

## BAYESIAN SPLINE METHOD FOR ASSESSING EXTREME LOADS ON WIND TURBINES

BY GIWHYUN LEE<sup>†</sup>, EUNSHIN BYON<sup>‡</sup>, LEWIS NTAIMO<sup>†</sup>, AND YU DING<sup>†</sup>

*Texas A&M University*<sup>†</sup> and *University of Michigan*<sup>‡</sup>

This 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 imposed on a turbine structure that the turbine would experience during its service life time. 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, using load data collected in a period much shorter than a turbine's service life. The spline method is applied to three sets of turbine's load response data to estimate the corresponding extreme loads at the roots of the turbine blades. Compared to the current industry practice, the spline method appears to provide better extreme load assessment.

**1. Introduction.** A wind turbine operates under various loading conditions in stochastic weather environments. The increasing size, weight, and length of components of utility-scale wind turbines escalate the stresses (or, loads, responses) imposed on the structure. As a result, modern wind turbines are prone to experiencing 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

---

\*This analysis has benefited from measurements downloaded from the internet database: “Database of Wind Characteristics” located at DTU, Denmark. Internet: <http://www.winddata.com/>. Wind field time series from the following sites have been applied: Roskilde, Denmark; Alborg, Denmark; and Tehachapi Pass, California, US. The authors would also like to acknowledge the generous support from their sponsors. Lee, Byon and Ding were partially supported by NSF grants CMMI-0727305 and CMMI-0926803.

*Keywords and phrases:* Bayesian spline regression, Extreme load, Monte Carlo integration, Reliability, Wind Power

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 million Newton-meter (MN-m), during a specific time interval. Then, we define the load exceedance probability as follows:

$$(1.1) \quad P_T = P[y > l_T],$$

where  $P_T$  is the target probability of exceeding the load level  $l_T$  (in the same unit 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 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  $P_T$ . The waiting time until this event happens should be longer than, or equal to, the service time. Therefore, a reasonable level of  $P_T$  can be found in the following way [IEC (2005); Peeringa (2003)]:

$$(1.2) \quad P_T = \frac{10}{T \times 365.25 \times 24 \times 60}.$$

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

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:

$$(1.3) \quad p(y) = \int p(y|\mathbf{x})p(\mathbf{x})d\mathbf{x}.$$

Here,  $p(\mathbf{x})$  is the joint probability density function of wind characteristics in a covariate vector  $\mathbf{x}$ . The conditional distribution of  $y$  given  $\mathbf{x}$ ,  $p(y|\mathbf{x})$  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.

The conditional distribution modeling in (1.3) 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  $p(y|\mathbf{x})$ . 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  $p(\mathbf{x})$ , 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.

For in-land turbines, the wind characteristic vector  $\mathbf{x}$  in general comprises two elements: (1) a steady state mean of wind speed and (2) the stochastic variability of wind speed [Bottasso, Campagnolo and Croce (2010); Ronold and Larsen (2000); Manuel, Veers and Winterstein (2001)]. 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)].

In this study, we propose a new procedure to estimate the long-term extreme load level  $l_T$  for wind turbines. The novelty of the new procedure is primarily regarding how to model the short-term distribution  $p(y|\mathbf{x})$ . Specially, we establish a load distribution for  $y|\mathbf{x}$  using spline models. As such, we label the resulting method *a Bayesian spline method for extreme loads*. In the remainder of the paper, we first provide some background information regarding wind turbine load responses and the datasets used in this study. In Section 3, we explain how the extreme load estimation problem is currently solved. We proceed to present the details of our spline method in Section 4. In Section 5, we compare the spline method with the method reviewed in Section 3, arguing that the spline method produces better estimates. Finally, we end the paper with some concluding remarks in Section 6.

**2. Background and datasets.** Figure 1 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.

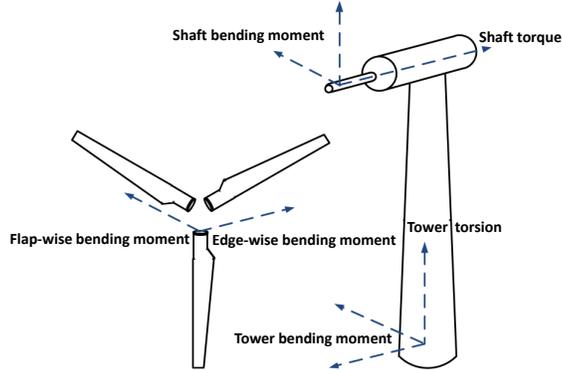


FIG 1. Illustration of structural loads at different components (The illustration is modified based on a figure originally available at [WindData].)

We only study in-land 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]. Table 1 summarizes the specification 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:

- *Pitch control*: To avoid production of excessive electricity, turbines hold the 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*: This is the lowest wind speed at a hub height at which a wind turbine starts to produce power.
- *Cut-out wind speed*: This is the speed beyond which a wind turbine shuts itself down to protect the turbine.
- *Rated wind speed*: This is 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 10-minute maximum blade-root flap-wise bending moment (hereafter, we call  $y$  a maximum load). 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*: 25 Hz = 15,000 measurements/10-min;
- *ILT2*: 32 Hz = 19,200 measurements/10-min;
- *ILT3*: 35.7 Hz = 21,420 measurements/10-min.

Here, 1 Hz means one measurement per second. The raw measured variables are  $v_{ij}$  and  $y_{ij}$ , where  $i = 1, \dots, n$  represents a 10-minute block of data and  $j = 1, \dots, N$  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:

$$(2.1) \quad v_i = \frac{1}{N} \sum_{j=1}^N v_{ij},$$

$$(2.2) \quad s_i = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (v_{ij} - v_i)^2} \quad \text{and}$$

$$(2.3) \quad y_i = \max \{y_{i1}, y_{i2}, \dots, y_{iN}\}.$$

**3. Literature review.** 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 IEC 61400-1:1999 are specified as discrete events based on design experiences and empirical models [Moriarty, Holley and Butterfield (2002)]. Veers and Butterfield (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, Holley

TABLE 1  
*Specifications of wind turbines in three datasets*

Wind turbine model (Name of dataset)	NEG-Micon/2750 (ILT1)	Vestas V39 (ILT2)	Nordtank 500 (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

and Butterfield (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 extreme 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.

After IEC 61400-1:2005 was issued, many studies were reported to devise and recommend statistical approaches for extreme load analysis [Freudenreich and Argyriadis (2008); Agarwal and Manuel (2008); Peeringa (2009); Moriarty (2008); Fogle, Agarwal and Manuel (2008); Regan and Manuel (2008); Natarajan and Holley (2008)]. These studies adopt a common framework, which we call *binning* method. The basic idea of the binning method is to discretize the domain of a wind profile vector  $\mathbf{x}$  into a finite number of bins. For example, one can divide the range of wind speed, 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  $y|\mathbf{x}$  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.

