Similar to the residual analysis described above, comparing DIFF_{BF} with DIFF_{AF} quantifies the increment rate. Denote further that $\text{DIFF} = \text{DIFF}_{\text{AF}} - \text{DIFF}_{\text{BF}}$, which quantifies the difference in power generation between before and after an upgrade or a fault occurrence when effects of weather profile are controlled for. The positive DIFF refers to an upgraded wind turbine that produces more power output under the same weather profile; the negative DIFF refers to the occurrence of a fault on a wind turbine.

V.4.   Test cases

We employ our procedure of turbine performance assessment for three different case studies: a case of vortex generator installation, a case of pitch angle adjustment, and a case of fault occurrence. The case study of vortex generator installation uses the datasets measured at EWT1 and CWT1. The second and third case studies use synthetic datasets generated by following the characteristics of real data observed from other wind turbines.

V.4.1.   Vortex generator installation

To test whether or not vortex generator installation changes power generation performance, we employ the testing procedure presented in Figure 24. Here, the training data $\mathcal{D}_{\mathrm{TR}}$ were collected from April 29, 2010 through June 03, 2011. The before-upgrade data $\mathcal{D}_{\mathrm{BF}}$ were collected from June 4, 2011 through June 18, 2011, and the after-upgrade data $\mathcal{D}_{\mathrm{AF}}$ were collected from June 20, 2011 through June 4, 2011. The vortex generators were installed on the blades of EWT1 on June 19, 2011.

Figure 25 presents the scatter plots between the power output y and the average wind speed v of EWT1 and CWT1. In this figure, blue circles and red triangles represent the data in $\mathcal{D}_{\mathrm{BF}}$ and $\mathcal{D}_{\mathrm{AF}}$, respectively. We cannot observe any obvious changes caused by wind turbine modification in these speed-versus-power plots.

Table 13 presents the t-test results based on residuals r^{BIN} and r^{KP}. If the power curve fitting model is perfectly unbiased, then $\bar{r}_{\mathrm{BF}}$ can be expected to have a zero average. Furthermore, if the upgrade is successful, then the mean of $\bar{r}_{\mathrm{AF}}$ would be a positive value. However, the third column in Table 13 shows that both the binning method and the kernel plus method do not guarantee zero averages for $\bar{r}_{\mathrm{BF}}$. For this reason, instead of using the one-sample t-test, we suggest using the two-sample

96

(a) EWT1   (b) CWT1

Fig. 25. Scatter plots of power output versus 10-minute average wind speed: blue circles and red triangles represent, respectively, data points in $\mathcal{D}_{\mathrm{BF}}$ and $\mathcal{D}_{\mathrm{AF}}$

t-test, which is to compare the residuals obtained from before and after an upgrade. The last two columns in Table 13 are the t-statistics and p-value calculated by using r_{AF} and r_{BF}. For EWT1, the binning method cannot detect any difference, while the kernel plus method can detect a statistically significant difference between the residuals associated with data in $\mathcal{D}_{\mathrm{AF}}$ and $\mathcal{D}_{\mathrm{BF}}$ for EWT1. When we quantify the differences of the power output by the vortex generator installation using (5.4), the power output appears to have increased by around 2.04%.

V.4.2. Pitch angle adjustment

We simulate data for the case of pitch angle adjustment based on Figure 21. Figure 21 illustrates that pitch angle adjustment increases turbine performance at wind speeds above 9 m/s. We assume that the pitch angle of EWT2 was adjusted on October 17, 2011. The before-upgrade data $\mathcal{D}_{\mathrm{BF}}$ were collected from October 2, 2011 through

Table 13. Comparing t-test results between two groups: before and after vortex generator installation

Model	Turbine	$\bar{r}_{\mathrm{BF}}$	$\bar{r}_{\mathrm{AF}}$	$\bar{r}_{\mathrm{AF}} - \bar{r}_{\mathrm{BF}}$	t-statistics	p-value
BIN	EWT1	-5.54	-8.48	-2.94	-0.55	0.5816
	CWT1	-4.86	-21.19	-16.33	-2.57	0.0104
KP	EWT1	-0.08	12.45	12.53	2.97	0.003
	CWT1	6.60	11.61	5.01	1.07	0.2836

October 16, 2011, and the after-upgrade data $\mathcal{D}_{\mathrm{AF}}$ were collected from October 17, 2011 through October 31, 2011. For the after-upgrade data $\mathcal{D}_{\mathrm{AF}}$, we multiply the original power outputs by a factor of 1.03 for those corresponding to wind speeds above 9 m/s. Denote by $\mathcal{D}_{\mathrm{AF}}^{S1}$ the resulting synthetic dataset. Here, the real increment rate for data in $\mathcal{D}_{\mathrm{AF}}$ is not 3% because only some of power outputs y are increased by less than 3% (for example, under wind speeds from 4 m/s to 8 m/s, there is almost no increase in the power output). Thus, we calculate the real increment rate directly based on the differences between original and simulated power outputs, and the result is 2.19%.

Figure 26 presents scatter plots between the power output y and the average wind speed v of EWT2 and CWT2. In Figure 26(a), we observe that rated power is changed, while improvements in power efficiency corresponding to rated wind speed below (9 m/s $-$ 13 m/s) are not obvious.

