Statistical Extrapolation Methods for Estimating Wind Turbine Extreme Loads

Patrick Ragan^{*} and Lance Manuel[†]

Dept. of Civil, Architectural, and Environmental Engineering, University of Texas, Austin, TX, 78712, USA

With the introduction of the 3rd edition of the International Electrotechnical Commission (IEC) Standard 61400-1, designers of wind turbines are now explicitly required, in one of the prescribed load cases, to use statistical extrapolation techniques to determine nominal design loads. In this study, we use field data from a utility-scale 1.5MW turbine sited in Lamar, Colorado to compare the performance of several alternative techniques for statistical extrapolation of rotor and tower loads-these include the method of global maxima, the peak-over-threshold method, and a four-moment process model approach. Using each of these three options, fifty-year return loads are estimated for the selected wind turbine. We conclude that the peak-over-threshold method is the superior approach, and we examine important details intrinsic to this method, including selection of the level of the threshold to be employed, the parametric distribution used in fitting, and the assumption of statistical independence between successive peaks. While we are primarily interested in the prediction of extreme loads, we are also interested in assessing the uncertainty in our predictions as a function of the amount of data used. Towards this end, we first obtain estimates of extreme loads associated with target reliability levels by making use of all of the data available, and then we obtain similar estimates using only subsets of the data. From these separate estimates, conclusions are made regarding what constitutes a sufficient amount of data upon which to base a statistical extrapolation. While this study makes use of field data in addressing statistical load extrapolation issues, the findings should also be useful in simulation-based attempts at deriving wind turbine design load levels where similar questions regarding extrapolation techniques, distribution choices, and amount of data needed are just as relevant.

I. Introduction

Given a limited amount of field data, our goal is to use statistical extrapolation techniques to predict 50-year Greturn levels of wind turbine rotor and tower loads. This is essentially the same task that is currently being required of turbine designers in Design Load Case 1.1 of IEC Standard 61400-1, 3rd edition¹, where extrapolation is to be applied with simulated loads data. Although Annex F of the IEC guidelines makes reference to the study by Moriarty et al², where peak-over-threshold extrapolations are demonstrated using several different distributions, details of the extrapolation procedure are left to the designer's discretion. In this study, we compare results obtained using several different extrapolation procedures, with the goal of recommending which procedure might best be suited in a given application.

Our first objective is to compare three fundamentally different approaches for extracting the information to be used in loads extrapolation from raw time-series data. In order of increasing use of data, they are:

- Method of Global Maxima In this method, only the single largest data point (load) from each ten-minute file is used, and statistical distributions for these ten-minute maxima are estimated directly.
- Peak-Over-Threshold (POT) Method Multiple peaks are extracted from each file. Specifically, the largest value between each successive upcrossing of the threshold is extracted. Distributions are fit to load exceedances over the selected threshold.
- Process Model Approach The entire time history is modeled as a random process, using the first four statistical moments and the mean crossing rate of the observed time series. Unlike the other methods, this

^{*}Graduate Research Assistant

[†]Associate Professor

approach does not explicitly discard any of the data in the time series, although it has the disadvantage of not directly modeling the largest peaks.

Refer to the simulation study by Fitzwater and Winterstein³ for further discussion on these three approaches.

After a comparison between these three methods is made, specifics of the peak-over-threshold method are examined in greater detail. The selection of the threshold level to be employed is discussed at length, and a procedure is described by which this selection may be optimized. The question of which parametric distribution may be most appropriate is also explored, both on theoretical and practical grounds. Finally, we discuss the effect of implementing a minimum time separation on load peaks over the selected threshold for the purpose of ensuring greater statistical independence between successive peaks, and we examine the implications of neglecting this issue.

After discussing all of these details regarding the peak-over-threshold method, it is then used in a study of the variability of long-term load predictions as a function of the amount of data included in the analyses. Standard errors on long-term loads predicted using subsets of the data are calculated and compared with predictions made using the entire dataset. Then, conclusions are made about the size of dataset needed to produce reliable statistical extrapolation of loads for design.

Importantly, this study uses rarely available field data on loads from a utility-scale 1.5MW turbine to demonstrate various loads extrapolation techniques. Since field measurement campaigns are expensive, it is more common that simulation studies are employed in load extrapolation in practice. Insights related to statistical extrapolation provided by our analyses using field data are however still relevant in situations where such loads data are obtained from simulation, though some differences may be expected. For instance, inflow turbulence character is likely to be more variable in field measurements and simplifying assumptions (in stationarity and coherence, for example) are generally made in spectral models for turbulence used in simulation. Also, load control algorithms may be quite differently represented in the field than in models used in simulation. Notwithstanding these differences, a critical study of load extrapolation techniques is of interest to analysts and the availability of useful field data motivates the present study.



Figure 1. The Instrumented 1.5MW Wind Turbine.

II. Experimental Data

The subject of this study is a utility-scale 1.5MW wind turbine located at a Great Plains site near Lamar, Colorado (see Fig. 1). The turbine has a hub height of 80 meters, and a rotor diameter of 70.5 meters. Approximately 17,000 ten-minute records were taken over a period of roughly four months between September 2004 and January 2005 (Zayas et al^4). A total of 67 channels at a sampling rate of 40 Hz provided various measurements of the turbine's inflow, control state, and structural response. For the purposes of this study, the following three measurements will be the subject of the extreme loads estimation:

- Edge bending moment at the blade root (EBM)
- Flap bending moment at the blade root (FBM)
- Resultant bending moment at the tower base (TBM).

For various reasons, much of the original time series data was unusable for our purposes. Of the original dataset of roughly 17,000 ten-minute records, a total of 2,485 were used in this study. Figure 2 shows the distribution of this available data binned according to hub-height ten-minute mean wind speed. Extreme load distributions will be estimated both for data in individual wind speed bins as well as for the aggregated data resulting from

integrating the bin-specific distributions. The dataset is unfortunately missing a large proportion of files around the rated wind speed (12 m/s), which means that estimated long-term load distributions will not be as heavily influenced by the near-rated wind speed bins as one might expect if a full dataset were available.

III. The Method of Global Maxima

A. The Generalized Extreme Value Distribution

We begin by demonstrating statistical extrapolation using the simplest approach—the method of global maxima—in which the Generalized Extreme Value (GEV) distribution is fit to the single largest load value from each ten-minute file. First, the theoretical background of the GEV is discussed briefly.



Figure 2. Distribution of the Available Ten-Minute Data Records, Binned by Hub-Height Mean Wind Speed.

We seek a statistical distribution for a random variable which is defined as the maximum of the n (equal to 24,000 here) random variables that make up a ten-minute time series for data sampled at 40 Hz. It may be shown (see Gumbel⁵) that as n approaches infinity, the maximum of the n random variables follows one of three Extreme Value distributions. The GEV distribution comprises these three Extreme Value distributions; so one can say that the maximum of n random variables follows the GEV distribution. Since this is only strictly true as n approaches infinity, we say the GEV is the correct distribution in an asymptotic sense. A commonly made analogy to this result is with the more

familiar Central Limit Theorem which states that the sum of *n* random variables (under certain conditions) follows a normal distribution in an asymptotic sense.