According to the classical extreme value theory [Coles (2001); Smith (1990)], the short-term distribution of  $y|\mathbf{x}$  can be approximated by a generalized extreme value (GEV) distribution. The probability density function of the GEV is

$$(3.1) \quad p(y) = \begin{cases} \frac{1}{\sigma} \exp \left[ - \left( 1 + \xi \left( \frac{y-\mu}{\sigma} \right) \right)^{-\frac{1}{\xi}} \right] \left( 1 + \xi \left( \frac{y-\mu}{\sigma} \right) \right)^{-1-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\ \frac{1}{\sigma} \exp \left[ -\frac{y-\mu}{\sigma} - \exp \left( -\frac{y-\mu}{\sigma} \right) \right] & \text{if } \xi = 0, \end{cases}$$

for  $\{y : 1 + \xi(y - \mu)/\sigma > 0\}$ , where  $\mu \in \mathfrak{R}$  is the location parameter,  $\sigma > 0$  is the scale parameter, and  $\xi \in \mathfrak{R}$  is the shape parameter that determines the weight of the tail of the distribution.  $\xi > 0$  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,  $\xi \rightarrow 0$ ) to the Gumbel distribution with a light upper tail [Coles (2001)].

One of 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 as a function of the distribution parameters as follows:

$$(3.2) \quad l_T = \begin{cases} \mu - \frac{\sigma}{\xi} \left[ 1 - (-\log(1 - P_T))^{-\xi} \right] & \text{if } \xi \neq 0, \\ \mu - \sigma \log[-\log(1 - P_T)] & \text{if } \xi = 0. \end{cases}$$

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 also recommended by IEC (2005). 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 bins). 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.

Despite its popularity, the binning method has obvious shortcomings in estimating extreme loads. A major limitation is 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), and consequently, some bins do not have enough data. Then, the binning method may end up with inaccuracy or big uncertainty in the estimates 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, he/she affords to use more bins; otherwise, fewer bins.

**4. Bayesian spline method for extreme load.** In this section, we present our new procedure of estimating the extreme load with two sub-models. The first sub-model (in Section 4.1) is the conditional maximum load model  $p(y|\mathbf{x})$ , and the second sub-model (in Section 4.3) is the distribution of wind characteristics  $p(\mathbf{x})$ . 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 2 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 ILT1, the load response of ILT2 has a large variation beyond the rated wind speed. This large variation can be attributed to its 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 2(c) does not have an obvious downward trend beyond the rated speed.

Figure 3 presents the scatter plots between the 10-minute maximum loads and the standard deviations of wind speed during the 10-minute intervals.

We also observe nonlinear relationships between them, especially for the new pitch-controlled ILT1. Figure 4 shows scatter plots of 10-minute standard deviation versus 10-minute average wind speed. Some previous studies [Moriarty, Holley and Butterfield (2002); Fitzwater, Cornell and Veers (2003)] suggest that the standard deviation of wind speed varies with the average wind speed, which appears consistent with what we observe in Figure 4.

4.1. *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 across the entire wind profile. What we propose here is to abandon the bins and instead use a non-homogeneous GEV distribution whose parameters are not constant but depend on weather conditions.

Our research started out with 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 limited amount 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 regions. 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  $y_1, \dots, y_n$  with corresponding covariate variables  $\mathbf{x}_1 = (v_1, s_1), \dots, \mathbf{x}_n = (v_n, s_n)$ , as defined in equations (2.1) and (2.2). We choose to model  $y_i$  with a GEV distribution:

$$(4.1) \quad y_i | \mathbf{x}_i \sim GEV(\mu(\mathbf{x}_i), \sigma(\mathbf{x}_i), \xi), \quad \sigma(\cdot) > 0,$$

where the location parameter  $\mu$  and scale parameter  $\sigma$  in this GEV distribution are a nonlinear function of wind characteristics  $\mathbf{x}$ . The shape parameter  $\xi$  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  $\xi$  fixed is to keep the final model from becoming overly complicated. Let us denote  $\mu(\mathbf{x}_i)$  and  $\sigma(\mathbf{x}_i)$  by

$$(4.2) \quad \mu(\mathbf{x}_i) = f(\mathbf{x}_i),$$

$$(4.3) \quad \sigma(\mathbf{x}_i) = \exp(g(\mathbf{x}_i)),$$

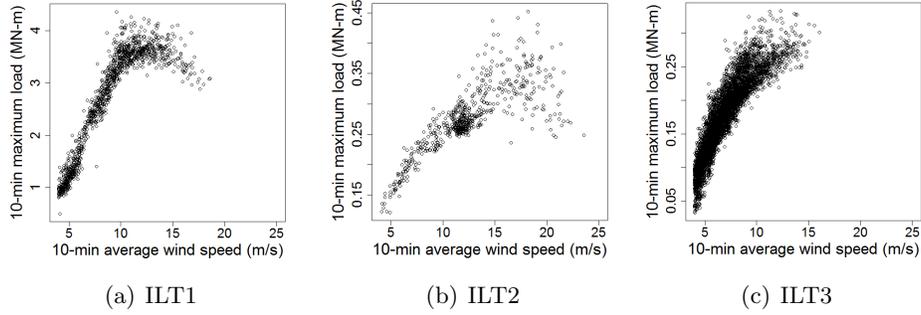


FIG 2. Scatter plots of 10-minute maximum load versus 10-minute average wind speed

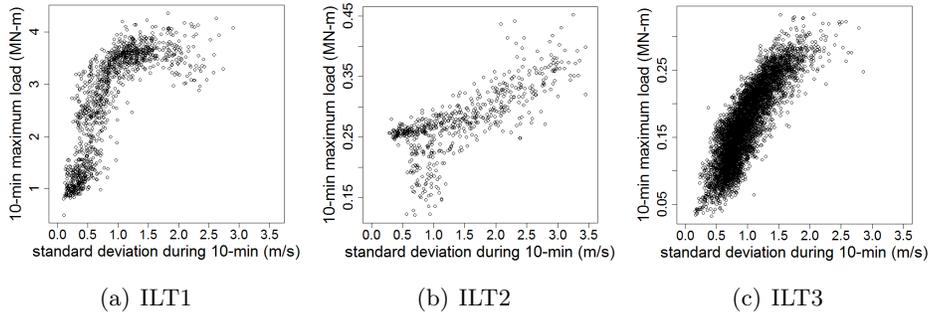


FIG 3. Scatter plots of 10-minute maximum load versus 10-minute standard deviations of wind speed

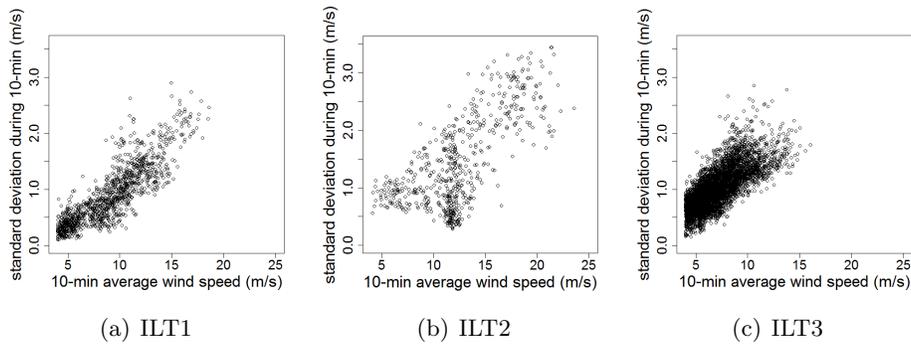


FIG 4. Scatter plots of 10-minute average wind speed versus 10-minute standard deviation of wind speed

where in (4.3), 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, Mallick and Smith (1998); Denison et al. (2002)] for capturing the nonlinearity between the load response and the wind-related covariates. The Bayesian MARS model has high flexibility. It includes the number and locations of knots as part of its model parameters and determines these from observed data. In addition, interaction effects among input factors can be modeled if choosing appropriate basis functions.