Table 14 shows the performance testing results of EWT2 and CWT2. The findings from this case study appear to be consistent with that from the vortex generator installation: the kernel plus method detects the change successfully, while the binning
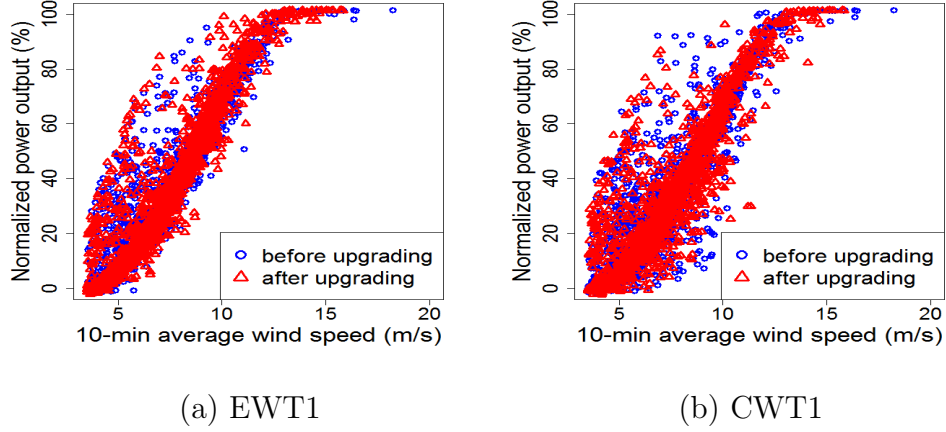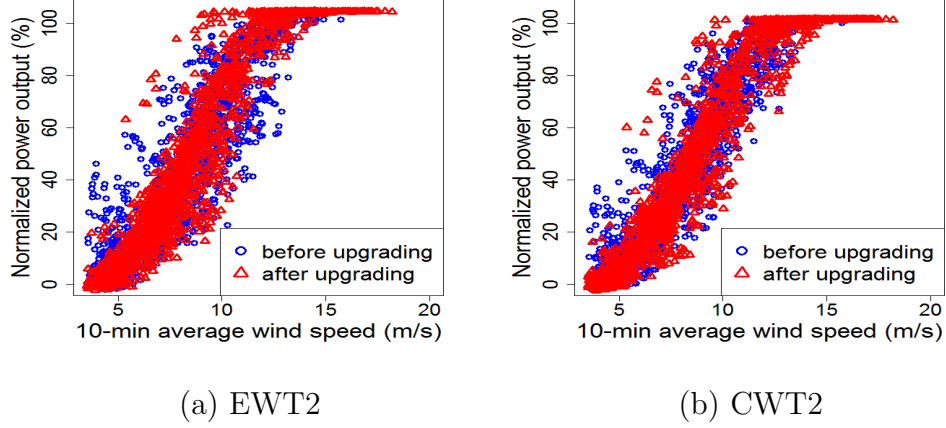
(a) EWT2   (b) CWT2

Fig. 26. Scatter plots of power output versus 10-minute average wind speed: for EWT2, blue circles and red triangles represent, respectively, data points corresponding to $\mathcal{D}_{\mathrm{BF}}$ and $\mathcal{D}_{\mathrm{AF}}^{S1}$; for CWT2, blue circles and red triangles represent, respectively, data points corresponding to $\mathcal{D}_{\mathrm{BF}}$ and $\mathcal{D}_{\mathrm{AF}}$

method does not. If we increase the multiplication factor from 1.03 to a bigger value when simulating the data for pitch angle adjustment, the binning method could also detect a more pronounced change. But we believe that a moderate change as simulated here is more realistic, and a more sensitive method like the kernel plus should be useful.

The average power outputs by our modification seems to be increased by around 2.50%, which is not much different from the real increment rate of 2.19%.

V.4.3. Fault detection

For fault detection, we simulate two datasets based on original testing dataset $\mathcal{D}_{\mathrm{AF}}$ of EWT3 by stipulating that a fault took place on October 1, 2011. The first simulated dataset, denoted by $\mathcal{D}_{\mathrm{AF}}^{S2}$, contains serious faults, which decreases turbine performance

Table 14. Comparing t-test results between two groups: before and
after pitch angle adjustment

Model	Turbine	$\bar{r}_{\text{BF}}$	$\bar{r}_{\text{AF}}$	$\bar{r}_{\text{AF}} - \bar{r}_{\text{BF}}$	t-statistics	p-value
BIN	EWT2	-16.39	-17.13	-0.74	-0.13	0.8949
	CWT2	-22.14	-24.86	-2.42	-0.43	0.6696
KP	EWT2	-11.46	5.95	17.41	2.91	0.0040
	CWT2	-6.25	-8.56	-2.31	-0.40	0.6926

by 50%, and the second dataset, denoted by $\mathcal{D}_{\text{AF}}^{S3}$, has soft faults, which lowers turbine performance by only 5%. Figure 27 and 28 show power curves and power residuals using the simulated datasets. As seen in Figure 27(a), the serious faults are not difficult to detect by visual inspection. When employing the existing outlier detection methods (Kusiak *et al.*, 2009, Osadciw *et al.*, 2010, Yan *et al.*, 2009), the first fault can easily be detected. The existing methods, however, become ineffective in detection of the soft faults of much smaller magnitude.