Mathematically, the Generalized Extreme Value Distribution is defined as follows:

$$F(x) = \exp\left[-\left(1-k\left(\frac{x-u}{a}\right)\right)^{1/k}\right], \ k \neq 0; \quad F(x) = \exp\left[-\exp\left(-\left(\frac{x-u}{a}\right)\right)\right], \ k = 0$$
(1)

where u, a, and k are location, scale, and shape parameters, respectively, of the distribution such that $-\infty < u < \infty$, $0 < a < \infty$, and $-\infty < k < \infty$. The shape factor, k, is of particular importance as its value defines the GEV distribution to be of Type 1, Type 2, or Type 3 (see Table 1). The distribution for k = 0 is the well-known Gumbel distribution, and it is easily shown that its formula results from a limiting argument with the formula for the case when $k \neq 0$.

Table 1. Properties of the GEV distribution Types 1, 2, and 3.

GEV Distribution Type	Name	Shape Factor	Lower Bound	Upper Bound
1	Gumbel Distribution	k=0	None	None
2	Frechet Family	<i>k</i> < 0	a/k + u	None
3	Weibull Family	$k \ge 0$	None	a/k + u

The use of the GEV distribution is well-established for modeling extremes of natural phenomena such as wind speeds and flood levels. In particular, the Gumbel distribution (Type 1 GEV, k = 0) is commonly used in various applications; for example, it was recommended in 1975 by the British National Environmental Research Council for use in extreme value problems in hydrology. Maximum values for random variables describing many physical processes, however, are likely to have upper bounds and may be more accurately represented by the generic GEV distribution which does not force k to be equal to zero. This is the argument made by Holmes and Moriarty⁶ for models of extreme wind gusts. It could similarly be argued that an upper-bound model is even more appropriate for wind turbine loads because of the limiting influence that modern wind turbine control systems are generally designed to have on loads. In one wind turbine-related study, Pandey and Sutherland⁷ showed that the Gumbel model was consistently conservative in predicting extreme loads using extrapolation.

Following the work of Hosking et al⁸, the Method of Probability Weighted Moments (or, equivalently, the Method of L-Moments) has become the most common approach used for GEV distribution parameter estimation. Using this procedure, GEV distribution fits were attempted to the binned load extremes data for EBM, FBM, and TBM (see Fig. 3). The distribution functions for all the wind speed bins have been plotted on a Gumbel scale which means that the case for k = 0 will appear linear, k < 0 shows negative curvature, and k > 0 shows positive curvature. Note that when a k > 0 fit is found, an upper bound represented by the light vertical dashed lines is indicated (as, for example, with the FBM load in the 13-15 m/s bin). Note that all values of the loads (bending moments here) have been normalized with respect to the largest value observed during the measurement campaign.

B. Integration of Conditional Load Distributions for Long-Term Loads

The individual GEV load distributions, conditional on wind speed, are used to estimate the overall probability that a random ten-minute maximum load, $M_{10 \text{ min}}$, is greater than any specified load level, x. This probability of exceedance is calculated as follows:



Figure 3. GEV distribution fits to global maxima for different wind speed bins.

4 American Institute of Aeronautics and Astronautics

$$1 - F(x) = P(M_{10\min} > x) = \int_{V} (1 - F(x | V)) f(V) dV$$
⁽²⁾

where F(x|V) is the cumulative conditional GEV distribution for the wind speed V, and f(V) is the probability density function for V. Practically speaking, since we have discretized V into 2 m/s wind speed bins, this integral reduces to:

$$1 - F(x) = P(M_{10\min} > x) = \sum_{V} (1 - F(x | V)) p(V)$$
(3)

 Table 2. Relative likelihoods of different wind speed bins based on a Rayleigh distribution.

Wind Speed Bin, $V(m/s)$	p(V)
<5	0.178
5-7	0.141
7-9	0.151
9-11	0.143
11-13	0.121
13-15	0.094
15-17	0.068
17-19	0.045
19-25	0.051
>25	0.007

where p(V) is the probability associated with the relevant wind speed bin. Values of p(V) were calculated based on a Rayleigh distribution of ten-minute mean wind speeds with a mean value of 9.71 m/s (see Table 2). It is assumed, for the purposes of this study, that only extreme loads during operation are of interest. Equivalently, we assume here that extreme loads cannot occur at wind speeds below cut-in (5 m/s) or above cutout (25 m/s), so that $F(x|V \le 5)$ and $F(x|V \ge 25)$ are each taken to be equal to unity. Results of the integration of the conditional load distributions for all three load types are summarized in Fig. 4.

It is assumed, for the purposes of this study, that only extreme loads during operation are of interest. Equivalently, we assume here that extreme loads cannot occur at wind speeds below cut-

in (5 m/s) or above cut-out (25 m/s), so that $F(x|V \le 5)$ and $F(x|V \ge 25)$ are each taken to be equal to unity. Results of the integration of the conditional load distributions for all three load types are summarized in Fig. 4.

Equation (3) may also be used to calculate any *R*-year return load, which is the load that is exceeded, on average, once every *R* years. Assuming that the occurrence of these large, rarely occurring loads is governed by a Poisson process, the waiting time between occurrences is exponentially distributed with mean *R*, so that the probability that the *R*-year return load is exceeded in a period of duration, *t*, is $[1-\exp(-t/R)]$. If *t* is very much smaller than *R*, this probability is very nearly equal to t/R, so that the probability that the 50-year return load is exceeded in a period of 10 minutes is $10/(50 \times 365.25 \times 24 \times 60) = 3.8 \times 10^{-7}$. Substituting this value in the left hand side of Eq. (3) yields the 50-year return load as the value of *x*.



Figure 4. Probability of load exceedances (integrated over all wind speeds) based on the method of global maxima.

Table 3 shows results of return load calculations at the 17-day, 1-year, 10-year, and 50-year levels. The 50-year load is of interest because it is used as the nominal design load in the IEC guidelines¹; the 17-day load has been included because our 2,485 ten-minute records are roughly equivalent to 17 days of continuous data, so we would expect the extrapolated load at the 17-day level to be close to the maximum observed normalized load, 1.0.

 Table 3. Loads associated with different return periods based on the method of global maxima.