Specifically, the Bayesian MARS models  $f(\mathbf{x})$  for the location parameter  $\mu$  and  $g(\mathbf{x})$  for the scale parameter  $\sigma$  are represented as a linear combination of the basis functions  $B_k^\mu(\mathbf{x})$  and  $B_k^\sigma(\mathbf{x})$ , respectively, as:

$$(4.4) \quad f(\mathbf{x}) = \sum_{k=1}^{K_\mu} \beta_k B_k^\mu(\mathbf{x}),$$

$$(4.5) \quad g(\mathbf{x}) = \sum_{k=1}^{K_\sigma} \theta_k B_k^\sigma(\mathbf{x}),$$

where  $\beta_k, k = 1, \dots, K_\mu$  and  $\theta_k, k = 1, \dots, K_\sigma$  are the coefficients of the basis functions  $B_k^\mu(\cdot)$  and  $B_k^\sigma(\cdot)$ , respectively, and  $K_\mu$  and  $K_\sigma$  are the number of the respective basis functions. According to the study by Denison, Mallick and Smith (1998), which proposed the Bayesian MARS, the basis functions are specified as follows:

$$(4.6) \quad B_k(\mathbf{x}) = \begin{cases} 1, & k = 1 \\ \prod_{j=1}^{J_k} [h_{jk} \cdot (x_{r(j,k)} - t_{jk})]_+, & k = 2, 3, \dots, K. \end{cases}$$

Here,  $[\cdot]_+ = \max(0, \cdot)$ ,  $J_k$  is the degree of interaction modeled by the basis function  $B_k(\mathbf{x})$ ,  $h_{jk}$  is the sign indicator, 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.

We here introduce an integer variable  $T_k$  to represent the types of basis functions used in equation (4.6). Since we consider two predictors  $v$  and  $s$  for inland turbines, there could be three types of basis functions, namely  $[\pm(v - *)]_+$  and  $[\pm(s - *)]_+$  for each explanatory variable, respectively, and  $[\pm(v - *)]_+[\pm(s - *)]_+$  for interactions between them. So we let  $T_k$  to take the integer value of 1, 2, or 3, to represent the three types of basis functions. That is,  $[\pm(v - *)]_+$  is represented by  $T_k = 1$ ,  $[\pm(s - *)]_+$  represented by  $T_k = 2$ , and  $[\pm(v - *)]_+[\pm(s - *)]_+$  represented by  $T_k = 3$ . When  $J_k = 1$

in equation (4.6), then the first two types of basis functions are used, while when  $J_k = 2$ , all three types of basis functions are used. In our model, we set  $J_k = 1$  or  $J_k = 2$  in the model of the location parameter  $\mu$  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,  $J_k = 1$  is set for its location parameter  $\mu$ . For the scale parameter  $\sigma$ , we always set  $J_k = 1$  for all three datasets.

Let  $\Psi_a = (\Psi_\mu, \Psi_\sigma, \xi)$  denotes all the parameters used in model (4.1), where  $\Psi_\mu$  and  $\Psi_\sigma$  include the parameters in function  $f(\cdot)$  and  $g(\cdot)$ , respectively. These parameters are grouped into two sets: (1) the coefficients of the basis functions in  $\beta = (\beta_1, \dots, \beta_{K_\mu})$  or  $\theta = (\theta_1, \dots, \theta_{K_\sigma})$ , and (2) the number and locations of the knots, and the types of basis function in  $\phi_\mu$  or  $\phi_\sigma$ , as follows:

$$(4.7) \quad \begin{aligned} \phi_\mu &= (K_\mu, \Lambda_2^\mu, \dots, \Lambda_{K_\mu}^\mu), \\ \text{where } \Lambda_k^\mu &= \begin{cases} (T_k^\mu, h_{1k}^\mu, t_{1k}^\mu), & \text{when } T_k^\mu = 1, 2; \\ (T_k^\mu, h_{1k}^\mu, h_{2k}^\mu, t_{1k}^\mu, t_{2k}^\mu), & \text{when } T_k^\mu = 3, \end{cases} \end{aligned}$$

and

$$(4.8) \quad \begin{aligned} \phi_\sigma &= (K_\sigma, \Lambda_2^\sigma, \dots, \Lambda_{K_\sigma}^\sigma), \\ \text{where } \Lambda_k^\sigma &= (T_k^\sigma, h_{1k}^\sigma, t_{1k}^\sigma), \quad \text{when } T_k^\sigma = 1, 2. \end{aligned}$$

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 (4.1), priors of the parameters involved should be specified. In this paper, we use uniform priors on  $\phi_\mu$  and  $\phi_\sigma$ ; see the detailed expression in Appendix A. Given  $\phi_\mu$  and  $\phi_\sigma$ , we specify the prior distribution for the parameters  $(\beta, \theta, \xi)$  as the unit-information prior, i.e., UIP (Kass and Wasserman, 1995), which is defined by setting the corresponding covariance matrix to be equal to the Fisher information of one observation.

*4.2. Sub-model 1: Posterior distribution of parameters.* The Bayesian MARS model treats the number and locations of the knots as random quantities. 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, researchers usually use a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm developed by Green (1995). The acceptance probability for a RJMCMC algorithm

includes a Jacobian term, which accounts for the change in dimension. However, under the assumption that the model space for parameters of varying dimension is discrete, there is no need for a Jacobian. In our analysis, this assumption is satisfied since we only consider probable models over all possible knot locations and numbers. Therefore, instead of using the RJMCMC algorithm, we use the reversible jump sampler (RJS) algorithm proposed in [Denison et al. \(2002\)](#). Since the RJS algorithm does not require new parameters to match dimensions between models and the corresponding Jacobian term to the acceptance probability, it is simpler and more efficient to execute.

To allow for dimensional changes, there are three actions in a RJS 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 change. 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 equal probability (i.e.,  $\frac{1}{3}$ ) to propose any of the three moves, and then, use the following acceptance probability  $\alpha$  for a proposed move from a model having  $k$  basis functions to a model having  $k^c$  basis functions:

$$(4.9) \quad \alpha = \min \{1, \text{the ratio of marginal likelihood} \times R\},$$

where  $R$  is a ratio of probabilities defined as:

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

We have  $R = 1$  for most cases, because the probabilities in the denominator and numerator are equal, except when  $k$  reaches either the upper or the lower bound.