Table 15 presents the performance testing results using the kernel plus model and the binning model. The kernel plus method appears to be effective for detecting the soft faults, while the binning method is not. The average power outputs in the two simulated datasets $\mathcal{D}_{\text{AF}}^{S2}$ and $\mathcal{D}_{\text{AF}}^{S3}$ appear to be decreased by around -50.51% and -6.2%, respectively. This is very similar to -50% and -5%, which are used for simulating the faults.

(a) Serious faults $\mathcal{D}_{\mathrm{AF}}^{S2}$ (b) Soft faults $\mathcal{D}_{\mathrm{AF}}^{S3}$
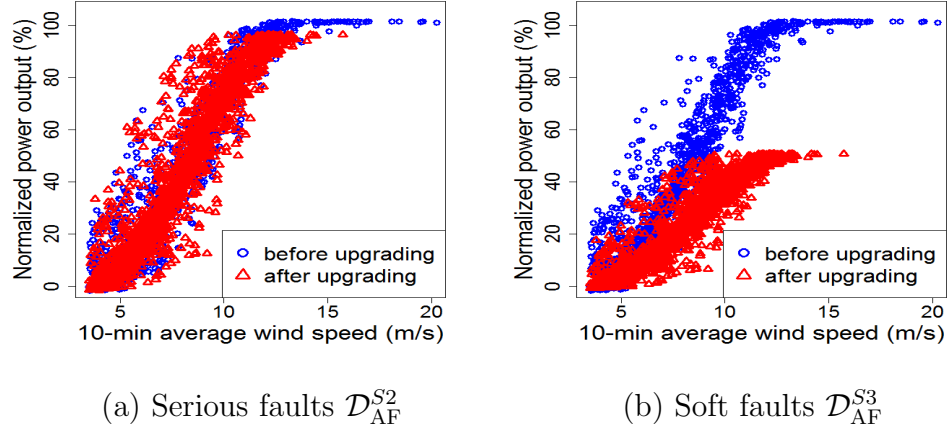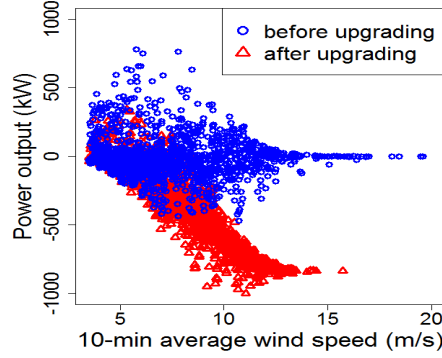
Fig. 27. Scatter plots of power output versus 10-minute average wind speed: blue circles and red triangles represent, respectively, data points corresponding to $\mathcal{D}_{\mathrm{BF}}$ and $\mathcal{D}_{\mathrm{AF}}^{S2}$ or $\mathcal{D}_{\mathrm{AF}}^{S3}$

Table 15. Comparing t-test results between two groups: before and after fault occurrence

Model	Turbine	$\bar{r}_{\mathrm{BF}}$	$\bar{r}_{\mathrm{AF}}$	$\bar{r}_{\mathrm{AF}} - \bar{r}_{\mathrm{BF}}$	t-statistics	p-value
	$EWT3^{S1}$	-14.36	-317.40	-303.05	-40.60	0.2200×10^{-15}
BIN	$EWT3^{S2}$	-14.36	-19.06	-4.70	-0.85	0.3943
	CWT3	-35.23	-6.64	28.59	3.99	0.6639×10^{-4}
	$EWT3^{S1}$	-1.65	-342.48	-340.83	-45.00	0.2200×10^{-15}
KP	$EWT3^{S2}$	-1.65	-44.17	-42.52	-8.98	0.2200×10^{-15}
	CWT3	-8.43	-1.39	7.04	1.11	0.2686

(a) BIN:50% reduction

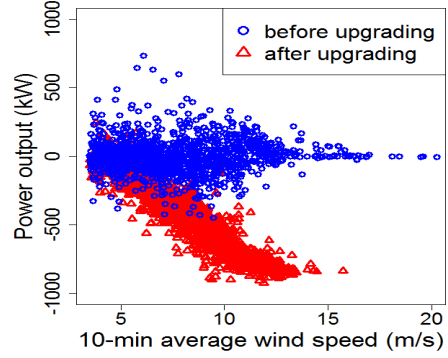(b) KP: 50% reduction
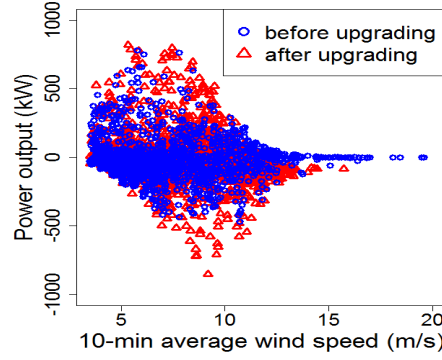
(c) BIN: 5% reduction
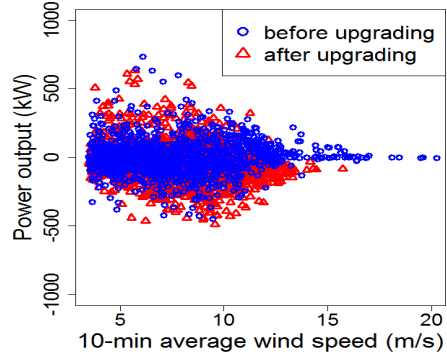
(d) KP: 5% reduction

