46th AIAA Aerospace Sciences Meeting and Exhibit
7 - 10 January 2008, Reno, Nevada

AIAA 2008-1339

Towards an Improved Understanding of Statistical Extrapolation for Wind Turbine Extreme Loads

Jeffrey Fogle*, Puneet Agarwal*, and Lance Manuel†
Dept. of Civil, Architectural, and Environmental Engineering, University of Texas, Austin, TX 78712, USA

One of the design load cases to be evaluated in the current edition of the design standard for wind turbines as prescribed by the International Electrotechnical Commission (IEC) requires that nominal loads associated with a return period of 50 years be established. This must usually be done by carrying out simulations of the aeroelastic response of the turbine. This design load case is for an ultimate limit state and in order to be able to estimate this rare nominal load, it is necessary to have, as starting point, extreme loads data that are of an adequate quantity and quality to facilitate robust long-term predictions of the load. Practitioners attempting to follow the guidelines set forth in the IEC standard have voiced concerns about aspects of the load extrapolation as expressed therein—for instance, questions have arisen related to the minimum number of ten-minute turbine response simulations that should be performed, about whether only a single (global) maximum load from each simulation should be saved or whether, alternatively, several time-separated (block) maxima may be preferred. Also, while it is clear that not all turbine load types are influenced by each wind speed between cut-in and cut-out to the same degree, no discussion about the desired effort required for each wind speed bin is presented; it is important that simulation effort be focused mainly on wind speed ranges that control the highest load for each load type. The present study attempts to answer all these questions. Using extreme load statistics (global and block maxima) for four different load measures derived from aeroelastic simulations on a 5MW turbine model, we study extreme loads as a function of wind speed (we call these short-term load distributions). For different block sizes (time separations), block maxima are tested for independence and empirical load distributions are compared when global and block maxima are used. We aggregate the short-term load distributions (conditional on wind speed) to the long-term level by integration over all wind speeds. We present convergence criteria which serve to assess whether or not an adequate number of simulations has been performed. Throughout, we highlight the importance of striving for efficiency with regard to the effort expended in running simulations by staying cognizant of important wind speed ranges that are design drivers for each load type. Together, all of the above lead to a proposal that we make for addressing load extrapolation that focuses on efficiency, that spells out how to employ either global or block load maxima, and that provides easy-to-use convergence criteria for deciding on an adequate number of simulations that must be performed before attempting long-term load prediction using extrapolation.

I. Introduction

Statistical extrapolation of wind turbine loads from limited simulations is required in order to predict rare long-term loads associated with an important design load case specified in the International Electrotechnical Commission standard for the design of wind turbines (IEC 61400-1: Ed. 3, 2005)[1]. A response simulation represents the stochastic response of a wind turbine to specified random environmental conditions. Each design load case (DLC) specifies the environmental conditions to be used for the aeroelastic simulations. One particular load case, DLC 1.1, will be the focus of our discussions; therefore, a brief background about it is appropriate here. DLC

*Graduate Research Assistant
†Associate Professor

1.1 deals with extreme loads that a wind turbine might experience during normal operation—when wind speeds range between cut-in and cut-out. DLC 1.1 also specifies that inflow conditions used in the simulations should represent those associated with near-neutral atmospheric conditions. A normal turbulence model (NTM) is prescribed in the IEC standard that must be used to represent the inflow turbulence fields pertinent to DLC 1.1. In this load case, the hub-height wind speed, V, averaged over ten minutes may be treated as a single random variable representing the environment; in the IEC standard, turbulence intensity needed for the NTM is specified in terms of this wind speed, depending on the class for which the turbine design is considered. DLC 1.1 requires that aeroelastic simulations be conducted over the entire power-producing wind speed range from cut-in to cut-out. It is convenient, as the IEC standard permits, to carry out simulations over discrete wind speed intervals or bins; typically, bins of 2 m/s size for V are employed.

Since DLC 1.1 relates to an ultimate limit state for design, it requires that a "nominal" load be established that has a low probability of occurrence. The standard states that this nominal load must have a return period of 50 years or, equivalently, that this load may be exceeded on average only once every 50 years[1]. We will refer to this nominal load as the 50-year load in the following. It is clear that prediction of the 50-year load needs to recognize the various wind speeds that will be encountered and their relative likelihoods. If load statistics or distributions are established separately for each wind speed bin, it is important that a sufficient number of simulations are carried out for each bin and that aggregation and proper weighting of loads from each bin are also done correctly. Clearly, it is computationally infeasible to carry out the large number of ten-minute simulations that would be needed to accumulate loads data that account for the actual duration that would match the target return period. Instead, a limited number of simulations are generally carried out; effort, though, must be judiciously expended in running simulations most carefully for wind speed ranges that bring about the rarest, largest loads as well as the most variable ones. Careful statistical extrapolation from such limited simulations can then make it possible to derive the required 50-year loads.

In this study, we address some concerns that have been raised with regard to experiences practitioners have had with attempts to address DLC 1.1 in the IEC standard[1]. The standard requires that the 50-year load be established but it does not unambiguously provide a procedure that will lead to this load from simulations. The guidelines that are provided are vague at best, for example, when addressing the issue of what represents a sufficient number of simulations to run. There are also no clear indications of what constitutes a check that the 50-year load when derived is a robust or stable estimate. The standard does not explicitly suggest that effort might best be focused on the most important bins (usually at or around rated wind speeds for some load types and at our around cut-out wind speeds for others); although this would be prudent. Finally, the standard does not clearly describe what extreme load statistics may be saved from each ten-minute simulation—the use of a single (global) maximum from each simulation needs to be considered against alternatives that employ several time-separated (block) maxima from each simulation; in the latter, the question of what constitutes a set of independent block maxima is important as it fundamentally affects the derivation of the 50-year load. The standard allows use of multiple maxima by methods such as the peak-over-threshold procedure but details are missing with regard to robust tests for independence. In this study, we address each of these issues. In brief, we address efficiency when we discuss which wind speed bins are design drivers for each load; we address convergence criteria that lead to approaches to quantify when an adequate number of simulations have been run that yield stable short-term empirical load distributions; and we address the issue of independence in block sizes and statistical tests for independence and also discuss the difference in load predictions based on the use of global and block maxima. Throughout, insights are provided to guide the effort involved in carrying out statistical loads extrapolation as required for DLC 1.1.

This work represents our continuing effort to improve understanding of statistical loads extrapolation as it applies to wind turbine design. The load simulation data sets were provided to the authors as well as to other members of a Loads Extrapolation Exercise (LE^3) working group that was formed at the request of the maintenance committee of the IEC 61400-1 turbine design standard. The simulated loads data sets were generated for a baseline 5MW wind turbine model developed at the National Renewable Energy Laboratory. All of the findings that are reported here are based exclusively on statistical studies on data from simulations with this turbine model alone.