The marginal likelihood in (4.9) can be expressed as follows:

$$(4.10) \quad p(\mathcal{D}_y | \phi_\mu, \phi_\sigma) = \int p(\mathcal{D}_y | \beta, \theta, \xi, \phi_\mu, \phi_\sigma) p(\beta, \theta, \xi | \phi_\mu, \phi_\sigma) d\beta d\theta d\xi.$$

where  $\mathcal{D}_y = (y_1, \dots, y_n)$  represents a set of observed load data. Since it is difficult to calculate the above marginal likelihood analytically in our study, we consider an approximation of  $p(\mathcal{D}_y | \phi_\mu, \phi_\sigma)$ . [Kass and Wasserman \(1995\)](#)

and Raftery (1995) showed that when UIP priors are used, the marginal log-likelihood, i.e.,  $\log(p(\mathcal{D}_y|\phi_\mu, \phi_\sigma))$ , can be reasonably approximated by the Schwarz information criterion (SIC) [Schwarz (1978)]. The SIC is expressed as:

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

where  $\hat{\beta}, \hat{\theta}, \hat{\xi}$  are the maximum likelihood estimators (MLEs) of the corresponding parameters obtained conditional on  $\phi_\mu$  and  $\phi_\sigma$ , 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  $\phi_\mu$  and  $\phi_\sigma$  in the RJS 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:

$$(4.11) \quad \log \frac{p(\mathcal{D}_y|\phi_\mu^c, \phi_\sigma)}{p(\mathcal{D}_y|\phi_\mu, \phi_\sigma)} \simeq \text{SIC}_{\phi_\mu^c, \phi_\sigma} - \text{SIC}_{\phi_\mu, \phi_\sigma}, \text{ and}$$

$$(4.12) \quad \log \frac{p(\mathcal{D}_y|\phi_\mu, \phi_\sigma^c)}{p(\mathcal{D}_y|\phi_\mu, \phi_\sigma)} \simeq \text{SIC}_{\phi_\mu, \phi_\sigma^c} - \text{SIC}_{\phi_\mu, \phi_\sigma}.$$

Then, we use two acceptance probabilities  $\alpha_\mu$  and  $\alpha_\sigma$  for accepting or rejecting a new state in  $\phi_\mu$  and  $\phi_\sigma$ , respectively. Using the SICs,  $\alpha_\mu$  and  $\alpha_\sigma$  are expressed as:

$$(4.13) \quad \alpha_\mu = \min \left\{ 1, \exp \left( \text{SIC}_{\phi_\mu^c, \phi_\sigma} - \text{SIC}_{\phi_\mu, \phi_\sigma} \right) \times R \right\}, \text{ and}$$

$$(4.14) \quad \alpha_\sigma = \min \left\{ 1, \exp \left( \text{SIC}_{\phi_\mu, \phi_\sigma^c} - \text{SIC}_{\phi_\mu, \phi_\sigma} \right) \times R \right\}.$$

In order to produce the samples from the posterior distribution of parameters in  $\Psi_a$ , we sequentially draw samples for  $\phi_\mu$  and  $\phi_\sigma$  by using the two acceptance probabilities, while marginalizing out  $(\beta, \theta, \xi)$ ; and then, conditional on the sampled  $\phi_\mu$  and  $\phi_\sigma$ , draw samples for  $(\beta, \theta, \xi)$  using a Normal approximation based on the maximum likelihood estimates and the observed information matrix. The detailed simulation procedure can be found in Step I of Appendix B.

4.3. *Sub-model 2: distribution of wind characteristics.* To find a site-specific load distribution, the distribution of wind characteristics  $p(\mathbf{x})$  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 speeds  $s$  in Figure 4, the distribution of wind characteristics  $p(\mathbf{x})$  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 a Rayleigh distribution (RAY) [IEC (2005)]. These two distributions are arguably the most widely used ones for this purpose. Carta, Ramirez and Velazquez (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 distribution of 10-minute average wind speed  $v$  is expressed as:

$$(4.15) \quad v_i \sim \mathcal{M}_v(\boldsymbol{\nu}),$$

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  $\nu_1$ ,  $\nu_2$  and  $\nu_3$  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 (2005)], which appears to be what researchers have commonly used; see, for example, Fitzwater, Cornell and Veers (2003). The distribution is characterized by a location parameter  $\eta$  and a scale parameter  $\delta$ . In the literature, both  $\eta$  and  $\delta$  are treated as a constant. But we observe that datasets measured at different sites have different relationships between the average wind speed  $v$  and the standard deviation  $s$ . Some of the  $v$ -versus- $s$  scatter plots show nonlinear patterns.

Motivated by this observation, we employ a Bayesian MARS model for modeling  $\eta$  and  $\delta$ , 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:

$$(4.16) \quad s_i|v_i \sim TN2(\eta(v_i), \delta(v_i)),$$

where  $\eta(v_i) = f_\eta(v_i)$  and  $\delta(v_i) = \exp(g_\delta(v_i))$ ,

where  $f_\eta$  and  $g_\delta$ , like their counterparts in (4.4) and (4.5), are linear combinations of the basis functions taking the general form (4.6). Notice that both of the functions have only one input variable, which is the average wind speed.

Let  $\Psi_\eta = (\beta_\eta, \phi_\eta)$  and  $\Psi_\delta = (\theta_\delta, \phi_\delta)$  denote the parameters in  $f_\eta(\cdot)$  and  $g_\delta(\cdot)$ . Since the basis functions  $f_\eta$  and  $g_\delta$  in (4.16) have only one input variable, only one type of basis function (i.e.,  $T_k = 1$ ) is needed. Hence,  $\phi_\eta$  and  $\phi_\delta$  are much simpler than  $\phi_\mu$  and  $\phi_\sigma$ , their counterparts in (4.7) and (4.8), and are expressed as follows:

$$(4.17) \quad \begin{aligned} \phi_\eta &= \left( K_\eta, \Lambda_2^\eta, \dots, \Lambda_{K_\eta}^\eta \right), \\ &\text{where } \Lambda_k^\eta = (T_k^\eta, h_{1k}^\eta, t_{1k}^\eta) \quad \text{and} \quad T_k^\eta = 1; \end{aligned}$$

and

$$(4.18) \quad \begin{aligned} \phi_\delta &= \left( K_\delta, \Lambda_2^\delta, \dots, \Lambda_{K_\delta}^\delta \right), \\ &\text{where } \Lambda_k^\delta = (T_k^\delta, h_{1k}^\delta, t_{1k}^\delta) \quad \text{and} \quad T_k^\delta = 1. \end{aligned}$$

We choose the prior distribution for  $(\beta_\eta, \theta_\delta)$  as UIP and the prior for  $(\phi_\eta, \phi_\delta)$  as uniform distribution, and solve this Bayesian MARS model by using a RJS algorithm, as in the preceding two sections. The predictive distributions of the average wind speed  $\tilde{v}$  and the standard deviation  $\tilde{s}$  are

$$(4.19) \quad p(\tilde{v}|\mathcal{D}_v) = \int p(\tilde{v}|\boldsymbol{\nu})p(\boldsymbol{\nu}|\mathcal{D}_v)d\boldsymbol{\nu}$$

$$(4.20) \quad p(\tilde{s}|\tilde{v}, \mathcal{D}_v, \mathcal{D}_s) = \int \int p(\tilde{s}|\tilde{v}, \Psi_\eta, \Psi_\delta)p(\Psi_\eta, \Psi_\delta|\mathcal{D}_v, \mathcal{D}_s)d\Psi_\eta d\Psi_\delta$$

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 Appendix B.