Fig. 28. Scatter plots of power output versus 10-minute average wind speed: blue circles and red triangles represent, respectively, power residuals r_{BF} and r_{AF} calculated from $\mathcal{D}_{\mathrm{BF}}$ and $\mathcal{D}_{\mathrm{AF}}^{S2}$ or $\mathcal{D}_{\mathrm{AF}}^{S3}$

V.5. Discussion

In this chapter, we use the kernel plus method based on the additive multivariate kernel method, described in Chapter IV, for the purpose of turbine performance assessment. This kernel plus method (additive multivariate kernel method) serves as a useful tool in producing the so-called endogenous power curve where the influence of

environmental factors can be controlled for so that the energy production efficiency of a turbine can be much better quantified. We present a new procedure including the kernel plus method to quantify the changes in turbine power generation associated with three different cases of upgrades or occurrence of faults. We found that the findings from the kernel plus method (additive multivariate kernel method) are consistent with the action of upgrading or occurrence of faults, while the findings from the binning method appears random, due to the large, uncontrolled uncertainty associated with this method. We want to note that even though we include four weather covariates in this chapter, the resulting kernel plus method (additive multivariate kernel method) is scalable and can easily incorporate many more covariates without running into the curse of dimensionality or data scarcity problems.

CHAPTER VI

CONCLUSION

In this chapter we first summarize the methods developed in this dissertation and highlight the contributions. We then provide some thoughts on future directions, including immediate extensions of the current work and other related problems.

VI.1. Summary

The objective of this dissertation research is to construct new conditional models for estimating extreme load levels and power curves under stochastic weather conditions. In modeling loads or power outputs, one challenge is to address stochastic uncertainty caused by a variety of environmental factors. In the wind industry, the current practice is to use the binning method, which is to discretize the domain of a few environmental factors and to estimate with each bin. The major limitation of the binning method lies in its rigid compartmentalization of data and its separate use of the data for each individual bin instead of borrowing strength from other bins. A binning method is effective when applied to single dimensional data but runs into the curse of dimensionality and loses its effectiveness when used for handling multivariate data. To address the issue, we develop a Bayesian spline model and an additive multivariate kernel model for the extreme load level and power curve, respectively.

In Chapter III, we provide a new extreme load analysis model estimating the site-specific structural design load (extreme load levels). Mathematically, the extreme load level is defined as an extreme quantile of the turbine's maximum load distribution. We consider the marginal maximum load distribution by using the conditional short-term distribution given weather profiles due to the non-stationary nature of the loads across

the entire wind profile. The novelty of our approach is in our proposed treatment to model the short-term distribution. To establish the desired non-homogeneous GEV distribution for the short-term distribution, we use a flexible Bayesian spline model. The location parameter and scale parameter in the short-term distribution are adequately modeled by the Bayesian MARS models. We also employ the Bayesian MARS model for estimating the distribution of wind characteristics. Finally, based on both the short-term distribution and wind characteristics distribution, we obtain an empirical predictive distribution of the extreme load level by using a RJMCMC algorithm. Our case study using one set of simulated data and three sets of real data show that the current industry method tends to overestimate the extreme quantile values.

In Chapter IV, we provide a new model to estimate the power curve more accurately. In the power curve estimation, the main challenge is to control for the influence from environmental effects such as wind direction, air density (temperature, pressure), standard deviation of wind speed, wind shear, and humidity. For handling the multi-dimensional power curve estimation problem in a reasonable computation time, we devise an additive multivariate kernel model. The advantage of our model is to capture the nonlinear relationships between environmental factors and the wind power output, as well as the high-order interaction effects among some of the environmental factors. The additive multivariate kernel model includes three-variable product kernels, which can capture the interaction effects between another environmental factor with wind speed and wind direction. Subsequently, all the multivariate kernels constitute an additive model. Then, we provide the point estimation and density estimation for power outputs by employing heuristic bandwidth algorithms. Finally, we evaluate the performance of our additive multivariate kernel with the RMSE and CRPS measure using real field data.

The power curve can be applied in practice for a variety of important tasks including wind power forecasting and assessing turbine performance efficiency. In Chapter V, we incorporate the power curve estimation model and present a new procedure for turbine performance assessment. Most of the existing studies use a constant threshold of power residuals over the whole wind speed region. Such an approach is usually ineffective in detecting small changes due to the high variance in the residuals. To address this issue, we suggest the use of the modified additive multivariate kernel model (kernel plus method) in computing the residuals, and then conduct a Student's t-test on the difference between the residuals before and after an upgrade or a fault occurrence. We evaluate the new procedure on three different case studies and substantiate the claim of benefit made earlier.