II. The LE^3 Dataset

The Loads Extrapolation Evaluation Exercise (LE^3) working group was constituted in order to address ongoing issues related to the IEC standard and particularly DLC 1.1 that deals with statistical loads extrapolation from limited simulation. The LE^3 working group approved a proposal to develop a database of simulated load time histories and summary statistics from a 5MW turbine model developed by the National Renewable Energy Laboratory (NREL). This model is based on an onshore version of NREL's baseline turbine model developed to

represent a utility-scale 5MW offshore wind turbine[2]; this onshore model has identical properties to the offshore turbine above the mudline. The turbine has a hub height of 90 meters and a rotor diameter of 126 meters. The machine is a variable-speed, collective pitch-controlled turbine with a rated wind speed of 11.5 m/s. The maximum rotor speed is 12.1 rpm. Moriarty[3] provides a detailed account of the turbine model and the various inflow conditions covered by the simulations. Inflow turbulence was simulated using TurbSim v12.0[4]; a Kaimal power spectrum, a shear exponent of 0.20, and a deterministic turbulence standard deviation (given V) was employed based on the NTM. The program, FAST v6.02b[5], was used to carry out aeroelastic simulations for hub-height wind speeds, V, varying between cut-in and cut-out wind speeds. For the IEC Class I-B site assumed, the ten-minute hub-height wind speed follows a standard Rayleigh distribution with mean equal to 10 m/s.

Two data sets were generated for the use by the LE[3] working group. The first data set consists of 14,400 ten-minute simulations or 1,200 simulations for each of twelve wind speeds ranging between 3 and 27 m/s. (When running TurbSim to generate inflow turbulence time histories, the target wind speeds were set at discrete values of 3 m/s, 5 m/s, etc. up to 25 m/s and were assumed to represent 2 m/s bins centered at the target values. Realized ten-minute average wind speeds varied slightly from the target values.) A second data set was generated by representing wind speeds according to a Rayleigh distribution for five full years and then carrying out aeroelastic response simulations for those inflow conditions.

In the present study, four representative and contrasting loads were analyzed from the first LE[3] data set. These loads include the out-of-plane bending moment (OOPBM) at a blade root, the out-of-plane blade tip deflection (OOPTD), the fore-aft tower bending moment (FATBM) at the base, and the in-plane blade bending moment (IPBM) at a blade root.

III. Statistical Load Extrapolation

The first step in statistical load extrapolation involves the identification of load extremes from the turbine simulations. Consider the case where the single largest (global) maximum is extracted from each ten-minute time series for a wind speed bin, V_k. The probability, $P(L > l|V_k)$, that a given load of interest, L, will exceed any specified load level, l, in ten minutes may be obtained by rank ordering the N_k real-valued global maxima (X_i; $i = 1$ to N_k) that are obtained by running N_k simulations for wind speed bin, V_k. In practice, once the load level, l, is specified, one can obtain the empirical short-term conditional distribution by finding two integers j and j+1 that are such that where $X_j \le l \le X_{j+1}$ where $1 \le j \le (N_k - 1)$. Since $P(L \le X_j|V_k) = j/(N_k+1)$ and $P(L \le X_{j+1}|V_k) = (j+1)/(N_k+1)$, one can obtain $P(L > l|V_k) = 1 - n_k(l)/(N_k+1)$ where $n_k(l)$ is obtained by interpolation. In summary, the empirical short-term load distribution for any bin, V_k, may be obtained as follows:

$$\begin{aligned} P(L > l \mid V_k) &= 1 - \frac{n_k(l)}{N_k+1}; \quad \text{where } n_k(l) = j + \frac{l - X_j}{X_{j+1} - X_j}, \text{if } X_j \le l \le X_{j+1} \text{ and } 1 \le j \le N_k - 1 \\ &= \frac{N_k}{N_k+1}; \text{ if } l < X_1 \\ &= \frac{1}{N_k+1}; \text{ if } l > X_{N_k} \end{aligned} \tag{1}$$

Alternately, one can set $n_k(l)$ equal to j in Eq. (1) above if linear interpolation is not desired; this will lead to a staircase-like empirical distribution function that is sometimes preferred in statistical analysis since it is more consistent with what was actually observed in the data. Figure 1 shows example empirical short-term distributions for two loads, OOPBM and OOPTD estimated using global maxima from 200 simulations in each wind speed bin.

The distribution given by Eq. (1) is termed a short-term distribution on the global maximum load, L, since it is a distribution conditional on wind speed. All the various wind speeds likely to be encountered need to be considered in order to yield the long-term distribution on L. In terms of a continuous random variable, V, the long-term distribution can be obtained as follows:

$$P(L > l) = \int_{V_{in}}^{V_{out}} P(L > l \mid V = v) f_V(v) dv \tag{2}$$

To evaluate Eq. (2) in order to obtain long-term distributions for turbine loads, one needs short-term distributions as well as the wind speed probability density function, $f_V(v)$; the latter is taken to be the Rayleigh density function for IEC Class I-B conditions in this study and only the mean value of V of 10 m/s is needed. One can use Eq. (2) to obtain the long-term load distribution for loads if parametric distribution fits are attempted to each empirical short-term distribution given by Eq. (1). This might be termed the "fitting-before-aggregation" approach.

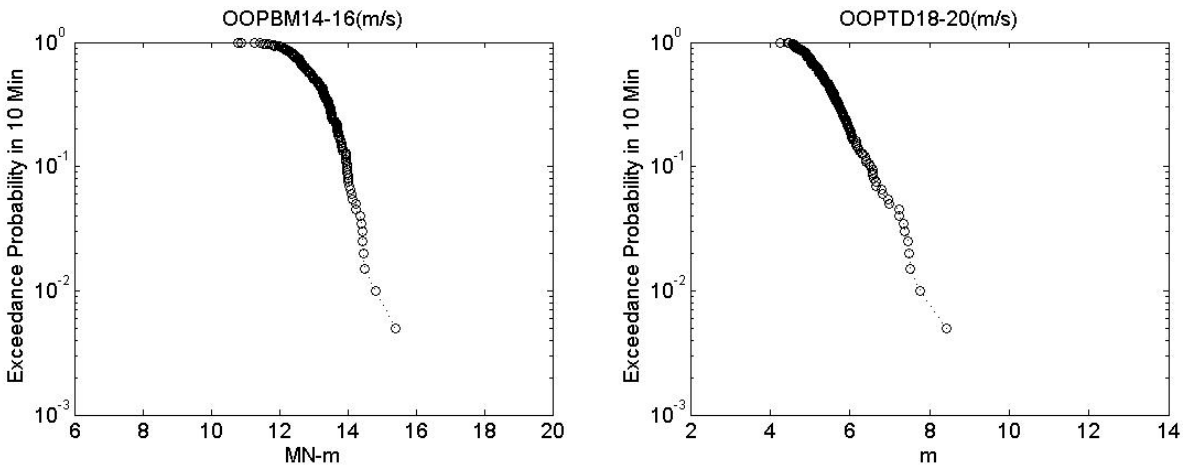


Figure 1: Example empirical short-term empirical distributions for OOPBM and OOPTD estimated using 200 simulations for each wind speed bin.