4.4. *Posterior predictive distribution of the extreme load level  $l_T$ .* We are interested in getting the posterior 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,  $p[\tilde{y}|\mathcal{D}, \Psi_a]$ , which is

$$(4.21) \quad p[\tilde{y}|\mathcal{D}, \Psi_a] = \int \int p[\tilde{y}|\tilde{v}, \tilde{s}, \Psi_a, \mathcal{D}]p[\tilde{v}, \tilde{s}|\mathcal{D}_v, \mathcal{D}_s]d\tilde{v}d\tilde{s},$$

where  $p[\tilde{v}, \tilde{s}|\mathcal{D}_v, \mathcal{D}_s]$  can be expressed as the product of (4.19) and (4.20).

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 Appendix B);
- Draw a set of samples from the posterior distribution of model parameters  $\Psi_a = (\Psi_\mu, \Psi_\sigma, \xi)$ ; this is realized by employing the RJS algorithm in Section 4.2 (or Step I in Appendix B);
- Given the above samples of wind characteristics and model parameters, we calculate  $(\mu, \sigma, \xi)$  that are needed in a GEV distribution; this yields a short-term distribution  $p[\tilde{y} | \tilde{v}, \tilde{s}, \Psi_a]$ ;
- Integrating out the wind characteristics  $(\tilde{v}, \tilde{s})$ , obtain the long-term distribution  $p[\tilde{y} | \mathcal{D}, \Psi_a]$ .
- Draw samples from  $p[\tilde{y} | \mathcal{D}, \Psi_a]$ , and compute a quantile value  $l_T[\Psi_a]$  corresponding to  $P_T$ .

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

## 5. Results.

5.1. *Model selection.* Table 2 presents the SIC values of the six candidate average wind speed models using different ILT datasets. 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 dominates 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 models*

Distributions	ILT1	ILT2	ILT3
W2	-2984	-1667	-12287
W3	<b>-2941</b>	<b>-1663</b>	<b>-11242</b>
RAY	-3120	-1779	-13396
LN3	-2989	-1666	-11444
G3	-2974	-1666	-11290
IG3	-2986	-2313	-11410

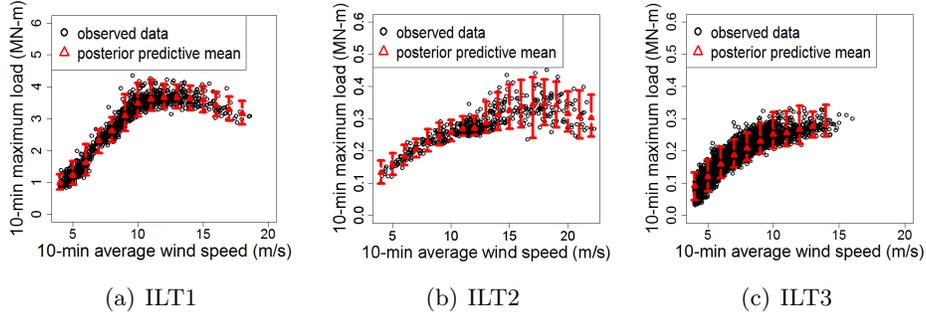


FIG 5. 95% point-wise credible intervals for different wind speeds

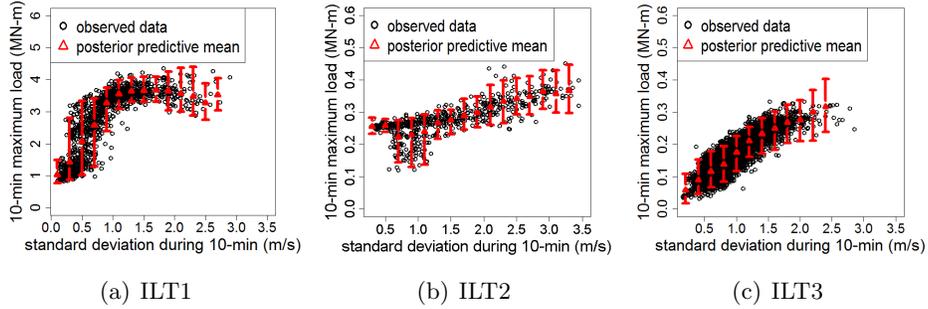


FIG 6. 95% point-wise credible intervals for different standard deviations

5.2. *Point-wise credible intervals.* As a form of checking the conditional maximum load model, we present in Figures 5 and 6 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,  $p(\tilde{y}|\mathbf{x})$ . 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:

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

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

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  $\mathcal{D}_{s_*} = \{(v_i, s_i) :$

$s_* - 0.05 < s_i < s_* + 0.05$ , 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 5) or a  $s$ -plot (Figure 6). To complete these 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.

5.3. *Comparison between the binning method and spline method for conditional maximum load.* 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  $p(y|\mathbf{x})$ , 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 levels in the long-term distribution, because of the relatively small amount of observation records. What we undertake in this section is to evaluate a method's performance of estimating the tail of the short-term distribution  $p(y|\mathbf{x})$ . We argued before that the short-term distribution underlies the difference between the proposed Bayesian spline method and the binning method. The comparison in this section is intended to show the advantage of the Bayesian spline method. In Section 5.5, we employ a simulation study that generates a much larger dataset, allowing us to compare the performance of two methods in estimating the extreme load level in the long-term distribution.

To evaluate the tail part of a conditional maximum load distribution, we 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:

$$\begin{aligned}
 & S_{\tau,b}(\hat{l}(\mathbf{x}_i), y(\mathbf{x}_i)) \\
 (5.1) \quad & = \begin{cases} \left( \mathbb{1}(\hat{l}(\mathbf{x}_i) \geq y(\mathbf{x}_i)) - \tau \right) \frac{1}{|b|} ([\hat{l}(\mathbf{x}_i)]^b - [y(\mathbf{x}_i)]^b) & \text{for } b \neq 0, \\ \left( \mathbb{1}(\hat{l}(\mathbf{x}_i) \geq y(\mathbf{x}_i)) - \tau \right) \log \left( \frac{\hat{l}(\mathbf{x}_i)}{y(\mathbf{x}_i)} \right) & \text{for } b = 0, \end{cases}
 \end{aligned}$$

where  $\hat{l}(\mathbf{x}_i)$  is the  $\tau$ -quantile estimation of  $p(y|\mathbf{x}_i)$  for a given  $\mathbf{x}_i$ ,  $y(\mathbf{x}_i)$  is the observed maximum load in the test dataset, given the same  $\mathbf{x}_i$ ,  $b$  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  $p(y|\mathbf{x})$ . For any  $\mathbf{x}_i$  in the test set, the  $\tau$ -quantile estimation  $\hat{l}(\mathbf{x}_i)$  can be computed using  $p(y|\mathbf{x})$ . And then, the GPL loss function value is taken as the average of all  $S_{\tau,b}$  values over the test set, as follows:

$$(5.2) \quad \bar{S}_{\tau,b} = \frac{1}{n_t} \sum_{i=1}^{n_t} S_{\tau,b}(\hat{l}_i(\mathbf{x}_i), y_i),$$

where  $n_t$  is the number of data points in a test set, and  $y_i$  is the same as  $y(\mathbf{x}_i)$ . We call  $\bar{S}_{\tau,b}$  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 mean score the mean score and use  $\bar{S}_{\tau,b}$  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 RJS algorithm in Section 4.2, we draw  $N_l = 100$  samples from the short term distribution. Accordingly, we can evaluate the quality of quantile estimations of the short term distribution for a  $\tau$  up to 0.99.