We believe that the proposed models contribute to wind power research in the following ways:

1. We provide a new extreme load analysis model estimating the site-specific structural design load (extreme load levels). This model allows us to calculate the structural design load for the same kind of wind turbine that will be built at a different site. The conditional distribution modeling for the structural design load is a necessary practice in the wind industry. A turbine needs to be assessed for its ability to resist the extreme loads under the specific wind profile at the site it will be installed. Turbine manufacturers usually test a small number of representative turbines at their own testing site, producing $p(y|\boldsymbol{x})$. When a turbine is to be installed at a commercial wind farm, the wind profile at the proposed installation site can be collected and substituted into (1.3) as $p(\boldsymbol{x})$, so that the site-specific extreme load can be assessed. Without the conditional distribution model, a turbine test would have to be done for virtually every new

wind farm. Doing so is very costly and thus uncommon.

2. We provide a new additive multivariate kernel model to estimate the power curve more accurately. This model includes more explanatory variables such as wind direction, air density (temperature, pressure), standard deviation of wind speed, wind shear, and humidity, which will eventually need to be controlled for and excluded from the final estimate of the endogeneous power curve. The contribution of this new power curve model is that it can help examine the relative importance of the environmental factors in terms of their ability to affect the operation efficiency of wind turbines, and to make wind power prediction more accurate.

3. We present case studies to showcase how the new kernel-based power curve model can be used to inform decision making for turbine performance assessments. This is done for quantifying production gain from turbine upgrades as well as production loss from turbine faults. Our case studies demonstrated the benefit of the enhanced power prediction capabilities of the new power curve model, as it can detect subtle, small changes associated with a turbine where the conventional binning method cannot.

VI.2. Suggestions for future research

In the first study of extreme load assessment, we employ a Bayesian approach which uses the SIC criterion to guide the selection of the model. Then, to evaluate the tail part of a conditional maximum load distribution, we apply a different criterion, the GPL. In the approach of the Bayesian decision theory, it is possible to incorporate the GPL loss function in the given problem and to find optimal estimators minimizing the posterior expected loss. Doing so will lead to a coherent, unified framework, which

also may give better results. It is worth exploring this new estimation framework in a future study.

In the second study of power curve estimation, whether or not we should have used wind measurements obtained right at the turbine site, as opposed to those obtained at the mast site, is a valid question. On commercial wind farms, almost every turbine is indeed equipped with an *in-situ* anemometer, which in theory can measure the wind speed and direction specific to that turbine. In practice, however, turbine-based anemometer data are not commonly used for power curve estimation or wind power prediction purposes. There have been some discussions concerning the pros and cons of using turbine-based anemometer data (Albers *et al.*, 1999, Hayes *et al.*, 2011). The main obstacle is that turbine-based anemometers are not calibrated to the industry standard, and as a result, their wind measurements tend to be inaccurate. Manually calibrating all the turbine anemometers is a very costly proposition for a commercial wind farm, spread over tens of miles and housing over hundreds of turbines. It is also the case that turbine anemometers can measure wind speed only at a single height (i.e., the hub height) and are thus useless in computing wind dynamics measures like wind shear. The wind measurements obtained from a mast are generally considered to be accurate, or at least as accurate as they can be, since the industry standard does require the masts to be calibrated periodically (IEC, 2005b). But entirely depending on wind measurements from a handful of meteorological masts is not ideal, either, especially when the turbines are located several kilometers away from the mast. We are not aware of a solution to address this problem, but this issue definitely deserves future research attention.

REFERENCES

Ackermann, T. (2005) *Wind Power in Power Systems*, John Wiley & Sons Ltd., West Sussex, UK.

Agarwal, P. and Manuel, L. (2008) Extreme loads for an offshore wind turbine using statistical extrapolation from limited field data. *Wind Energy*, **11**, 673–684.

Albers, A., Klug, H. and Westermann, D. (1999) Power performance verification, in *Proceedings of European Wind Energy Conference (EWEC 99)*, Nice, France, pp. 657–660.

AWEA (2012) *U.S. Wind Industry Fourth Quarter 2011 Market Report*, American Wind Energy Association (AWEA), Washington, D.C.

Belghazi, O. and Cherkaoui, M. (2012) Pitch angle control for variable speed wind turbines using genetic algorithm controller. *Journal of Theoretical and Applied Information Technology*, **39**, 6–10.

Bossanyi, E. (2003) GH bladed theory manual, Technical Report 282/BR/009, Garrad Hassan and Partners Ltd.

Bottasso, C., Campagnolo, F. and Croce, A. (2010) Computational procedures for the multi-disciplinary constrained optimization of wind turbines, Technical Report DIA-SR 10-02, Dipartimento di Ingegneria Aerospaziale, available at http://www.aero.polimi.it/ bottasso/DownloadArea.html.

Carta, J., Ramírez, P. and Velázquez, S. (2008) Influence of the level of fit of a density probability function to wind-speed data on the WECS mean power output estimation. *Energy Conversion and Management*, **49**, 2647–2655.

Chattopadhyay, D. (2004) Life-cycle maintenance management of generating units in a competitive environment. *Power Systems, IEEE Transactions on*, **19**, 1181–1189.

Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, New York.