Load Type	$L_{17 \text{ day}}$	$L_{1 \text{ yr}}$	$L_{10 \text{ yr}}$	$L_{50 \text{ yr}}$
FBM	0.95	1.02	1.08	1.13
EBM	0.97	1.02	1.04	1.05
TBM	0.99	1.26	1.59	1.83

C. Discussion of Results

Examining first the long-term results in Fig. 4 and Table 3, our initial impression is that the extrapolated tower bending loads are unrealistically high, while the flap and edge loads on the blade are somewhat lower than we would expect. What is most objectionable about these results, however, is that the lower wind speed bins are dominating the extrapolation, which is physically unrealistic. Table 4 demonstrates this best by displaying the relative contributions from each wind speed bin to the summation in Eq. (3) at the 5-year level. Referring back to Fig. 3, some of the unexpected behavior indicated above can be attributed to overestimation of long-term loads in the lower

wind speeds (5-7 and 7-9 m/s bins) at least for FBM and TBM, and systematic underestimation of long-term loads for the higher wind speed bins.

this load occurs (global maxima method).								
Wind Speed Bin (m/s)	FBM	EBM	TBM					
5-7	0.0653	0.0000	0.0004					
7-9	0.8218	0.0000	0.9996					
9-11	0.1130	0.5525	0.0000					
11-13	0.0000	0.0000	0.0000					
13-15	0.0000	0.4475	0.0000					
15-17	0.0000	0.0000	0.0000					
17-19	0.0000	0.0000	0.0000					
19+	0.0000	0.0000	0.0000					

Table 4.	Probability	that the der	rived 50-yr	load	arises
fr	om different ^v	wind speed	bins given	that	
41.	to load a server	. (alah al		(h a d	

The distributions for FBM and TBM in the two lowest wind speed bins appear much flatter than the others; this might also suggests that the observed loads data for these bins were perhaps not large enough to be affected by the limiting influence of the turbine's controller. Information about the way in which the controller might prevent excessive loads on the turbine is not included in these distributions, and the extrapolations overestimate long term-loads as a result.

Loads data for the higher wind speed bins, on the other hand, appear to have been affected a great deal by the controller, as evidenced by the strong positive curvatures of the distributions in Fig. 3. These distributions have the opposite problem to those for the

lower wind speed bins, as they consistently underestimate long-term loads. This underestimation is painfully obvious in some cases—for example, in some bins, the predicted upper bound by the model (shown by the light vertical dashed line) is actually lower than several *observed* values for that bin, which means that such fitted GEV distributions based on the global maxima have a zero probability of being correct. Closer inspection revealed that these large observed loads that exceed the GEV-predicted upper bounds occurred in rare, but real, conditions, usually accompanied by particularly strong wind gusts which the controller is unable to respond to quickly enough. Also, these large loads occurred too infrequently in the sample to influence the fitted distributions in a marked way so as to raise the return load predictions to higher, more appropriate levels.

The short story, though, is that the method of global maxima did not perform well with these loads data. We could point to the GEV distributions as the source of the problem, and seek other distribution types or parameter estimation methods that might improve the results. Parameter estimation via maximum likelihood methods, for instance, might at least prevent fitted distributions from having upper bounds below an observed data point. However, we do not pursue such options here because we believe that a more significant drawback to the global maximum method described here lies more likely in the fundamental inefficiency of employing only one peak load value from each ten-minute segment of data and attempting to use that to gain long-term load information.

IV. Peak-Over-Threshold Method

The peak-over-threshold method has an advantage over the global maximum method of being able to make use of multiple peaks from some ten-minute time series files—those with higher loads—and perhaps none from other files during which lower, less relevant loads were recorded. By choosing a load threshold level that is sufficiently high, the method selectively extracts the largest, most important peaks from the dataset, allowing extrapolation to be based on a relevant and homogeneous sample of data points.

A. Choice of Statistical Distribution

The Generalized Pareto distribution (GPD) was introduced by Pickands⁸, who demonstrated that it is the appropriate distribution for exceedances over a threshold under the same basic conditions for which the GEV is the appropriate distribution for global maxima. The only additional assumptions are that upcrossings of the threshold are Poisson and, thus, the exceedances themselves are independent. Provided that these conditions are true, the GPD and GEV are intimately related, and even share the same shape factor, k.

Mathematically, the Generalized Pareto distribution is defined by

$$F(x) = 1 - \left(1 - \frac{kx}{a}\right)^{1/k}, \ k \neq 0; \quad F(x) = 1 - \exp\left(\frac{-x}{a}\right), \ k = 0$$
(4)

where x is the amount of load exceedance over a chosen threshold, u. The parameters a and k are the scale factor and the shape factor, respectively, with $0 < a < \infty$ and $-\infty < k < \infty$. The shape factor, k, carries the same importance as with the GEV distribution; its value defines the nature of the upper tail and the existence or not of an upper bound. The case with k = 0 is the exponential distribution with mean value, a, and the distribution formula then follows from a limiting argument applied to the $k \neq 0$ case. In all cases, the distribution has a lower bound of zero since any value that does not exceed the threshold, u, is by definition excluded from the distribution. The properties of all 3 GPD types are summarized in Table 5.

GPD Type	Name	Shape Factor	Lower Bound	Upper Bound
1	Gumbel	k=0	0	None
2	Frechet Family	$k \le 0$	0	None
3	Weibull Family	<i>k</i> > 0	0	a/k

Table 5. Properties of the Generalized Pareto Distribution Types 1, 2, and 3.

the GPD has been discussed in detail by Hosking et al⁹, where the method of probabilityweighted moments is compared to the conventional method of moments as well as

Parameter estimation for

to the method of maximum likelihood. The conventional method of moments is the simplest to apply and is shown to outperform the other methods for distributions whose shape factor, k, is near or slightly above zero, as is the case for most of the fits associated with our data.

The GPD is the standard for POT analysis in many applications such as for the prediction of flood levels and wind gusts. See, for instance, Davison and Smith¹⁰, Holmes and Moriarty⁶, Brabson and Palutikof¹¹, or Ryden¹². Despite this as well as its strong theoretical foundation, it has not been systematically employed to extrapolate structural loads on a wind turbine from POT data.

The Weibull 3-Parameter distribution (W3P) is another option for use with POT data, and was one of several employed in the study by Moriarty et al^2 . The W3P distribution is given as follows:

$$F(x) = 1 - \exp\left(-\left(\frac{x - u_0}{a}\right)^k\right)$$
(5)

It is important that the location parameter, u_0 , in Eq. (5) is not confused with the selected threshold, u. Although there is no theoretical justification for using the W3P for peak-over-threshold data, it does have the advantage of offering greater flexibility than the GPD, since it has three parameters instead of two. (Specifically, with the method of moments, for example, this enables the sample skewness of the data to influence the shape of the fitted distribution.) A detailed comparison of long-term load predictions based on the GPD and the W3P will be presented in a subsequent section. The results shown in the remainder of this section are based on the W3P, with parameter estimation using the method of moments (see Moriarty et al² for details).