We first take a look at the comparisons in Figure 7, which compares the PL loss (i.e.,  $b = 1$ ) of both methods as  $\tau$  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 value 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  $\tau$  is approaching 0.99 in Figure 7, 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  $\tau$ , so that their differences are compressed in the figure. If one looks at the solid 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 in the cases of ILT1 and ILT3 datasets. When  $\tau$  gets larger, the spline method produces a significant improvement over the binning method, with a reduction of PL loss ranging from 33% to 50%. The trend is different when using the ILT2 dataset. But still, the spline method can reduce the mean scores of the PL loss from the binning method by 8% to 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.

We compute the mean scores of the GPL loss under three different power parameters  $b = 0, 1, 2$  for each method. Table 3 presents the results under  $\tau = 0.9$ , while Table 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, these results clearly show the improvement achieved by employing the Bayesian spline method.

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

In order to understand the difference between the spline method and binning method, we compare the 0.99 quantiles of the 10-minute maximum

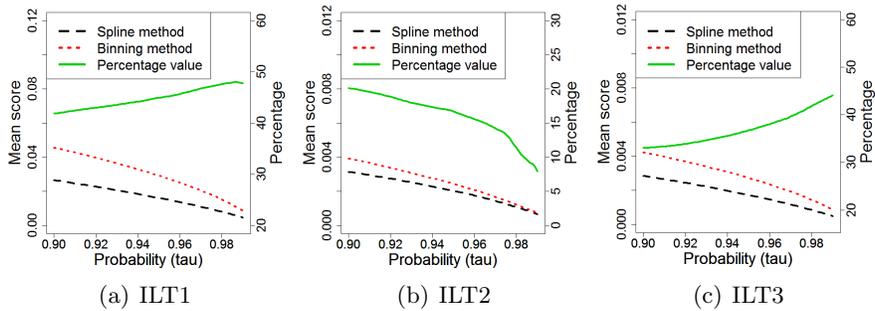


FIG 7. Comparison of PL function: the left Y-axis represents the mean score values and the right Y-axis represents the percentage values, which are the reduction in the mean scores when the spline method is compared with the binning method.

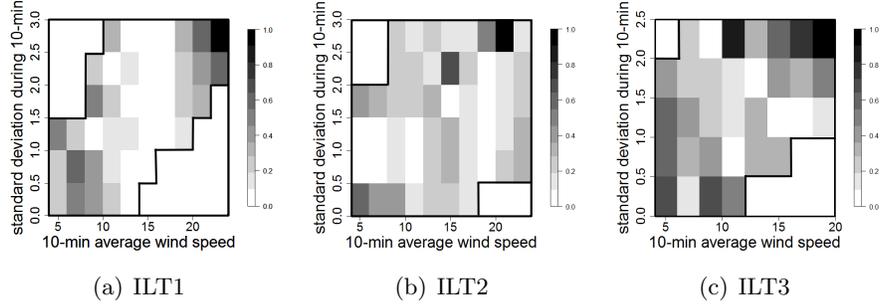


FIG 8. Comparison of the 0.99-quantiles between binning method and spline method

load conditional on a specific wind condition. This is done by computing the difference in the quantile values of conditional 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 8 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 weather conditions, the spline method is able to make better use of the available data and overcome the limited data problem for rare weather events.

5.4. *Estimation of extreme load.* Finally, Tables 5 and 6 show the estimates of the extreme load levels  $l_T$ , corresponding to  $T = 20$  and  $T = 50$  years, respectively. The values in parenthesis are the 95% credible (or confidence) intervals.

We observe that the extreme load levels  $l_T$  obtained by the binning method are generally higher than those obtained by the spline method. 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 detailed procedure for

TABLE 5  
*Estimates of extreme load levels ( $l_T, T = 20$  years), unit: MN-m*

Datasets	Binning method	Spline method
ILT1	6.455 (6.063, 7.092)	4.750 (4.579, 4.955)
ILT2	0.752 (0.658, 0.903)	0.576 (0.538, 0.627)
ILT3	0.505 (0.465, 0.584)	0.428 (0.398, 0.463)

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

Datasets	Binning method	Spline method
ILT1	6.711 (6.240, 7.485)	4.800 (4.611, 5.019)
ILT2	0.786 (0.682, 0.957)	0.589 (0.547, 0.646)
ILT3	0.527 (0.480, 0.621)	0.438 (0.405, 0.476)

computing the binning method’s confidence interval is included in Appendix C.

5.5. *Simulation of extreme load.* In this section, a simulation study is undertaken to assess the estimation accuracy of extreme load level in 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) Generate a sample  $x_i$  from a 3-parameter Weibull distribution. Then sample  $x_{ij}, j = 1, \dots, 1,000$ , from a normal distribution having  $x_i$  as its mean and a unit variance. The set of  $x_{ij}$ ’s represents the different wind speeds within a bin.
- (b) Draw the samples  $y_{ij}$  from a normal distribution with its mean as  $\mu_{ij}^s$  and its standard deviation as  $\sigma_{ij}^s$ , which are expressed as follows:

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

$$\sigma_{ij}^s = 0.1 \times \log(x_{ij}).$$

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 so that the simulated  $y$  looks like the actual mechanical load response. While many of the parameters used above do not have any physical meaning, some of them do; for instance, the “17” in “ $x_i < 17$ ” bears the meaning of the rated wind speed.

- (c) Find the maximum value  $y_i = \max\{y_{i,1}, \dots, y_{i,1000}\}$ , 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.
- (d) Repeat (a) through (c) for  $i = 1, \dots, 1000$  to produce the training dataset with  $n = 1,000$  data pairs, and denote this dataset by  $\mathcal{D}_{TR} = \{(x_1, y_1), \dots, (x_{1000}, y_{1000})\}$ .

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 4.4, which produces the posterior 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 also obtain 100 different samples of these quantiles.

Figure 9(a) shows a scatter plot of the simulated  $x$ 's and  $y$ 's in  $\mathcal{D}_{TR}$ , which resembles the load responses we saw previously. Figures 9(b) and (c) present the extreme load levels estimated by the two 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 yields wider confidence intervals than the spline method. Furthermore, the degree of overestimation appears to increase as the probability corresponding to an extreme quantile value goes smaller. This observation confirms what we observed in Section 5.4 using the field data. This simulation result suggests that using the binning method for extreme load estimation is not a good practice.

**6. Summary.** This study presents a Bayesian spline method for estimating the extreme load on wind turbines. The spline method essentially supports a non-homogeneous GEV 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 in-land wind turbine load response data and making comparisons with the binning method.