Alternatively, one could obtain the long-term distribution by collecting data from all the wind speed bins together (this can be conceived of as putting all the data into one box). Assume that the maxima in bin i, are rank-ordered such that if there are total of N_i extremes, then $l_{1,1} \le l_{1,2} \le l_{1,3} \ldots \le l_{1,N}$, the notation used implies that $l_{i,k}$ is the kth rank-order maximum from bin i Note that the total number of load maxima from all bins, N, is equal to $\sum_{i=1}^{N_B} N_i$. If the number of simulations run in each bin is proportional to the actual likelihood of that bin, then the empirical long-term distribution is given simply as follows:

$$P(L \le l) = \sum_{i=1}^{N_B}\left(\sum_{k=1}^{N_i} \frac{I[l_{i,k} \le l]}{N+1}\right)$$
$$\text{where } I[l_{i,k} \le l] = 1, \quad \text{if} \quad l_{i,k} \le l$$
$$= 0, \quad \text{otherwise} \tag{3a}$$

If the number of simulations run in each bin is not in proportion to the actual likelihood of that bin, an appropriate weighting needs to be included to yield the following expression for the empirical long-term distribution:

$$P(L \le l) = \sum_{i=1}^{N_B}\left(\sum_{k=1}^{N_i} \frac{I[l_{i,k} \le l]}{N+1}\right) \cdot w_i = \sum_{i=1}^{N_B}\left(\sum_{k=1}^{N_i} \frac{I[l_{i,k} \le l]}{N+1}\right) \cdot \frac{N \cdot P(V_i)}{N_i} \tag{3b}$$

where w_i refers to the weight assigned to the ith wind speed bin in the summation. This weight is determined as the ratio of the required number of samples, $N{\cdot}P(V_i)$, that should have been obtained from bin, i, to the actual number N_i that were collected for that bin (where $P(V_i)$ represents the Rayleigh distribution-based probability of occurrence of wind speeds in the ith wind speed bin and, again, N_B represents the total number of bins that need to be considered between cut-in and cut-out). Equations (3a) and (3b) can serve as an alternative to the parametric "fitting-before-aggregation" approach followed with Eq. (2). In fact, using Eq. (3a) or (3b) as appropriate, can be considered the "aggregation-before-fitting" approach since no short-term distribution fits are necessary; rather, the data are empirical aggregated according to Eqs. (3a) or (3b), and then only for the sake of extrapolation (discussed next) can parametric fits be applied to the empirical long-term distribution.

Note that Eqs. (3a) and (3b) provide empirical expressions for the long-term distribution of the global maximum of ten-minute maximum load, L. This distribution can be used to derive the 50-year load by noting that the 50-year load has a return period of 50 years. If one assumes independence between global ten-minute maxima, the desired 50-year load, l_{50}, must be such that the probability of its exceedance in ten minutes is 10 min/(50 yr × 365.25 days/yr × 24 hr/day × 60 min/hr) = 1/2,629,800 = 3.8×10^{-7}. Clearly, in order to predict loads with such low probabilities of exceedance, statistical extrapolation will be necessary from the limited simulations that will be carried out. As an example, empirical long-term aggregated distributions computed on the basis of Eq. (3b) using 200 simulations for each wind speed bin are presented in Fig. 2 for two loads, OOPBM and OOPTD. To emphasize the need for statistical extrapolation, the desired 50-year probability level is also shown in Fig. 3. It is clear that extrapolation

over several orders of magnitudes is necessary between the lowest empirical load exceedance probability and the desired 50-year level.

It is worthwhile to note that in the "aggregation-before-fitting" approach suggested by Eqs. (3a) and (3b), since the data are more heterogeneous as they represent different wind speed bins, parametric fits can focus on the tails of the empirical data. Extrapolation to the desired 50-year return period level can follow directly with either of the two approaches. Ragan and Manuel[6] provide examples of the use of Generalized Extreme Value distribution fits to field data on loads (global maxima) from a utility-scale wind turbine based on the "fitting-before-aggregation" approach. The small amounts of data usually available in empirical short-term data often lead to fits of poor quality[6]; moreover, such fits are needed for all wind speed bins even where loads are not large. The "aggregation-before-fitting" approach involves fitting to distributions on long-term loads; aggregated data are generally larger in number and given that large rare loads are of interest, fitting can be concentrated in the tail which is most useful for extrapolation. In this study, we focus on the approach of aggregating data first to arrive at empirical long-term distributions that can then be used in fitting and extrapolation.