B. Return Loads and Integration of Long-Term Distributions

Once a distribution has been fit to the peaks above a selected threshold within a wind speed bin, a bin-specific *R*-year return load can be found by using Eq. (5) to solve for the excess over the threshold, *x* for the desired probability. This requires setting the fractile level, $F = 1 - 1/(\lambda R)$, where λ is the crossing rate of the chosen threshold so that λR is the number of peaks above the threshold expected in *R* years. The desired load level itself, x_R , is the excess added back to the original threshold,

$$x_R = a \left[\ln(\lambda R) \right]^{1/k} + u_0 + u \tag{6}$$

A bin-specific return load is the load that would be expected to occur at an average rate of once every R years if the wind conditions in that given wind speed bin remained constant at all times. Despite this awkward physical interpretation, such values can be useful in comparing extrapolated results from different bins. Of course, what we are ultimately interested in is the long-term distribution integrated over all bins and the overall return loads derived from this integrated load distribution. For this, we use a calculation similar to Eq. (3), yielding

$$1 - F(x) = P(M_t > x) = \sum_{V} P(M_t > x | V) p(V) = \sum_{V} (1 - (F_V(x - u_V))^{\lambda_t}) p(V)$$
(7)

where the random variable M_t is the maximum load in time t (equal to 10 minutes here), F_V is the (conditional) distribution function for POT loads in bin V, u_V is the threshold chosen for bin V, and λ_V is the crossing rate of u_V in bin V.

C. Choice of Threshold

Before we begin the peak-over-threshold analysis, an obvious and important question relates to how the threshold level should be chosen. Figure 5 shows a sample flap bending moment (FBM) time series that has been reproduced three times. In each case, peaks have been extracted according to a different threshold; the total number of peaks, *n*, extracted in each case is shown above each plot. The top plot uses a threshold which is simply the mean value of the time series; the second plot uses a threshold at the level of the mean plus 1.4 times the standard deviation, and the bottom plot uses a threshold at the level of the mean plus 1.9 standard deviations.



Figure 5. Sample peak-over-threshold data for FBM in the 9-11 m/s bin: three alternative threshold levels.

The first threshold (i.e., the mean value) is clearly below what may be considered an ideal level for POT analysis. The most obvious problem is that many of the largest peaks in the time series-between around the 200 and 300 second marks-are not extracted and hence not retained as POT data. This is because the floor (troughs) of the time series over this period of time is above the threshold, which in turn means that there are no upcrossings of the threshold over this time segment (recall that only the largest value between successive upcrossings is defined as a peak over the threshold and hence extracted for further statistical analysis). In addition, between the 300 and 450 second marks, there are many peaks extracted that are not very near to the top of the time series (where we would expect important peaks to be) but are instead close to the bottom; these are the result of successive upcrossings of the threshold in the times series that are extremely close together, and are associated with minor peaks that we would prefer to omit from our extrapolation. Both of these problems

highlight the danger of choosing a threshold level that is too low—namely, the largest peaks, which we intuitively would like to retain for use in our extrapolation, are not all included, while many minor peaks, which we may not want to retain, are included. This results in a grossly heterogeneous dataset that makes fits to statistical distributions very difficult.

The bottom two plots in Fig. 5 do a much better job of retaining all of the largest peak load values, although the middle plot does pick up a few of the undesirable minor peaks around the 275 and 500 second marks. The threshold level in the bottom plot appears to be the most successful in terms of retaining only the most important largest peaks. Of course, if we use a threshold level that is too high, our sample of peaks might end up being too small and our extrapolation might become susceptible to sampling errors due to small POT sample size.

Still, based on our studies to date, we believe that it is important to use relatively high thresholds in order for distribution fits to avoid modeling errors associated with heterogeneous sets of peaks. But how high is high enough? Moriarty et al² suggest the μ + 1.4 σ level (corresponding to the middle plot in Fig. 5), reasoning that this is just below the level of peaks for a signal that is a deterministic sine wave, whose amplitude is exactly $\sqrt{2}$ (1.414) standard deviations above its mean. One would expect this threshold to work quite well for edge bending moment, which is affected most by the periodic influence of gravity loading. In some cases though, such as for the flap bending moment time series in Fig. 5, the process bears little resemblance to a deterministic sine wave, and there are perhaps more ideal threshold levels that can be chosen.

An often missed point related to the use of POT data in extrapolation is that it is necessary to choose a single threshold level that must be used for all the time series in a given wind speed bin because all of the peaks from these time series in the bin will be fit to the same distribution, which models the exceedance of peaks over that single selected threshold level. If we speak of thresholds in terms of the sample mean and standard deviation from an entire bin's data, the most appropriate threshold is even more likely to be greater than 1.4 standard deviations above the mean, because the larger peaks will likely come from time series (in the bin) whose ten-minute mean is higher than the mean for the entire bin. As an example, for the time series in Fig. 5 that comes from the 9-11 m/s wind speed bin, the ten-minute mean and standard deviation values are 0.427 and 0.111, respectively. For the entire 9-11 m/s bin (475 ten-minute files), the mean and standard deviation values are 0.391 and 0.100, respectively. So what we referred to as the $\mu + 1.4\sigma$ level for the threshold in Fig. 5 is actually, in the context of the entire bin, 1.91

standard deviations above the mean: $0.427 + 1.4 \times 0.11 = 0.391 + 1.91 \times 0.100$. Similarly, what we referred to as the μ + 1.9 σ level in Fig. 5 is actually, in the context of the entire bin, 2.46 standard deviations above the mean.

In any case, it is not obvious in advance which threshold should be chosen in order to extract a homogeneous set of the most relevant peaks that one might like to use in statistical extrapolation. For this reason, we do not select the threshold *a priori*. Instead, we search for an "optimal threshold," using the following procedure:



Figure 6. Comparison of distribution fits for various thresholds for FBM in the 9-11 m/s bin.



Figure 7. Summary of distribution results for various thresholds for FBM in the 9-11 m/s bin.

9 American Institute of Aeronautics and Astronautics

- 1. Choose a range of possible thresholds for each bin. We use 20 equally spaced thresholds between the 0.5 and 1.0 load levels; another option would be to use equally spaced thresholds between the $\mu + 1.4\sigma$ level and the maximum observed load value (or simulated value, if one is using simulations) for the bin.
- 2. For each threshold choice, peaks are extracted from all the time series in the bin and a distribution is fit. As discussed earlier, we use the Weibull 3-parameter distribution.
- 3. Use some goodness-of-fit criterion to determine the optimal threshold for the bin. In our case, we select the threshold for which the loads from the fitted distribution have the lowest mean square error with respect to the observed loads data at the same fractiles. To avoid making conclusions based on too small a dataset (sampling errors), threshold levels for which fewer than 10 peaks result are excluded from consideration.

This procedure is illustrated in Figs. 6 and 7 for flap bending moment in the 9-11 m/s bin. For this bin, the threshold level of 0.71 has the lowest mean squared error; hence, it is considered to be the optimal threshold. This threshold is 3.2 standard deviations above the bin mean of 0.39 and, interestingly, is so high that none of the peaks from the sample time series of Fig. 5 are extracted for this bin (implying only that the sample series that was shown in Fig. 5 had peak load levels that are not high enough to merit including them as representative of the bin in long-term load extrapolation; this is indeed the way less important data can and should be systematically omitted). The 0.53 threshold level corresponds to the $\mu + 1.4\sigma$ level and its use leads to a mean squared error that is roughly 10 times greater than that with the optimal threshold; moreover, it predicts a bin-specific 50-year return load of 1.29 versus 1.05 predicted by a fit using the optimal threshold.