The popularity of the binning method in industrial practice is due to the simplicity of its idea and procedure. However, simplicity of a procedure

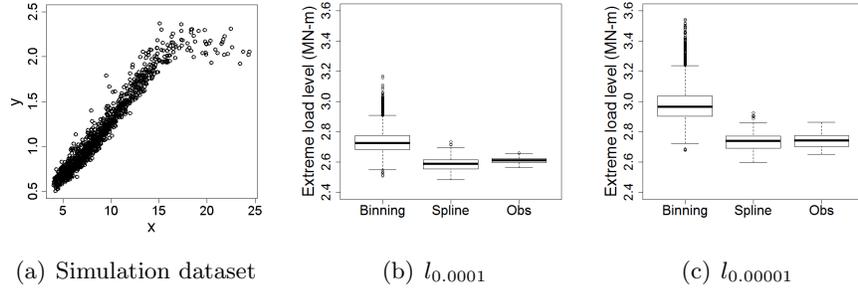


FIG 9. *Simulation dataset, estimated and observed extreme quantile values: (a) An example of a simulated dataset. (b) and (c) Boxplots for the distribution of the binning estimate, the Bayesian spline estimate, and the respective sample quantile across 100 simulated datasets.*

should not be mistaken as simplicity of a model. Suppose that one uses a  $6 \times 10$  grid to bin the two-dimensional wind covariates (as we did in this study) and fixes the shape parameter  $\xi$  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  $\xi$  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 RJS algorithm is between 12 and 18. The number of model parameters  $d_k$  in (4.11) is 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.

## References.

- 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.
- BOTTASSO, C. L., CAMPAGNOLO, F. and CROCE, A. (2010). Computational procedures for the multi-disciplinary constrained optimization of wind turbines, Technical Report, Dipartimento di Ingegneria Aerospaziale. Available at <http://www.aero.polimi.it/bottasso/DownloadArea.htm>.
- CARTA, J. A., RAMIREZ, P. and VELAZQUEZ, S. (2008). Influence of the level of fit a density probability function to wind-speed data on the WECS mean power output estimation. *Energy Conversion and Management* **49** 2647-2655.
- COLES, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, New York.
- DENISON, D. G. T., MALLICK, B. K. and SMITH, A. F. M. (1998). Bayesian MARS. *Statistics and Computing* **8** 337-346.

- DENISON, D. G. T., HOLMES, C. C., MALLICK, B. K. and SMITH, A. F. M. (2002). *Bayesian Methods for Nonlinear Classification and Regression*. Wiley, Chichester, UK.
- FITZWATER, L. M., CORNELL, C. A. and VEERS, P. S. (2003). Using environmental contours to predict extreme events on wind turbines. In *Proceedings of the 2003 ASME Wind Energy Symposium, AIAA Paper-2003-865*. Reno, Nevada.
- FITZWATER, L. M. and WINTERSTEIN, S. R. (2001). Predicting design wind turbine loads from limited data : comparing random process and random peak models. In *Proceedings of the 2001 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.
- GNEITING, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association* **106** 746-762.
- GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711-732.
- IEC, (1999). *IEC 61400-1 Ed 2: Wind Turbines-Part1: Design Requirements*. International Electrotechnical Commission, Geneva, Switzerland.
- IEC, (2005). *IEC 61400-1 Ed 3: Wind Turbines-Part1: Design Requirements*. International Electrotechnical Commission, Geneva, Switzerland.
- KASS, R. E. 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.
- LI, G. and SHI, J. (2010). Application of Bayesian model averaging in modeling long-term wind speed distributions. *Renewable Energy* **35** 1192-1202.
- MANUEL, L., VEERS, P. S. and WINTERSTEIN, S. R. (2001). Parametric models for estimating wind turbine fatigue loads for design. *ASME Journal of Solar Energy Engineering* **123** 346-355.
- MORIARTY, P. (2008). Database for validation of design load extrapolation techniques. *Wind Energy* **11** 559-576.
- MORIARTY, P., HOLLEY, W. E. and BUTTERFIELD, S. (2002). Effect of turbulence variation on extreme loads prediction for wind turbines. *ASME Journal of Solar Energy Engineering* **124** 387-395.
- NATARAJAN, A. and HOLLEY, W. E. (2008). Statistical extreme load extrapolation with quadratic distortions for wind turbines. *ASME Journal of Solar Energy Engineering* **130** 031017:1-7.
- PEERINGA, J. M. (2003). Extrapolation of extreme responses of a multi megawatt wind turbine, Technical Report, Energy Research Centre of the Netherlands. Available at <http://www.ecn.nl/docs/library/report/2003/c03131.pdf>.
- PEERINGA, J. M. (2009). Comparison of extreme load extrapolations using measured and calculated loads of a MW wind turbine, Technical Report, Energy Research Centre of the Netherlands. Available at <http://www.ecn.nl/docs/library/report/2009/m09055.pdf>.
- RAFTERY, A. E. (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. *ASME Journal of Solar Energy Engineering* **130** 031011:1-15.
- RONOLD, K. O. and LARSEN, G. C. (2000). Reliability-based design of wind-turbine rotor blades against failure in ultimate loading. *Engineering Structures* **22** 565-574.
- SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6** 461-464.

- SMITH, R. L. (1990). Extreme value theory. In *Handbook of Applicable Mathematics* **7** 437-471. John Wiley, England.
- SØRENSEN, J. D. and NIELSEN, S. R. K. (2007). Extreme wind turbine response during operation. *Journal of Physics, Conference Series* **75** 012074:1-7.
- VEERS, P. S. and BUTTERFIELD, S. (2001). Extreme load estimation for wind turbines: Issues and opportunities for improved practice. In *Proceedings of the 2001 ASME Wind Energy Symposium, AIAA Paper-2001-0044*. Reno, Nevada.
- WINDDATA, Available at <http://www.winddata.com>, Accessed July-2010.

## APPENDIX A: PRIORS

In this appendix, we specify priors for parameters used in the basis functions as follows:

$$\phi = (K, \Lambda_2, \dots, \Lambda_K)$$

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

$$p(K) = \frac{1}{n}, \quad K = \{1, \dots, n\}$$

$$p(T_k) = \begin{cases} 1, & T_k = \{1\} & \text{for } \phi_\eta \text{ and } \phi_\delta, \\ \frac{1}{2}, & T_k = \{1, 2\} & \text{for } \phi_\mu \text{ in ILT2 and all } \phi_\sigma, \\ \frac{1}{3}, & T_k = \{1, 2, 3\} & \text{for } \phi_\mu \text{ in ILT1 and ILT3,} \end{cases}$$

$$p(h_{.k}) = \frac{1}{2}, \quad h_{.k} = \{+1, -1\}$$

$$p(t_{.k}) = \frac{1}{n}, \quad t_{.k} = \{v_1, \dots, v_n\} \text{ or } \{s_1, \dots, s_n\}$$

## APPENDIX B: IMPLEMENTATION DETAILS OF THE SPLINE METHOD

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

1. *Step I*: construct the posterior 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  $t$ ,  $K_\mu$  and  $K_\sigma$  are equal to the number of basis functions specified in  $\phi_\mu^{(t)}$  and  $\phi_\sigma^{(t)}$ . Find the MLEs of  $\beta^{(t)}, \theta^{(t)}, \xi^{(t)}$

and the inverse of the negative of Hessian matrix, given  $\phi_\mu^{(t)}$  and  $\phi_\sigma^{(t)}$ .