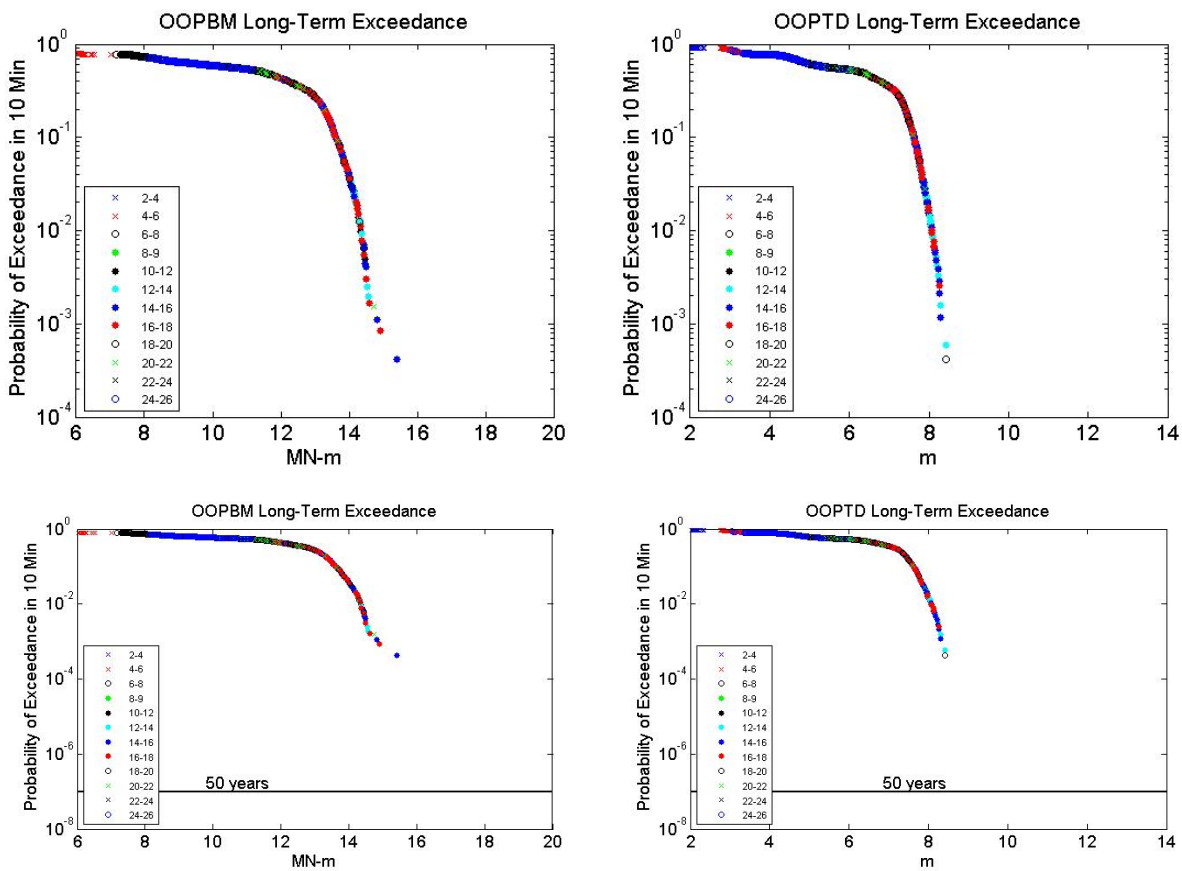


Figure 2: Example empirical long-term empirical distributions for OOPBM and OOPTD estimated using 200 simulations for all the wind speed bins considered in aggregation. Probability levels associated with a 50-year return period for loads are shown for comparison.

IV. The Relative Importance of Different Wind Speeds to Turbine Load Extremes

To compare the relative importance of different wind speed bins on load extremes, it is of interest to study load extreme statistics as a function of wind speed. With a little effort (i.e., limited simulations), it is often possible to identify which wind speeds can cause the largest turbine loads on average and, equally important, which ones show the greatest load variability. The largest loads are associated with the lowest probability of exceedance and, as such, are closest to the rare probability levels to which extrapolation is needed; the wind speeds where these largest loads occur are therefore of obvious interest. Empirical short-term distributions need to be well estimated in these bins in particular. Even if loads realized in some bins are not among the largest, if variability in extremes from simulations in those bins is large, they can have a significant influence on distribution tails and, hence, on extrapolation.

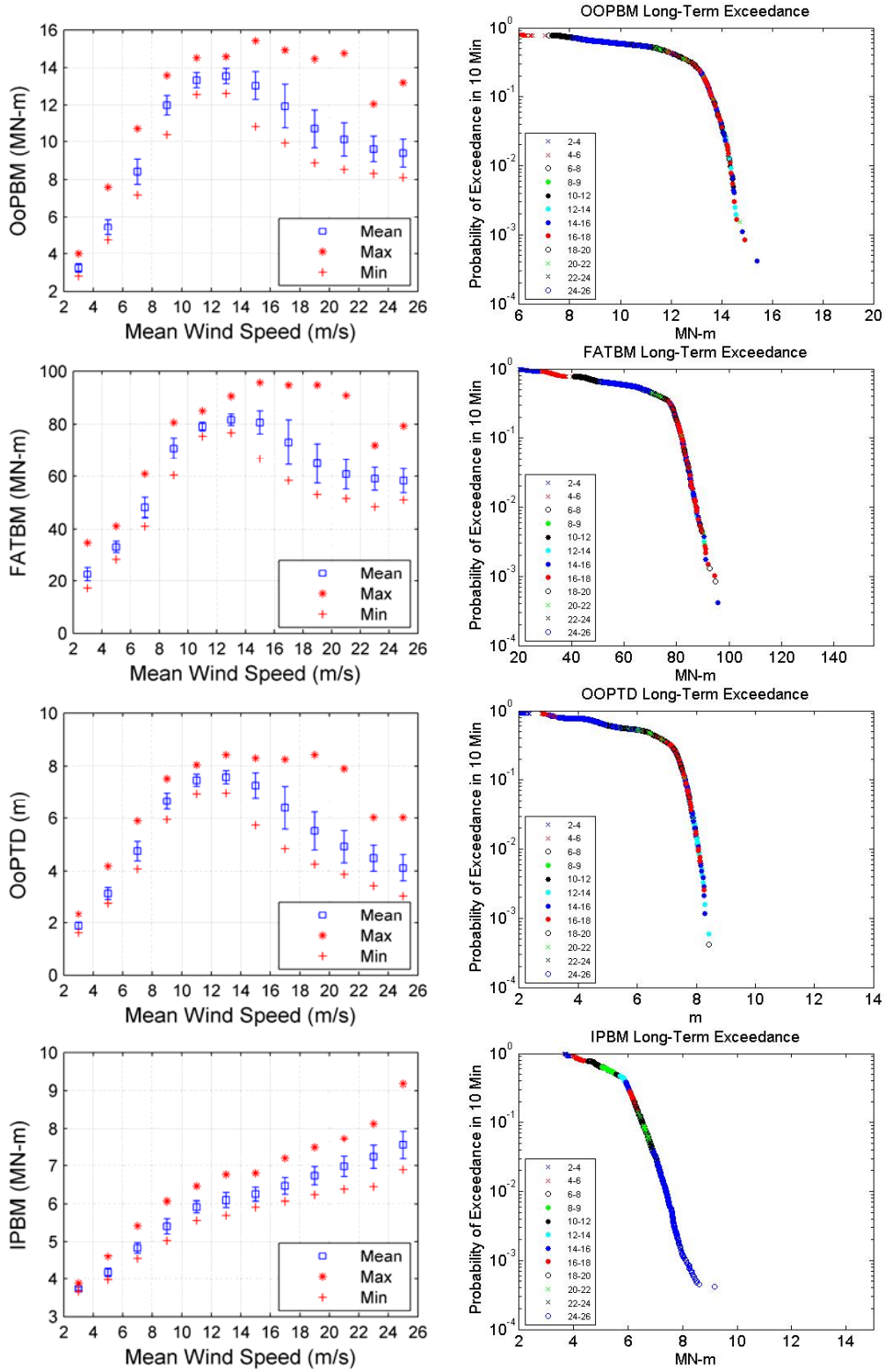


Figure 3: Distribution of short-term load maxima (left figures) as a function of wind speed for four loads, OOPTD, FATBM, OOPBM, and IPBM based on 200 simulations per wind speed bin. Long-term aggregated distributions (right figures) for the loads showing specific load values contributed to by each wind speed bin.