Denison, D., Holmes, C., Mallick, B. and Smith, A. (2002) *Bayesian Methods for Nonlinear Classification and Regression*, John Wiley & Sons Ltd., Chichester, UK.

Denison, D., Mallick, B. and Smith, A. (1998) Bayesian MARS. *Statistics and Computing*, **8**, 337–346.

DOE (2008) *20% Wind Energy by 2030*, U.S. Department of Energy (DOE), Washington, D.C.

Fan, J. and Yim, T. (2004) A crossvalidation method for estimating conditional densities. *Biometrika*, **91**, 819–834.

Fitzwater, L., Cornell, C. and Veers, P. (2003) Using environmental contours to predict extreme events on wind turbines, in *Proceedings of the ASME Wind Energy Symposium*, AIAA Paper-2003-865, Reno, Nevada.

Fitzwater, L. and Winterstein, S. (2001) Predicting design wind turbine loads from limited data: Comparing random process and random peak models, in *Proceedings of the ASME Wind Energy Symposium*, AIAA paper-2001-0046, Reno, Nevada.

Fogle, J., Agarwal, P. and Manuel, L. (2008) Towards an improved understanding of statistical extrapolation for wind turbine extreme loads. *Wind Energy*, **11**, 613–635.

Freudenreich, K. and Argyriadis, K. (2008) Wind turbine load level based on extrapolation and simplified methods. *Wind Energy*, **11**, 589–600.

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M. and Draxl, C. (2011) The state-of-the-art in short-term prediction of wind power: A literature overview, Technical Report ANEMOS Project, Risø National Laboratory.

Gill, S., Stephen, B. and Galloway, S. (2012) Wind turbine condition assessment through power curve copula modeling. *Sustainable Energy, IEEE Transactions on*, **3**, 94–101.

Gneiting, T. (2011) Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**, 746–762.

Gneiting, T., Larson, K., Westrick, K., Genton, M. and Aldrich, E. (2006) Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space-time method. *Journal of the American Statistical Association*, **101**, 968–979.

Gneiting, T. and Raftery, A. (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378.

Green, P. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

Hall, P., Racine, J. and Li, Q. (2004) Cross-validation and the estimation of conditional probability. *Journal of the American Statistical Association*, **99**, 1015–1026.

Hayes, B., Ilie, I., Porpodas, A., Djokic, S. and Chicco, G. (2011) Equivalent power curve model of a wind farm based on field measurement data, in *Proceedings of the IEEE Trondheim PowerTech 2011*, Trondheim, Norway, pp. 1–7.

Hering, A. and Genton, M. (2010) Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, **105**, 92–104.

111

Hyndman, R., Bashtannyk, D. and Grunwald, G. (1996) Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, **5**, 315–336.

IEC (1999) *IEC 61400-1 Ed 2: Wind Turbines-Part1: Design Requirements*, International Electrotechnical Commission, Geneva, Switzerland.

IEC (2005a) *IEC 61400-1 Ed 3: Wind Turbines-Part1: Design Requirements*, International Electrotechnical Commission, Geneva, Switzerland.

IEC (2005b) *IEC 61400-12-1 Ed 1: Wind Turbines-Part12-1: Power Performance Measurements of Electricity Producing Wind Turbines*, International Electrotechnical Commission, Geneva, Switzerland.

Jeon, J. and Taylor, J. (2012) Using conditional kernel density estimation for wind power density forecasting. *Journal of the American Statistical Association*, **107**, 66–79.

Kass, R. and Wasserman, L. (1995) A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, **90**, 928–934.

Kusiak, A., Zheng, H. and Song, Z. (2009) On-line monitoring of power curves. *Renewable Energy*, **34**, 1487–1493.

Li, G. and Shi, J. (2010) Application of Bayesian model averaging in model long-term wind speed distributions. *Renewable Energy*, **35**, 1192–1202.

Mahalanobis, P.C. (1936) On the generalised distance in statistics, in *Proceedings of the National Institute of Science of India*, pp. 49–55.

Manuel, L., Veers, P. and Winterstein, S. (2001) Parametric models for estimating wind turbine fatigue loads for design. *Journal of Solar Energy Engineering: Transactions of the ASME*, **123**, 346–355.

Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J. and Conzelmann, G. (2009) Wind power forecasting state-of-the-art, Technical Report ANL/DIS-10-1, Argonne National Laboratory, available at http://www.dis.anl.gov/pubs/65613.pdf.

Moriarty, P. (2008) Database for validation of design load extrapolation techniques. *Wind Energy*, **11**, 559–576.

Moriarty, P., Holley, W. and Butterfield, S. (2002) Effect of turbulence variation on extreme loads prediction for wind turbines. *Journal of Solar Energy Engineering: Transactions of the ASME*, **124**, 387–395.

Nadaraya, E. (1964) On estimating regression. *Theory of Probability and Its Applications*, **9**, 141–142.

Natarajan, A. and Holley, W. (2008) Statistical extreme load extrapolation with quadratic distortions for wind turbines. *Journal of Solar Energy Engineering: Transactions of the ASME*, **130**, 031017:1–7.

Nielsen, T., Nielsen, H. and Madsen, H. (2002) Prediction of wind power using time-varying coefficient functions, in *Proceedings of the 15th IFAC World Congress on Automatic Control*, Barcelona, Spain.