D. Discussion of Results

Using the procedure described, optimal thresholds are algorithmically found for each of the eight wind speed bins and for each of the three load types resulting in the load distributions shown in Fig. 8. A quick visual scan of the plots gives the impression that the distributions generally fit the observed peaks very well, with the possible exception of the edge bending moment for the 7-9 m/s bin. An investigation into the various threshold levels attempted for this bin (i.e., viewing plots similar to Figs. 6 and 7) reveals that there are higher thresholds that appear to fit the data better but these thresholds have mean squared errors that are slightly higher than that for the optimal threshold used in the fit for Fig. 8. This suggests that a simple mean squared error may not always be the best measure of the goodness of fit for a distribution. An evaluation procedure that weights errors in the upper tail more heavily may be one alternative for a goodness-of-fit criterion; the Kolmogorov-Smirnov test, which focuses on the worst fit of any of the data points, is another. In our case, though, the repercussions of the one poor fit with the 7-9 m/s bin for edge bending moment are quite small—we do not expect the 7-9 m/s bin to influence the overall results very much, especially when the conditional load distributions for each bin are integrated by the relative likelihood of the bin in order to yield overall long-term loads.

Figure 9 along with Tables 6 and 7 summarize the final results from the peak-over-threshold analysis, and are analogous to Figure 4 and Tables 3 and 4, respectively, that summarized results from the global maxima analysis. The 50-year return loads from the POT approach appear to be at far more reasonable levels than were seen with the GEV distribution fits seen in the global maxima method, and the wind speed bins that contribute most to the target exceedance probability for the long-term 50-year loads are correctly determined to be the bins for which many of the largest observed loads did in fact occur. In short, the peak-over-threshold method with Weibull 3-parameter fits and optimal thresholds selected as described earlier, appears to show considerable promise as a reliable method to extrapolate long-term wind turbine loads.

V. Process Model Approach

For a Gaussian random process, X(t), with zero mean, unit variance, and mean upcrossing rate v, the maximum load in time *T* is a random variable with the following cumulative distribution function (representing the probability of non-exceedance of any specified level, *x*, of the process:

$$F(x) = \exp\left(-vT\exp\left(-x^2/2\right)\right) \tag{8}$$

For physical processes that are non-Gaussian, such as would be the case for the blade or tower bending moments on a wind turbine in general, the maximum load may be found by relating the non-Gaussian process to an associated Gaussian process using Hermite polynomials (and moments of the process that are of higher than second order). Following the work of Winterstein^{13,14}, we use a four-moment Hermite transformation, which allows us to estimate the distribution of a ten-minute maximum load given the first four moments and the mean upcrossing rate of the load process. See the cited references^{13,14} for further details.



Figure 8. Weibull 3-parameter fits to peak-over threshold data with optimal thresholds.

11 American Institute of Aeronautics and Astronautics



Figure 9. Probability of load exceedances (integrated over all wind speeds) based on the peak-over-threshold method.

Figure 10 shows an example of this method applied to a sample flap bending moment time series in the 15-17 m/s bin. Given the time series, we estimate the mean (μ), standard deviation (σ), skewness (α_3), kurtosis (α_4), and mean upcrossing rate (ν). As the histogram in Fig. 10 shows, this time series is not exactly Gaussian but neither is it dramatically different from a Gaussian process; its skewness and kurtosis values are -0.112 and 2.625, respectively, in comparison to 0 and 3.0 for a Gaussian process. Based on the four-moment Hermite transformation model, the distribution function for the ten-minute maximum of this process is obtained, as shown in the bottom plot of Fig. 10. The median (0.5 fractile) of the maximum distribution, 0.84, is very close to the

observed ten-minute maximum, 0.86, and the 50-year level (fractile level of 3.8×10^{-7}) is 1.01. This means that if the process remained stationary with the observed four moments and crossing rate, a load of only 1.01 would be expected to occur, on average, once every 50 years.

Table	6.	Loads	asso	ciated	with	different	return
periods	b٤	ised on	the	peak-o	ver-t	hreshold	method.

Load Type	$L_{17 \text{ day}}$	$L_{1 \text{ yr}}$	$L_{10 \text{ yr}}$	$L_{50 \text{ yr}}$
FBM	0.99	1.07	1.14	1.18
EBM	1.01	1.09	1.15	1.19
TBM	1.01	1.10	1.18	1.24

Although the results for this example time series seem reasonable, when the results over all of the observed time series are integrated, long-term load predictions are found to be overly conservative (see Table 8 which can be directly compared to Tables 3 and 6 from the global maximum and the POT methods, respectively). There are a number of problems that contribute to these conservative predictions, some of which are enumerated briefly below:

- Highly non-Gaussian processes (such as the EBM process) cannot be effectively represented by this method, which is intended to be used mainly for mild perturbations to Gaussian processes.
- There is a sizable region of skewness and kurtosis combinations (many of which occurred in our data sets) for which the Hermite transformation is non-monotonic (or nearly so), and cannot be used.
- The model is highly sensitive to non-stationarity (the wind turbine loads data often do not display desirable stationary characteristics).
- The statistical moments are highly variable, particularly so for the higher moments (skewness and kurtosis), and this is especially true since we are using field data. Large skewness and kurtosis values are often associated with unrealistically large load predictions.

For these various reasons, the process model approach is not recommended for the extrapolation of long-term wind turbine loads.

-			-
Wind Speed Bin (m/s)	FBM	EBM	TBM
5-7	0.0000	0.0000	0.0000
7-9	0.0000	0.0000	0.0000
9-11	0.0001	0.0075	0.0001
11-13	0.0000	0.0002	0.0002
13-15	0.0000	0.0140	0.8358
15-17	0.0007	0.0008	0.0000
17-19	0.7073	0.0246	0.0130
19+	0.2919	0.9529	0.1509

Table 7. Probability that the derived 50-yr load arises from different wind speed bins given that this load occurs (peak-over-threshold method).

VI. Investigation of Details related to the Peak-Over-Threshold Method

Of the three methods presented, the peak-overthreshold method has been found to be the superior choice for conducting statistical loads extrapolations. Therefore, we return to the subject of peak-overthreshold modeling to further investigate several of the details associated with this procedure.

A. Choice of Threshold Level

Justification for a procedure that uses an optimal threshold was presented earlier (in Section IV), as

were the overall results based on this procedure for the field data considered here. We now investigate how these results would change if a constant threshold of $\mu + 1.4\sigma$ were used instead, as suggested by Moriarty et al² and in Annex F of the IEC Standard 61400-1, Ed. 3¹.