From Fig. 3, it is possible to identify those wind speed bins associated with the largest loads and greatest short-term extreme variability and also to identify those bins that most influence the tails of the long-term load distributions. This is useful since summary plots such as these make it possible to focus efforts in the most important bins. For OOPBM and FATBM, important controlling wind speed bins are in the 14-22 m/s range with perhaps the dominant winds being closer to the 14-16 m/s bin. The OOPTD is controlled by somewhat lower wind speeds; the largest loads occur between 10 and16 m/s. The IPBM is clearly dominated by wind speeds close to the cut-out wind speed of 25 m/s. While these controlling wind speed bins were identified by carrying out 200 simulations pre bin, it is possible to identify important bins with considerably less simulation effort. From Fig. 3, it is clear that wind speeds below 10 m/s do not contribute large loads to any of the loads discussed; moreover, they also do not exhibit large variability in load extremes. In the long-term distributions shown in Fig. 3, these bins contribute load values to the left of the "knee" of these empirical distributions; as such, they are not important in extrapolation and a minimal effort in terms of simulation may be justified for these bins when deriving short-term distributions based on Eq. (1). In carrying out simulations with a view towards extrapolation, it is worthwhile to understand turbine load extremes as a function of wind speed in this manner to avoid excessive computational effort.

V. Block Maxima and the Issue of Independence

An alternative to the use of only a single maximum (i.e., the global maximum) from each simulated ten-minute time series is to extract several extremes from each time series in a systematic manner. While this can be done by methods such as the peak-over-threshold procedure[6], there are simpler methods as well. For instance, one could split or partition the time series into individual non-overlapping blocks of constant duration. From each of these blocks, then, a single largest value is extracted; together these extracted extremes constitute a set of block maxima. Figure 4 shows an example OOPBM load time series with one-minute block maxima indicated by circles and the single ten-minute global maximum indicted by an asterisk.

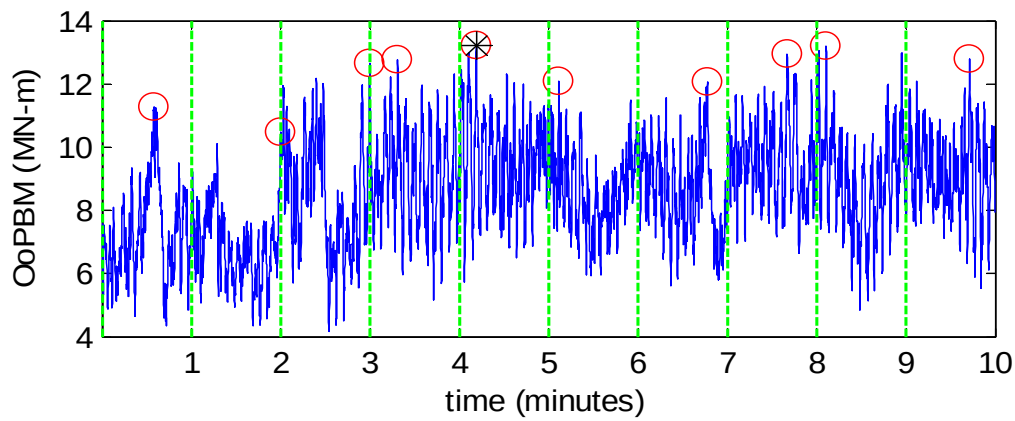


Figure 4: Example OOPBM load time series showing global and block maxima.

Figure 4 illustrates that from a single load time series of 10-minute duration, more extremes data may be extracted if block maxima are employed as opposed to global maxima in extrapolation. As long as the block maxima can be shown to be mutually independent, short-term global maxima (L) distributions as in Eq. (1) can be related to short-term block maxima (L_{block}) distributions as follows:

$$P(L < l \mid V_k) = \left[P(L_{block} < l \mid V_k)\right]^n \text{ or } F_L(l \mid V_k) = \left[F_{L_{block}}(l \mid V_k)\right]^n \tag{4}$$

where $F_L(\,)$ and $F_{L_{block}}(\,)$refer to the cumulative distribution functions for L and L_{block}, respectively.

In terms of probabilities of exceedance of any load level, l, one can also write:

$$P(L > l \mid V_k) = 1 - \left[1 - P(L_{block} > l \mid V_k)\right]^n \tag{5}$$

If one is interested in the p-quantile ten-minute maximum load, l_p, defined such that $F_L(l|V_k) = p$, the adjusted load quantile in terms of the block maximum distribution must be adjusted as follows:

$$l_p = F_L^{-1}(p) = F_{L_{block}}^{-1}(p^{1/n}) \tag{6}$$

where n represents the number of blocks contained in 10 minutes. As might be expected, the non-exceedance probability, p, for global maximum needs to be adjusted to a rarer non-exceedance probability level, $p^{1/n}$, for block maximum if it is to correspond to the same load. So, while a greater amount of extremes data are extracted when block maxima are employed and hence lower probability levels can be reached, the same p-quantile needs to be sought farther in the tail of the block maxima distribution. For instance, if the 80th percentile ten-minute maximum load is required (which corresponds to a non-exceedance probability of 0.80 for global maxima), when one-minute block maxima are used, the corresponding non-exceedance quantile for block maxima is $0.80^{1/10}$ or 0.978 which is considerably farther in the tail of the distribution of the block maxima. Figure 5 illustrates this effect for OOPBM load maxima extracted from six simulations of a single wind speed bin. The asterisks represent global maxima and the circles, block maxima. To highlight the point that the global maxima are also block maxima, the specific global maxima are shown twice to indicate where they appear in block maxima distribution. Some extracted block maxima are higher than a few global maxima and arguably better defined tail trends are seen in the block maxima distribution. However, as can be seen, the 80th percentile ten-minute maxima load is at an exceedance probability level of 1 - 0.978 or 0.022 if the block maxima distribution is used. The 80th percentile load itself is read off at roughly the same level with either distribution.

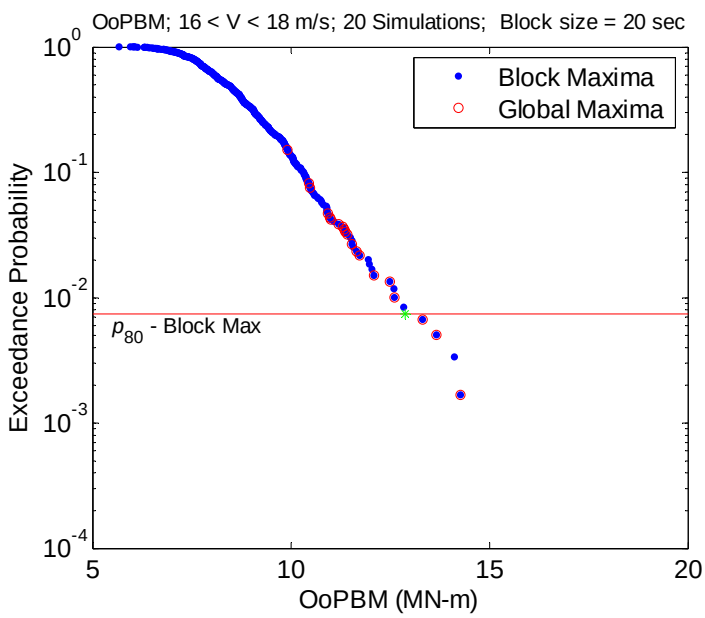


Figure 5: Comparison of block and global maximum probability levels associated with a given load quantile for OOPBM loads (in MN-m)..