- (c) Generate  $u_\mu^1$  uniformly on  $[0, 1]$  and choose a move in the RJS 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  $\phi_\mu^* = \text{BIRTH-proposal}(\phi_\mu^{(t)})$ , which is to augment  $\phi_\mu^{(t)}$  with a  $\Lambda_{K_\mu+1}^\mu$  that is selected uniformly at random;
  - Else if  $(b_{K_\mu} \leq u_\mu^1 \leq b_{K_\mu} + r_{K_\mu})$  then go to DEATH step, denoted by  $\phi_\mu^* = \text{DEATH-proposal}(\phi_\mu^{(t)})$ , which is to remove from  $\phi_\mu^{(t)}$  with a  $\Lambda_k^\mu$  where  $2 \leq k \leq K_\mu$  that is selected uniformly at random;
  - Else go to MOVE step, denoted by  $\phi_\mu^* = \text{MOVE-proposal}(\phi_\mu^{(t)})$ , which first do  $\phi_\mu^\dagger = \text{DEATH-proposal}(\phi_\mu^{(t)})$  and then do  $\phi_\mu^* = \text{BIRTH-proposal}(\phi_\mu^\dagger)$ .
- (d) Find the MLEs  $(\beta^*, \theta^*, \xi^*)$  and the inverse of the negative of Hessian matrix, given  $\phi_\mu^*$  and  $\phi_\sigma$ .
- (e) Generate  $u_\mu^2$  uniformly on  $[0, 1]$  and compute the acceptance ratio  $\alpha_\mu$  in (4.13), using the results from (b) and (d).
- (f) Accept  $\phi_\mu^*$  as  $\phi_\mu^{(t+1)}$  with probability  $\min(\alpha_\mu, 1)$ . If  $\phi_\mu^*$  is not accepted, let  $\phi_\mu^{(t+1)} = \phi_\mu^{(t)}$ .
- (g) Generate  $u_\sigma^1$  uniformly on  $[0, 1]$  and choose a move in the RJS 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  $\phi_\sigma^* = \text{BIRTH-proposal}(\phi_\sigma^{(t)})$ , which is to augment  $\phi_\sigma^{(t)}$  with a  $\Lambda_{K_\sigma+1}^\sigma$  that is selected uniformly at random;
  - Else if  $(b_{K_\sigma} \leq u_\sigma^1 \leq b_{K_\sigma} + r_{K_\sigma})$  then go to DEATH step, denoted by  $\phi_\sigma^* = \text{DEATH-proposal}(\phi_\sigma^{(t)})$ , which is to remove from  $\phi_\sigma^{(t)}$  with a  $\Lambda_k^\sigma$  where  $2 \leq k \leq K_\sigma$  that is selected uniformly at random;
  - Else go to MOVE step, denoted by  $\phi_\sigma^* = \text{MOVE-proposal}(\phi_\sigma^{(t)})$ , which first do  $\phi_\sigma^\dagger = \text{DEATH-proposal}(\phi_\sigma^{(t)})$  and then do  $\phi_\sigma^* = \text{BIRTH-proposal}(\phi_\sigma^\dagger)$ .

- (h) Find the MLEs  $(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*, \xi^*)$  and the inverse of the negative of Hessian matrix, given  $\phi_\mu^{t+1}$  and  $\phi_\sigma^*$ .
  - (i) Generate  $u_\sigma^2$  uniformly on  $[0, 1]$  and compute the acceptance ratio  $\alpha_\sigma$  in (4.14), using the results from (d) and (h).
  - (j) Accept  $\phi_\sigma^*$  as  $\phi_\sigma^{(t+1)}$  with probability  $\min(\alpha_\sigma, 1)$ . If  $\phi_\sigma^*$  is not accepted, let  $\phi_\sigma^{(t+1)} = \phi_\sigma^{(t)}$ .
  - (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 calculate a sample of  $\mu$  and  $\sigma$  using (4.4) and (4.5), respectively, for each pair of the  $N_w \times N_{sw}$  samples of  $(v, s)$  obtained in Step II. This generates  $N_w \times N_{sw}$  samples of  $\mu$  and  $\sigma$ .
  - (m) Draw  $N_l$  samples for the 10-minute maximum load  $\tilde{y}$  from each GEV distribution with  $\mu_i$ ,  $\sigma_i$ , and  $\xi_i$ ,  $i = 1, \dots, N_w \times N_{sw}$ , where  $\mu_i$  and  $\sigma_i$  are among  $N_w \times N_{sw}$  samples obtained in (l), and  $\xi_i$  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 4.3.
  - (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 MLEs 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 RJS algorithm again, namely (a) through (k) in Step I, to get one posterior sample of  $\Psi_\eta = (\beta_\eta, \phi_\eta)$  and  $\Psi_\delta = (\theta_\delta, \phi_\delta)$ .
- (f) Take the posterior sample of  $\Psi_\eta$  and  $\Psi_\delta$ , obtained in (e), and calculate a sample of  $\eta$  and  $\delta$  using (4.16) for each sample of  $v$ . This generates  $N_w$  samples of  $\eta$  and  $\delta$ .
- (g) Draw a sample for the standard deviation of wind speed  $\tilde{s}$  from each truncated normal distribution with  $\eta_i, \delta_i, i = 1, \dots, N_w$ . Using the  $N_w$  samples of  $\eta$  and  $\delta$  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  $N_l = 100$ .

#### APPENDIX C: 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  $\mu$  and  $\sigma$  (recall that  $\xi$  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  $\mu$  and  $\sigma$  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  $\mu$  and  $\sigma$  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  $\tilde{y}$ 's from all bins are pooled together to estimation  $l_T$ .

Specially, we go through the following steps, where  $\Phi_c$  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  $p[\tilde{v}, \tilde{s} | \mathcal{D}_v, \mathcal{D}_s]$  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  $\mu$  and  $\sigma$  for that specific bin from a multivariate normal distributions taking the MLE as its mean

and the inverse of the negative of Hessian matrix as its covariance matrix. Not all the bins have data. For those which do not have data, its  $\mu$  and  $\sigma$  are a weighted average of all non-empty bins with the weight related to the inverse squared distance between bins, as we explained above. Collectively,  $\Phi_c$  contains all the  $\mu$ 's and  $\sigma$ '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  $\mu$  and  $\sigma$  in  $\Phi_c$  is chosen; doing this yields the short-term distribution  $p[\tilde{y}|\tilde{v}, \tilde{s}, \Phi_c]$  for that specific bin.
- Draw  $N_l$  samples of  $\tilde{y}$  from  $p[\tilde{y}|\tilde{v}, \tilde{s}, \Phi_c]$  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  $l_T[\Phi_c]$  corresponding to  $P_T$ .
- Repeat the above procedure  $M_l$  times to get the median 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 spline method's implementation.

DEPARTMENT OF INDUSTRIAL AND  
SYSTEMS ENGINEERING,  
TEXAS A&M UNIVERSITY,  
COLLEGE STATION, TX 77843-3131  
E-MAIL: [giwhyun@tamu.edu](mailto:giwhyun@tamu.edu)  
[ntaimo@tamu.edu](mailto:ntaimo@tamu.edu)  
[yuding@iemail.tamu.edu](mailto:yuding@iemail.tamu.edu)

DEPARTMENT OF INDUSTRIAL AND  
OPERATIONS ENGINEERING,  
UNIVERSITY OF MICHIGAN,  
ANN ARBOR, MI 48109  
E-MAIL: [ebyon@umich.edu](mailto:ebyon@umich.edu)