Osadciw, L.A., Yan, Y., Ye, X., Benson, G. and White, E. (2010) Wind Turbine Diagnostics Based on Power Curve Using Particle Swarm Optimization, in *Wind*

Power Systems (Green Energy and Technology), edited by Wang, L., Singh, C. and Kusiak, A., Springer-Verlag, Berlin Heidelberg, pp. 151–165.

Øye, S. (1995) The effect of vortex generators on the performance of the ELKRAFT 1000kW turbine, in *Proceedings of the 9th IEA Symposium on the Aerodynamics of Wind Turbines*, Stockholm, Sweden.

Peeringa, J. (2003) Extrapolation of extreme responses of a multi megawatt wind turbine, Technical Report ECN-C-03-131, Energy Research Centre of the Netherlands, available at http://www.ecn.nl/docs/library/report/2003/c03131.pdf.

Peeringa, J. (2009) Comparison of extreme load extrapolations using measured and calculated loads of a MW wind turbine, Technical Report ECN-M-09-055, Energy Research Centre of the Netherlands, available at http://www.ecn.nl/docs/library/report/2009/m09055.pdf.

Pinson, P., Nielsen, H., Madsen, H. and Nielsen, T. (2008) Local linear regression with adaptive orthogonal fitting for wind power application. *Statistics and Computing*, **18**, 59–71.

Raftery, A. (1995) Bayesian model selection in social research. *Sociological Methodology*, **25**, 111–163.

Regan, P. and Manuel, L. (2008) Statistical extrapolation methods for estimating wind turbine extreme loads. *Journal of Solar Energy Engineering: Transactions of the ASME*, **130**, 031011:1–15.

Rehman, S. and Al-Abbadi, N. (2005) Wind shear coefficients and their effect on energy production. *Energy Conversion and Management*, **46**, 2578–2591.

Ronold, K. and Larsen, G. (2000) Reliability-based design of wind-turbine rotor blades against failure in ultimate loading. *Engineering Structures*, **22**, 565–574.

Rosenblatt, M. (1969) Conditional probability density and regression estimates, in *Multivariate Analysis II*, edited by Krishnaiah, P., Academic Process, New York, pp. 25–31.

Ruppert, D., Sheather, S. and Wand, M. (1995) An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, **90**, 1257–1270.

Sánchez, I. (2006) Short-term prediction of wind energy production. *International Journal of Forecasting*, **22**, 43–56.

Schwartz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Smart Blade (2013) Custom vortex generators, retrieved Feburary 20, 2013, from http://smart-blade.com/products-services/vortex-generators.html.

Smith, R. (1990) Extreme value theory, in *Handbook of Applicable Mathematics*, edited by Ledermann, W., chapter 14, John Wiley & Sons, Inc., New York, pp. 437–471.

Snedecor, G. and Cochran, W. (1989) *Statistical Methods*, 8 edition, Iowa State University Press, Iowa City, Iowa, USA.

Sørensen, J. and Nielsen, S. (2007) Extreme wind turbine response during operation. *Journal of Physics: Conference Series*, **75**, 012074:1–8.

Stephen, B., Galloway, S., McMillan, D., Hill, D. and Infield, D. (2011) A copula model of wind turbine performance. *Power Systems, IEEE Transactions on*, **26**, 965–966.

Tindal, A., Johnson, C., LeBlanc, M., Harman, K., Rareshide, E. and Graves, A. (2008) Site-specific adjustments to wind turbine power curves, in *Proceedings of the American Wind Energy Association(AWEA) WINDPOWER Conference & Exhibition*, Houston, Texas, pp. 1–11.

Uluyol, O., Parthasarathy, G., Foslien, W. and Kim, K. (2011) Power curve analytic for wind turbine performance monitoring and prognostics, in *Proceedings of the Annual Conference of the prognostics and Health Management Society*, Montreal, Canada.

Veers, P. and Butterfield, S. (2001) Extreme load estimation for wind turbines: Issues and opportunities for improved practice, in *Proceedings of the ASME Wind Energy Symposium*, AIAA Paper-2001-0044, Reno, Nevada.

Walford, C. (2006) Wind turbine reliability: understanding and minimizing wind turbine operation and maintenance costs, Technical Report SAND2006-1100, Sandia National Laboratories, available at http://prod.sandia.gov/techlib/access-control.cgi/2006/061100.pdf.

Wan, Y., Ela, E. and Orwig, K. (2010) Development of an equivalent wind plant power curve, Technical Report NREL/CP-550-48146, National Renewable Energy Laboratory, available at http://www.nrel.gov/docs/fy10osti/48146.pdf.

Wang, L., Tang, X. and Liu, X. (2012) Blade design optimization for fixed-pitch fixed-speed wind turbines. *International Scholarly Research Network Renewable Energy*, **2012**, 682859:1–8.

Watson, G. (1964) Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, **26**, 359–372.

WindData (2010) Database on wind characteristics, retrieved June 2, 2010, from http://www.winddata.com, Accessed July-2010.

Yan, Y., Osadciw, A., Benson, G. and White, E. (2009) Inverse data transformation for change detection in wind turbine diagnostics, in *Proceedings of the 22nd IEEE Canadian Conference on Electrical and Computer Engineering*, Newfoundland, Canada, pp. 944–949.