As was stated before, Eqs. (4) to (6) are valid as long as block maxima selected from each time series are independent of each other. Intuitively, it may be expected that smaller block sizes will lead to greater dependence among the block maxima. Statistical tests for independence represent the only objective means of assessing the extent of independence or lack thereof in a sample of block maxima from load simulations.

A. Test for Independence

Several tests to evaluate independence between two random variables are available in the literature. We focus here on a test proposed by Blum et al[7]. Details related to this test along with examples may be found in Hollander and Wolfe[8]; that reference also provides a correction for a typographical error in an equation in Blum et al[7]. Blum's test has been demonstrated by Skaug and Tjøstheim[9] to work to test for independence for time series data, for which it was not originally developed.

Two random variables, X and Y, may be stated to be independent of one another if the product of their marginal probability distribution functions is equal to their joint distribution. According to Blum's test for independence, the null hypothesis, H_0, is that the two variables X and Y are independent. Thus, we have in terms of cumulative distribution functions:

$$H_0 : F_{X,Y}(x, y) = F_X(x)F_Y(y) \tag{7}$$

Blum's test makes use of a test statistic, B, that must be checked against a critical value, B_{cr}, at any specified significance level. This test statistic is computed as follows:

$$B = \frac{1}{2}\pi^4 N \sum_{j=1}^{N} \frac{\left(N_1 N_4 - N_2 N_3\right)^2}{N^5} \tag{8}$$

where N is the sample size for both X and Y. The quantities, N_1 to N_4, are computed for all values of j from 1 to N or effectively for all choices of $(X, Y) = (x_j, y_j)$ such that

N_1: number of (x, y) pairs such that $x \le x_j$ and $y \le y_j$.
N_2: number of (x, y) pairs such that $x > x_j$ and $y \le y_j$.
N_3: number of (x, y) pairs such that $x \le x_j$ and $y > y_j$.
N_4: number of (x, y) pairs such that $x > x_j$ and $y > y_j$.

If the value of B value as computed from Eq. (8) is higher than B_{cr}, then the null hypothesis is rejected and the two variables, X and Y, are not independent at the specified significance level.

Useful illustrative examples of the use of this test involve examining this B statistic for paired data that are known to be either strongly dependent or independent. Note that for a bivariate Gaussian distribution, correlation implies independence (the distribution is completely defined by a correlation coefficient and the first two marginal moments of each variable). If Blum's test is carried out for two jointly distributed Gaussian random variables that are strongly correlated, the B statistic is likely to be large; the opposite is true if the correlation is weak.

Variables, X and Y, assumed jointly Gaussian with a correlation coefficient of 0.9 were simulated; a scatter plot of the data is shown in Figure 6(a). Note that for any (x_j, y_j) pair that is part of this data set (where by design, X and Y, are strongly correlated and, thus, dependent), the values of N_1 and N_4 are generally much larger than N_2 and N_3; hence, the computed B value is large. For this case, B is equal to 31.8, which is much larger than the critical value, B_{cr}, of 4.23 at a 1% significance level. Hence, the independence (null) hypothesis is rejected. Another extreme case is considered where the variables, X and Y, assumed jointly Gaussian with a correlation coefficient of 0.05 were simulated; a scatter plot of the data is shown in Figure 6(b). Note that for any (x_j, y_j) pair that is part of this data set, this time, the values of N_1 and N_4 are generally of similar values to those of N_2 and N_3; hence, the computed B value is relatively smaller than in the first case. For this case, B is equal to 2.10, which is smaller than the critical value, B_{cr}, of 4.23 at a 1% significance level. Hence, the null hypothesis of independence is not rejected. These two examples serve to illustrate the use of Blum's test for independence between two random variables in general.

The independence of block maxima for different block sizes may be studied in a similar manner to that used for the preceding illustrative example. Blum's test statistic, the B value, for block maxima may be computed by forming lag-one vectors, X and Y, from all the block maxima in each ten-minute time series. The B value may be computed for these lag-one extremes to test if they are independent. It is expected that these extremes will become more independent as the block size is increased. At a certain optimum block size, computed B values will fall below the critical value, B_{cr}. While it is possible to study the B values for each simulation corresponding to a given wind speed, it is more instructive to study these B values (and, thus, independence) statistically as a function of block size by considering multiple simulations at a time in each wind speed bin; this makes it possible to consider scatter or uncertainty in the B values over different simulations. To this end, the mean, μ_B, and standard deviation, σ_B, of the B values from 200 simulations are computed for four different load measures, OOPBM, FATBM, OOPTD, and IPBM. Note that even with a smaller number of simulations, on the order of 15 to 20, statistics of the B values are quite stable and 200 simulations are not really needed. The mean B values with error bars representing one standard deviation are shown in Figure 7 for the four load types and for three different wind speed bins, 10-12 m/s, 16-18 m/s, and 22-24 m/s. As expected, mean values of B decrease monotonically with increasing block size. Even if the more stringent $(\mu_B+\sigma_B)$ level is checked versus the critical value, B_{cr}, at the 1% significance level, independence of block maxima is virtually assured for block sizes longer than 30 seconds for all four load types and in all three wind speed bins. Summarized in Table 1 are the appropriate block sizes based on criteria where either μ_B or $(\mu_B+\sigma_B)$ values are compared with B_{cr} at the 1% significance level. Clearly, loads in some wind speed bins exhibit greater dependence than others (e.g., low wind speed bins), but it appears that, at least with the LE^3 loads data set, one could

safely use block sizes of around 40-60 seconds, extracting between 10 and 15 extremes (block maxima) from each ten-minute time series and use these extremes to establish short-term load distributions.