Figure 10. Distribution function based on the four-moment process model (sample FBM time series in the 15-17 m/s bin).

Table 8.	Loads a	ssociated	with o	different	return
periods	based on	the proc	ess me	odel app	roach.

 L_{1v}

1.26

 $L_{10 y_1}$

1.42

L17 day

1.06

Load Type

FBM

Figures 11 and 12 demonstrate the difference in distribution fits for the 9-11 m/s and the 19+ m/s bins, respectively. Note first that the distributions for edge bending moment are reasonably similar with either threshold choice, which is not surprising given that the μ + 1.4 σ threshold was originally proposed with a deterministic sine wave in mind. For the FBM and TBM loads, however, the differences in fits due to the two thresholds are sometimes rather large, especially for the lower wind speed bin. The 9-11 m/s bin is intended to be representative of such lower wind speed bins while the 19+ m/s bin is intended to be representative of the higher wind speed bins, where the differences in fits are much smaller. Difference in the nature of the loads distributions for the low and high wind speed bins are also highlighted by studying Table 9 which shows that the optimal thresholds tend to be closer to the $\mu + 1.4\sigma$ level (though usually somewhat higher) for the higher wind speed bins while at lower wind speeds, the optimal threshold is at significantly higher levels. The results in Table 9 suggest that the optimal threshold for good fits is almost always greater than 1.4 standard deviations above the mean.



L<u>50 yr</u>

1.52

As a final comparison between fits based on use of an optimal threshold versus the constant $\mu + 1.4\sigma$ threshold, Table 10 shows overall return loads (integrated over all wind speeds) calculated using both approaches. The differences in the results for the FBM and EBM loads are not dramatic; the slightly larger FBM loads for the $\mu +$ 1.4σ threshold choice are associated with the over-predicting of loads for the 5-7 and 7-9 m/s bins. The overestimation of TBM in these bins is more severe, though, and the overall return loads for TBM with the $\mu + 1.4\sigma$ threshold are, as a result, overly conservative especially at the longer return periods.



a μ + 1.4 σ threshold for the 19+ m/s bin.

B. Choice of Distribution

In Section IV, we described how the Generalized Pareto Distribution (GPD) should theoretically be the preferred model for peak-over-threshold (POT) data, and we referenced several studies where the GPD has been applied with such data. However, all the results that were presented for the POT analyses were based on fitting a Weibull 3-parameter (W3P) model to the data. We now revisit this choice of distribution to see how the results would change if the GPD had been used instead. Figure 13 shows such a comparison of fitted loads distributions for the 13-15 m/s bin. Parameter estimates for both two distribution model choices were obtained using the method of moments.

Table 9. Optimal Thresholds: Number ofStandard Deviations above the Mean

	5-7	7-9	9-11	11-13	13-15	15-17	17-19	19+
FBM	5.55	4.72	3.19	2.16	2.07	1.99	2.01	2.41
EBM	1.43	1.31	1.53	1.45	1.33	1.75	1.58	1.57
TBM	6.50	5.10	2.96	1.71	0.97	1.83	2.30	2.26

In each case, optimal threshold levels were found separately for the two distributions so that associated datasets of peaks over the selected threshold are not necessarily the same in the top and bottom plots (corresponding to the W3P and the GPD, respectively) for each load type. (The threshold level that is optimal

for a W3P fit is not always the same threshold level that is optimal for a GPD fit.) Also, recall that the GPD has an upper bound when the shape factor, k, is greater than zero. This upper bound is represented in Fig. 13 by the vertical dashed lines seen at the right end of the distributions for FBM and EBM. The GPD fit to the TBM data has a negative k value and, hence, has no upper bound.

Table 10. Loads associated with different return periods based on the POT method: optimal threshold versus μ + 1.4 σ threshold.

Load Type	Threshold	$L_{17 \text{ day}}$	$L_{1 \text{ yr}}$	$L_{10 \text{ yr}}$	$L_{50 \text{ yr}}$
EDM	Optimal	0.99	1.07	1.14	1.18
ГDIVI	μ + 1.4 σ	1.01	1.11	1.18	1.24
EDM	Optimal	1.01	1.09	1.15	1.19
EDIVI	μ + 1.4 σ	1.00	1.07	1.12	1.15
TDM	Optimal	1.01	1.10	1.18	1.24
IDN	μ + 1.4 σ	1.04	1.28	1.56	1.81

A visual comparison between the two distribution fits gives the impression that the Weibull 3-parameter is the more "neutral" choice. That is, when the data at this scale appear to be close to linear, the distribution follows this near-linear trend quite well. On the other hand, the Generalized Pareto Distribution fits exhibit larger curvatures and, thus, more dramatic behavior. In the case of flap bending moment loads, the GPD fit predicts an

upper bound for the load that is only slightly higher than the maximum observed load (during the field measurement campaign), which is a result we cannot have great confidence in. At the other end of the spectrum, if one studies the tower bending moment data, the GPD fit shows an extreme negative curvature and, as a result, predicts unrealistically large long-term loads. This striking difference in the behavior of the GPD fits for the FBM and TBM loads is especially disconcerting since the data for these two load types do not visually appear to be dramatically different.

Table 11 summarizes overall return loads (integrated over all wind speeds) calculated using the two distribution models. The predictions with the Weibull 3-parameter model are slightly more conservative for the FBM and EBM loads because W3P distribution fits are unbounded while the GPD fits are bounded from above. For the TBM loads,

on the other hand, the GPD results are overly conservative because the fitted distributions are all unbounded and, as was seen in the plots for the 13-15 m/s bin (Fig. 13), the negative curvature can sometimes be very severe and the fitted GPD distribution can be wildly unbounded then.



Figure 13. Weibull 3-parameter vs. Generalized Pareto distribution fits for loads in the 13-15 m/s bin.

In summary, the GPD model has the advantage of having a strong theoretical basis but, in practice, the fitted distributions are rather erratic yielding very low upper bounds in some cases, and explosive unboundedness in others. So while the model is the theoretically appropriate one to use with peak-over-threshold data, its use can occasionally lead to large errors. On the other hand, the Weibull 3-parameter model appears to be more robust in the face of sampling variability (as was seen with the field data analyzed here). Hence, despite the lack of mathematical justification for applying this model to peak-over-threshold data, doing so appears to lead only to small, conservative errors, and the use of this W3P model is recommended for wind turbine load extrapolation studies.

 Table 11. Loads associated with different return periods

 based on the POT method: Weibull 3-parameter

 versus Generalized Pareto distributions.