Zhang, X., Zhang, J. and Gockenbach, E. (2009) Reliability centered asset management for medium-voltage deteriorating electircal equipment based Germany failure statistics. *Power Systema, IEEE Transactions on*, **24**, 721–728.

Zhu, X. and Genton, M. (2012) Short-term wind speed forecasting for power system operations. *International Statistical Review*, **90**, 2–23.

APPENDIX A

OBTAINING CONFIDENCE INTERVALS FOR THE POT METHOD

In this appendix, a procedure to calculate the confidence intervals for the POT method is presented. The POT method employs a Generalized Pareto (GP) distribution to model the tail of unconditional maximum loads. The cumulative distribution of the GP is

$$
\begin{aligned}
G_{\gamma,\chi}(y) &= 1 - \left(1 + \frac{\gamma(y-u)}{\chi}\right)^{-1/\gamma} && \text{if } \gamma \neq 0, \\
&= 1 - \exp\left(-\frac{y-u}{\chi}\right) && \text{if } \gamma = 0. && \text{(A.1)}
\end{aligned}
$$

where u is the threshold, $\chi > 0$ is the scale parameter, and γ is the shape parameter. Also, (A.1) is valid when $y - u \geq 0$ for $\gamma \geq 0$ and $0 \leq y - u \leq -\chi/\gamma$ for $\gamma < 0$. The threshold u is chosen graphically by looking at the threshold choice plots and mean residual life plots. Then, assuming that u is known, the other parameters are estimated using the maximum likelihood method.

Because of the assumption that the parameters χ and γ do not depend on wind covariates, the extreme load level l_T, given the target probability P_T, can be expressed directly in terms of function of parameters u, χ and γ as follows:

$$
l_T = u + \frac{\chi}{\gamma}(P_T^\gamma - 1). \tag{A.2}
$$

To calculate the confidence intervals for the POT method, we go through the following procedure:

- Draw a sample of χ and γ from a multivariate normal distributions taking the MLE as its mean and the inverse of the negative of the Hessian matrix as its covariance matrix.

- Compute the quantile value $l_T(\chi, \gamma)$ by using (A.2).

- Repeat the above procedure M_l times to get the mean and confidence intervals of l_T.

Our implementation here uses the same M_l as that used in the implementation in the spline method.

APPENDIX B

OBTAINING CONFIDENCE INTERVALS FOR THE BINNING METHOD

To calculate the confidence intervals for the binning method, we follow a procedure similar to the one used for calculating the credible intervals in the spline method. The difference is mainly that in the binning method, the parameters used in the GEV distribution, namely μ and σ (recall that ξ is fixed as a constant across all the bins), are sampled using only the data in a specific bin. For those bins which do not have data, its μ and σ are a weighted average of all non-empty bins with the weight related to the inverse squared distance between bins, following the approach used by Agarwal and Manuel (2008). Once a sample of μ and σ is obtained for a specific bin, the resulting local GEV is used to sample $\tilde{y}$ in that bin. Do this for all the bins, and the $\tilde{y}$'s from all bins are pooled together to estimate l_T.

Specially, we go through the following steps, where $\mathbf{\Phi}_c$ denotes the collection of the parameters associated with all local GEV distributions used in all bins.

- Draw $M_w \times N_w$ samples from the joint posterior predictive distribution $p\left[\tilde{v}, \tilde{s} | \mathcal{D}_v, \mathcal{D}_s\right]$ of wind characteristics $(\tilde{v}, \tilde{s})$; this step is the same as in the spline method.

- Using the data in a bin, draw a sample of μ and σ for that specific bin from a multivariate normal distributions taking the MLE as its mean and the inverse of the negative of the Hessian matrix as its covariance matrix. Not all the bins have data. For those which do not have data, its μ and σ are a weighted average of all non-empty bins with the weight related to the inverse squared distance

120

between bins, as we explained above. Collectively, $\boldsymbol{\Phi}_c$ contains all the μ's and σ's from all the bins.

- Decide which bins the wind characteristic samples $(\tilde{v}, \tilde{s})$'s fall into. Based on the specific bin in which a sample of $(\tilde{v}, \tilde{s})$ falls, the corresponding μ and σ in $\boldsymbol{\Phi}_c$ is chosen; doing this yields the short-term distribution $p\left[\tilde{y}|\tilde{v}, \tilde{s}, \boldsymbol{\Phi}_c\right]$ for that specific bin.

- Draw N_l samples of $\tilde{y}$ from $p\left[\tilde{y}|\tilde{v}, \tilde{s}, \boldsymbol{\Phi}_c\right]$ for each of the total $M_w \times N_w$ samples of $(\tilde{v}, \tilde{s})$. This produces a total of $M_w \times N_w \times N_l$ $\tilde{y}$ samples.

- One can then compute the quantile value $l_T[\boldsymbol{\Phi}_c]$ corresponding to P_T.

- Repeat the above procedure M_l times to get the mean and confidence intervals of l_T.

Our implementation here uses the same $M_w, M_l, N_w,$ and N_l as those used in the implementation in the spline method.