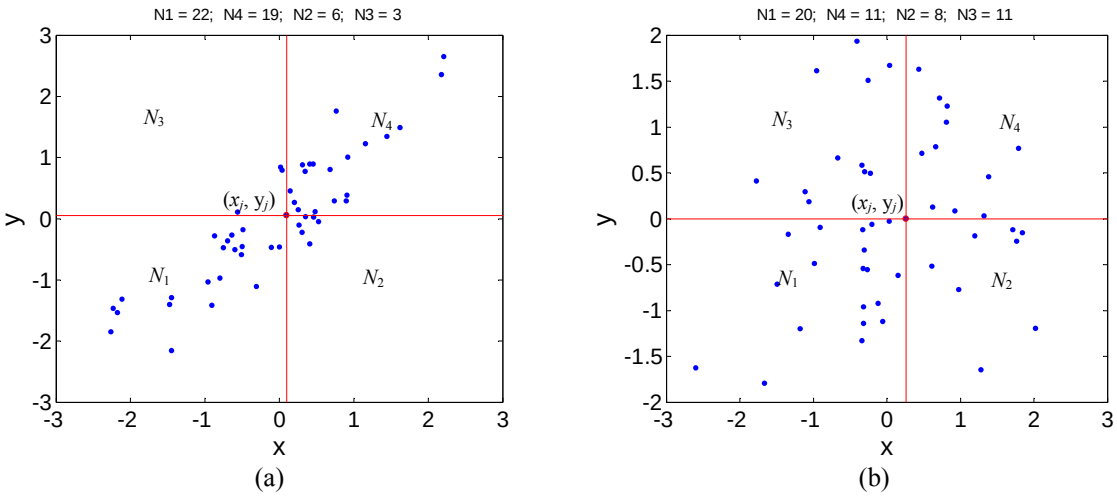


Figure 6: Scatter plot of simulated samples of two bivariate Gaussian random variables with correlation coefficients of (a) 0.9 and (b) 0.05. Also indicated are the values of N_1, N_2, N_3, and N_4 as computed when carrying out Blum's test for independence.

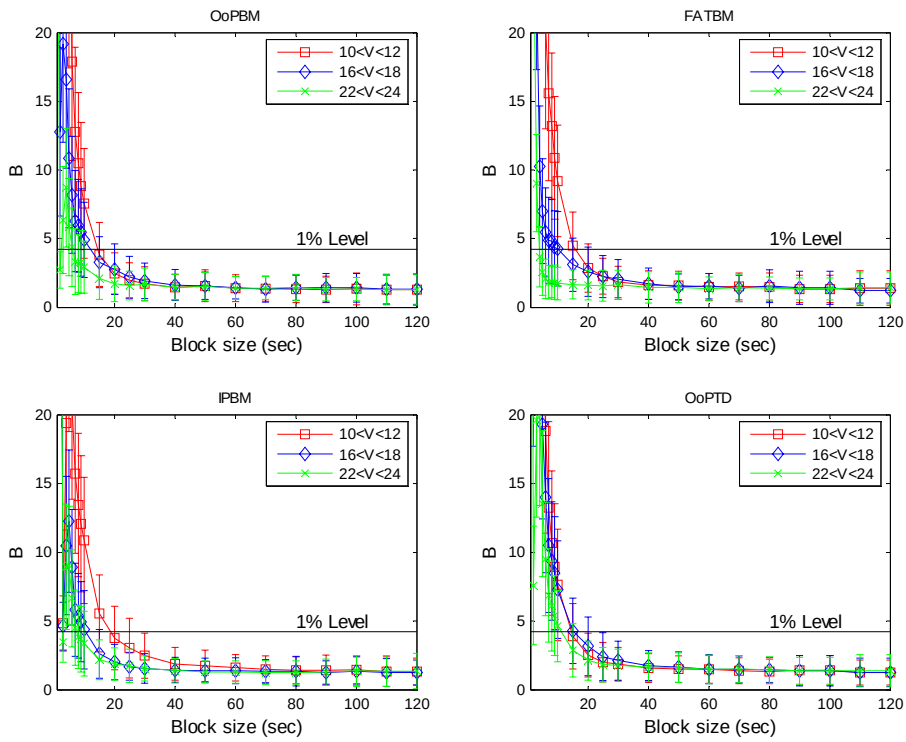


Figure 7: Variation of Blum's test *B* statistic for four loads as a function of block size (computed from 200 ten-minute time series for each load type and in three wind speed bins).

Table 1: Suggested block sizes (in seconds) for independent block maxima based on mean (μ_B) and mean plus one standard deviation ($\mu_B + \sigma_B$) values from 200 simulations and tested at the 1% significance level.

Wind speed (m/s)	OOPBM		FATBM		IPBM		OOPTD	
	μ_B	(μ_B+σ_B)	μ_B	(μ_B+σ_B)	μ_B	(μ_B+σ_B)	μ_B	(μ_B+σ_B)
2<V<4	50	70	30	50	30	60	50	70
4<V<6	40	60	25	40	40	60	40	60
6<V<8	40	60	30	50	30	50	40	60
8<V<10	40	50	40	50	25	40	40	50
10<V<12	15	20	20	25	20	30	15	20
12<V<14	20	30	20	30	15	20	25	40
14<V<16	20	30	20	30	15	20	20	30
16<V<18	15	25	15	25	15	20	20	25
18<V<20	15	20	6	15	10	15	15	25
20<V<22	7	20	5	6	9	15	15	20
22<V<24	7	15	4	5	8	15	15	20
24<V<26	6	15	4	5	7	15	15	20

B. Discussion on Independence

Returning to our earlier discussion on understanding which wind speed bins bring about the largest loads, it was suggested that lower wind speeds contribute almost no useful information to the tails of long-term load distributions. As such, increasing block sizes to ensure independence in these low wind speed bins offers no benefit to our ultimate goal of statistical loads extrapolation. Table 2 shows that if only the most important wind speed bins are considered (the ones that cause the largest loads), block maxima from block sizes as short as 10, 10, 15, and 15 seconds may be considered acceptably independent for OOPBM, FATBM, IPBM, and OOPTD, respectively (based on comparing μ_B levels from 20 simulations versus the 1% significance level B_{cr} value of 4.23).

Table 2: Average values of *B* from block maxima with different block sizes based on 20 simulations of four different loads for the most important wind speed bins.

Block Size (sec)	OOPBM	FATBM	IPBM	OOPTD
	16<V<18	16<V<18	24<V<26	16<V<18
5	9.65	6.91	8.37	18.50
10	4.19	3.72	4.52	6.42
15	3.18	2.73	2.31	3.81
20	3.15	2.95	1.90	3.19
30	2.15	2.12	1.79	2.34
60	1.34	1.39	1.33	1.53

Note: The critical value, B_{cr}, at the 1% significance level is 4.23.