Load Type	Distribution	$L_{17 \text{ day}}$	$L_{1 \text{ yr}}$	$L_{10 \text{ yr}}$	$L_{50 \text{ yr}}$
EDM	Weibull 3-P	0.99	1.07	1.14	1.18
ΓΟΝΙ	GPD	0.98	1.04	1.08	1.11
EDM	Weibull 3-P	1.01	1.09	1.15	1.19
EDIVI	GPD	1.00	1.07	1.11	1.14
TBM	Weibull 3-P	1.01	1.10	1.18	1.24
	GPD	1.04	1.30	1.93	2.80

C. Statistical Independence between Peaks in POT Data

When peak-over-threshold data are extracted so that distribution fits to them can be sought, it is assumed that the retained peaks from the underlying load process are statistically independent. One of the advantages of choosing a threshold that is sufficiently high is that the peaks over that selected threshold indeed do tend to be less correlated with other peaks including ones nearby. However, the clustering of peaks may still be observed. If the

correlation between successive peaks is positive (and large), the indication is that the assumption of independence required by our statistical models may not be strictly valid.

It has been suggested that a minimum time separation between successive peaks should be enforced when carrying out POT data analysis (see, for example, Brabson and Palutikof¹¹). For wind turbine loads, this minimum time separation could be based, say, on an assumed wind gust duration or on any important/dominant natural periods of vibration associated with relevant turbine components and the load in question. Regardless of how such a time separation is selected, imposing such a requirement has the advantage of increasing the likelihood of having independence between peaks, at the expense of possibly losing some data.

To implement a time separation (of, say, *t* seconds) on an existing peak-over-threshold dataset, the following two steps can be applied: first, identify clusters of peaks—i.e., any two peaks that occur within *t* seconds of each other are considered to be part of the same cluster; then, discard all of the peaks except the largest in the cluster.

Figure 14 demonstrates the results of this procedure applied to the FBM time series data with the μ + 1.4 σ threshold that were presented in Fig. 5. For both cases (with and without any enforced time separation), the correlation coefficient between successive peaks was calculated. Before any time separation requirement between peaks was enforced, there were 63 extracted peaks (see middle plot of Fig. 5) for which the correlation coefficient between successive peaks was 0.212; with an enforced time separation of 10 seconds (see Fig. 14), there were just four peaks remaining, for which the correlation coefficient was -0.850. (Large negative correlation between peaks is

purely coincidental and has little physical significance; the main goal is to reduce large positive correlations, because these are associated with dependence between the peaks.) The implication of this example is that it is possible to reduce the dependence between peaks, but this comes at a large cost in terms of the reduction of the size of the dataset.



Figure 14. Peak-over-threshold data for FBM in the 9-11 m/s bir (circled peaks satisfy a minimum time separation of 10 seconds—compare with Fig. 5).

The relationship between the correlation of successive peaks and different choices for enforced time separations between peaks as well as different threshold levels is shown in Fig. 15 for flap bending moment loads in the 17-19 m/s bin. With low threshold levels, correlation between peaks is generally quite high and positive, but it reduces noticeably by imposing a minimum time separation. At high threshold levels, correlations are quite low and insensitive to time

separation. Similar plots for lower wind speed bins (not shown) show correlations between successive peaks that are higher across the board, and do not approach zero even for time separations as large as 20 seconds. For FBM and EBM in both low and high wind speed bins, the largest drop in correlation occurs between the 2- and 4-second separation levels indicating that, if a time separation is to be enforced, 4 seconds may be an appropriate choice for such a separation.



Figure 15. Correlation between successive peaks over a threshold versus time separation and threshold level for FBM in the 17-19 m/s bin.

An important final question regarding the imposition of a minimum time separation with POT data is obviously related to whether or not final results on longterm loads are influenced or not by Fifty-year return loads have it. been computed for ten different imposed time separations varying between 2 and 20 seconds (in increments of 2 seconds) and the results are summarized in Table 12. In each case, the calculations were performed in the same manner as in Section IV-namely, by using an optimal threshold level together with the Weibull 3-parameter distribution-so that the results may be fairly compared. (Note that the results for a zero time

separation are the same as those presented in Table 6 earlier.) Overall, the variation in predicted 50-year return loads with choice of time separation is seen to be quite small. In fact, comparing the results based on a 20-second time separation with those based on no enforced separation at all reveals adjustments to the 50-year loads of only - 4.2%, -1.5%, and +2.4% for the FBM, EBM, and TBM, respectively.

In summary, although the imposition of a minimum time separation does reduce dependence between successive peaks, this comes at some cost in terms of the associated reduction in the size of the dataset which, for the purposes of extrapolation, would require that more data be collected. Importantly, though, neglecting to force a minimum time separation or, equivalently, the use of ordinary peak-over-threshold data, does not appear to seriously bias predictions of long-term loads. This suggests that any requirement of a minimum time separation between peaks in POT data on loads may very well be an unnecessary complication to the procedure.

Table 12. Fifty-year return loads, using peak-over-threshold data with various imposed time separations between peaks.

Time separation (s)	FBM	EBM	TBM
0	1.18	1.19	1.24
2	1.17	1.19	1.25
4	1.15	1.17	1.35
6	1.15	1.18	1.33
8	1.14	1.18	1.32
10	1.14	1.19	1.31
12	1.14	1.19	1.30
14	1.13	1.18	1.29
16	1.13	1.18	1.28
18	1.14	1.18	1.28
20	1.14	1.17	1.27

VII. Extrapolations based on Limited Data

Finally, we address how load predictions based on statistical extrapolation are influenced by the size of the dataset. Obviously, a larger dataset can be expected to yield more reliable results than a smaller dataset. Here, we seek to quantify the degree to which an increase in the dataset size dataset is accompanied by a corresponding decrease in the statistical uncertainty of the resulting extrapolation. Interestingly, the 3rd edition of the IEC 61400-1 guidelines¹ only requires that the equivalent of six ten-minute simulations be run per wind speed bin in extrapolation for load extremes. It is not surprising that some have found this number to be insufficient to ensure convergence to a stable long-term distribution for some wind turbine loads, but how much data is really necessary to provide an acceptable amount of stability in such load predictions?

In this study, we use varying sized subsets of the overall Lamar dataset to extrapolate long-term loads (using the peak-over-threshold method with optimal thresholds and the Weibull 3-parameter distributions), and long-term load predictions are then compared to those based on use of the entire dataset (these were summarized in Table 6). Specifically, subsets of 50, 100, 500, and 1,000 files are selected randomly, without replacement, from the overall dataset of 2,485 files. This is repeated 100 times for each of the four subset sizes and, in each case, the final result— i.e., the 50-year return load—is recorded. Results presented in Fig. 16 show, for each data subset size, the range of predictions that result from the 100 repetitions, as well as standard error bars on both sides of the mean value. The horizontal dashed line is the 50-year return load predicted with the overall dataset. Table 13 shows the standard errors on predicted loads arising from the 100 repetitions for each data subset size. As expected, the standard errors trend downwards with increasing sample size, but this trend is fairly gradual.



Figure 16. Fifty-year return loads: range of predictions, mean and standard error bars resulting from subsets of various sizes, based on 100 repetitions. (The horizontal dashed line is the prediction based on the entire dataset.)

Although all of the results presented in this paper are based on field data, in practice the vast majority of statistical extrapolations are performed using simulated loads data. When simulated loads data are used, one can choose how the data are distributed among the various wind speed bins, and a logical choice would be to simulate an equal number of datasets in each wind speed bin. To be consistent with such choices and to address the question of what might constitute an appropriate number of simulations to perform, we repeat the calculations described above varying the number of data used in long-term load prediction but, this time, we randomly draw an equal number of files from each wind speed bin, again repeating our draws 100 times.

Standard errors on 50-year load predictions based on various subsets of data with equal distributions over all bins are presented in Table 14. Again, convergence is not extremely fast and, in general, the standard errors are slightly smaller but on the same order as those in Table 13. In both cases, there is no simple answer to the question of how much data is necessary to ensure that the extrapolated load predictions will have indeed converged; standard errors in Tables 13 and 14 are intended merely to quantify the manner in which uncertainty in extrapolated load predictions are affected by the size of the dataset. When simulation studies are undertaken, it may be more meaningful to assess convergence of extrapolated load distributions for one wind speed bin at a time. Thus, one might consider varying the number of simulated data sets to assure convergence of conditional distributions (such as in Fig. 8) first rather than examine overall (integrated) long-term loads as was done here.

Table 13. Standard errors on 50-year return load predictions based on varying total number of data subsets drawn from all bins (100 repetitions).

Load Type	Subset Size (# 10 minute files)			
Loau Type	50	100	500	1000
FBM	0.103	0.103	0.055	0.028
EBM	0.116	0.098	0.061	0.041
TBM	0.149	0.196	0.167	0.074

Table 14. Standard errors on 50-year return load predictions based on varying number of data subsets drawn from each wind speed bin (100 repetitions).

Load Type	Subset Size (# 10 minute files per bin)			
	6	25	50	100
FBM	0.085	0.050	0.041	0.029
EBM	0.100	0.080	0.057	0.020
TBM	0.150	0.140	0.135	0.100

VIII. Conclusions

Field data on flap, edge, and tower bending loads from a utility-scale wind turbine were used to perform statistical extrapolation of long-term wind turbine loads using three alternative procedures: the method of global maxima, the peak-over-threshold method, and a process model approach. For the peak-over threshold method, issues of threshold choice, distribution model, and independence of peaks were explored in detail. Finally, the uncertainty in long-term loads was studied as a function of the quantity of data used in the extrapolation. General conclusions and recommendations are as follows:

- The peak-over-threshold method yields far superior results in comparison to the other methods.
 - The use of an "optimal" threshold leads to oosing a threshold such as $\mu + 1.4\sigma$ beforehand

better fits of the distributions to data in comparison to choosing a threshold such as $\mu + 1.4\sigma$ beforehand.

•

- The Weibull 3-parameter distribution performs consistently well for peak-over-threshold (POT) data, although it is unsupported by theory. The Generalized Pareto distribution has a stronger theoretical basis for use with POT data, but performs erratically in some cases, particularly for tower bending moment.
- The requirement of a minimum time separation between peaks in the peak-over threshold method has a very slight impact on extrapolated long-term load predictions, and has the disadvantage of significantly reducing the available amount of data. Such a requirement may thus be an unnecessary complication.
- Uncertainties on extrapolated results decrease gradually with increasing size of dataset.

Additional studies with different field datasets and/or using simulated loads data are necessary to corroborate the conclusions reached here which were all based on limited field data. Especially if a similar study is conducted with simulated loads data, useful insights can be gained that can help in evaluating design load cases that deal with statistical loads extrapolation.

Acknowledgments

The authors gratefully acknowledge the financial support provided by Contract No. 30914 from Sandia National Laboratories and by a CAREER Award (No. CMS-0449128) from the National Science Foundation. They also are pleased to acknowledge useful discussions with Herb Sutherland, Paul Veers, and Jose Zayas at Sandia during the course of this work.

References

¹IEC 61400-1 Ed. 3, *Wind Turbines - Part 1: Design Requirements*, International Electrotechnical Commission, IEC, 2005. ²Moriarty, P.J., Holley, W.E., and Butterfield, S.P., "Extrapolation of Extreme and Fatigue Loads using Probabilistic Methods," NREL/TP-500-34421, 2004.

³Fitzwater, L. M. and Winterstein, S. R., "Predicting Design Wind Turbine Loads from Limited Data: Comparing Random Process and Random Peak Models," *Proceedings of the ASME Wind Energy Symposium*, AIAA, Reno, NV, 2001, pp. 355-364.

⁴Zayas, J. R., Sutherland, H. J., and Patton, T. C., "Continuous Long-Term Test Program on a Utility-Scale Wind Turbine," *Proceedings of the ASME Wind Energy Symposium*, AIAA, Reno, NV, 2006.

⁵Gumbel, E.J., *Statistics of Extremes*, Columbia University Press, New York, NY, 1958.

⁶Holmes, J.D., and Moriarty, W.W., "Application of the Generalized Pareto Distribution to Extreme Value Analysis in Wind Engineering," *Journal of Wind Engineering and Industrial Aerodynamics*, 83, pp. 1-10, 1999.

⁷Pandey, M.D. and Sutherland H.J., "Probabilistic Analysis of LIST Data for the Estimation of Extreme Design Loads for Wind Turbine Components," Proceedings of the ASME Wind Energy Symposium, AIAA, Reno, NV, pp. 259-269, 2003.

⁷Hosking, J.R.M., Wallis, J.R., and Wood, E.F., "Estimation of the Generalized Extreme Value Distribution by the Method of Probability Weighted Moments," *Technometrics*, 27, pp. 251-261, 1985.

⁸Pickands, J., "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics*, 3(1), pp. 119-131, 1975.

⁹Hosking, J.R.M., and Wallis, J.R., "Parameter and Quantile Estimation for the Generalized Pareto Distribution," *Technometrics*, 29, pp. 339-349, 1987.

¹⁰Davison, A.C., and Smith, R.L., "Models for Exceedances over High Thresholds," J.R. Statist. Soc. B, 52(3), pp. 393-442, 1990.

¹¹Brabson, B.B., and Palutikof, J.P., "Tests of the Generalized Pareto Distribution for Predicting Extreme Wind Speeds," *Journal of Applied Meteorology*, 39, pp. 1627-1640.

¹²Ryden, J., "Aspects on Estimation of Return Values for Extreme Wind Speeds," *Proceedings of the Sixteenth International Offshore and Polar Engineering Conference*, San Francisco, CA, 2006, pp. 81-85.

¹³Winterstein, S.R., "Moment-Based Hermite Models of Random Vibration," *Report No. 219*, Dept. of Structural Engineering, Technical University of Denmark, Lyngby, Denmark, 1987.

¹⁴Winterstein, S.R., "Nonlinear Vibration Models for Extremes and Fatigue," *Journal of Engineering Mechanics*, 114(10), pp. 1772-1790, 1988.