It is instructive to see how a simpler statistical measure, such as the sample correlation coefficient between lag-one extremes, *X* and *Y*, as used before in the independence test, varies as the block size is changed. Such measures are easier to use than the Blum's test for independence described above; this makes them appealing for studying independence. However, it is important to note that an indication of lack of correlation does not guarantee independence. Nevertheless, sample correlation coefficients of lag-one block maxima (averaged over 20 simulations) for four load types and for the most important wind speed bins for each load are presented in Table 3. It is clear that sample correlation coefficient values on lag-one block maxima, as was the case with *B* values, generally decrease with increasing block size. However, no obvious acceptable correlation coefficient level on lag-one extremes can be claimed as a demarcation point for accepting lack of correlation among block maxima. While no attempt is made to address this issue, it can be seen that for the optimum block lengths needed to accept the independence criterion at the 1% significance level (see Table 2), correlation coefficients are estimated to be 0.33, 0.32, 0.17, and 0.35, respectively, for OOPBM, FATBM, IPBM and OOPTD. Thus, one might make a very general

statement, at least based on this LE^3 loads data set, that if correlation coefficients are about 30% or smaller between lag-one block maxima, then the selected block sizes selected reasonably well lead to independent block maxima. Note, however, that there is no implied direct relation—theoretical or empirical—between independence test *B* values and correlation coefficients.

Table 3: Averaged sample correlation coefficients from lag-one block maxima with different block sizes based on 20 simulations of four different loads for the most important wind speed bins.

Block size (sec)	OoPBM	FATBM	IPBM	OoPTD
	16<V<18	16<V<18	22<V<24	16<V<18
5	0.39	0.33	0.34	0.53
10	0.33	0.32	0.27	0.41
15	0.29	0.24	0.17	0.35
20	0.26	0.24	0.11	0.30
30	0.22	0.20	0.09	0.25
60	0.11	0.03	0.03	0.11

It has already been established that independence of block maxima is required for the mathematical relationships expressed in Eqs. (4) to (6) to hold. However, it is also important and perhaps more useful to discuss what effect if any the assumption of independence (rightly or wrongly) has on extrapolated loads. This can be done by studying the effect of assuming independence, or of assuming a block size assures independence, on the tails of short-term load distributions. Table 4 shows estimates of the 84th percentile global maximum load for four load types obtained using block maxima distributions along with the fractile adjustment given by Eq. (6). Block sizes are varied from 5 to 60 seconds for the construction of this table.

Table 4: Estimates of the 84th percentile global maximum load for four different load types as obtained from block maxima distributions using different block sizes.

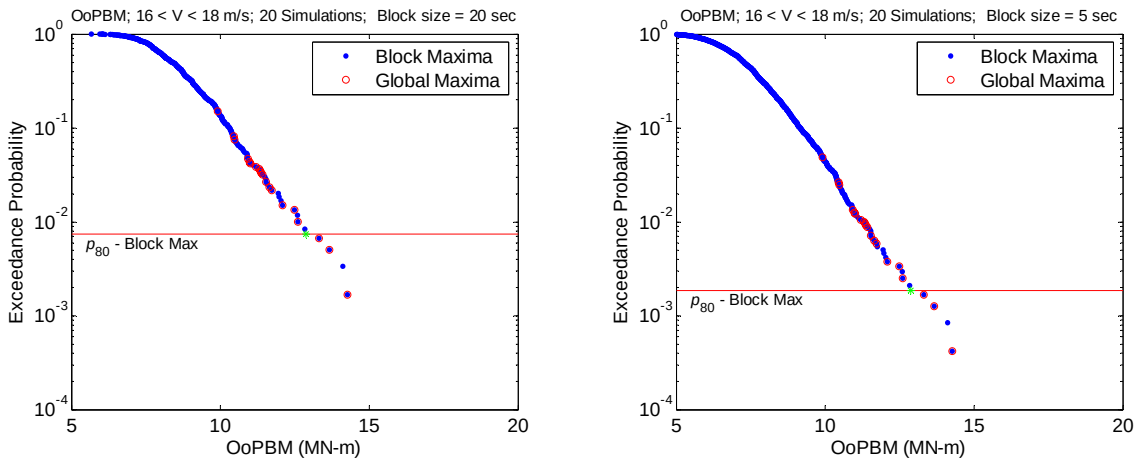
Block size (sec)	OOPBM	FATBM	IPBM	OOPTD
	MN-m	MN-m	MN-m	m
	16<V<18	16<V<18	24<V<26	16<V<18
5	12.87	78.91	7.47	7.31
10	12.87	78.92	7.47	7.31
15	12.87	78.92	7.47	7.31
20	12.88	78.93	7.47	7.31
30	12.88	78.94	7.47	7.32
60	12.89	78.97	7.48	7.32

Examining Table 4 reveals that the selection of block size has almost no effect on short-term loads at fairly rare fractile levels. This suggests that while smaller block sizes do not guarantee independence, they do not predict grossly inaccurate loads, at least at the 84th non-exceedance probability level for global maxima. This suggests that one could use very small block sizes without concern for independence but no new information is gained by the significantly larger sample of extremes that would result with this small block size. Stated differently, the findings summarized in Table 4, based on this LE^3 data set, suggest somewhat surprisingly that enforcing independence among the block maxima will likely have no significant impact on long-term aggregated distributions and hence on extrapolation. Figure 8 illustrates this issue by showing two short-term empirical distributions based on two different block sizes, one of 20-second duration that guarantees independent block maxima and the other of 5-second duration that almost certainly does not.

Short-term distributions on block maxima for the two block sizes are adjusted based on Eq. (5) so as to be directly represented in terms of exceedance probability in ten minutes. From Fig. 8, it is seen that even with a block length as short as 5 seconds, the tail of the short-term distribution of OOPBM for this 16-18 m/s wind speed bin is almost identical to the one obtained by using global maxima directly from 20 simulations. This confirms our statement that assumption of independence among block maxima, even if not justified, has insignificant impact on

predicted rare load fractiles. At the same time, this leads one to the conclusion that empirical distributions based on the use of larger sized samples of block maxima offer no advantages over the use of smaller samples of global maxima. It should be pointed out that these findings are based only on studies with the LE^3 loads data sets.

Finally, we might state that instead of focusing on the question of independence among extremes, it may be more beneficial to focus on establishing stable tails of the short-term empirical loads distributions. This point has been emphasized several times where we have focused attention on the importance of short-term distribution tails and their propagation to the aggregated long-term distribution curve and its use in extrapolation. In the following, we discuss procedures for controlling the uncertainty associated with these rare empirical short-term load fractiles.



(a) (b)

Figure 8: Short-term distributions for OOPBM for (a) a 20-second block size where independence is verified (see Table 2); and (b) a 5-second size where block maxima are dependent.

VI. Convergence Criteria

From the preceding discussions, it has been established that adjusting the block size when extracting extremes data from load time series has little impact on the tails of short-term distributions. Accordingly, now our focus shifts to whether or not the use of additional simulations, with global maxima extracted from each simulation, can help to better define distribution tails and, hence, the aggregated long-term distribution and the extrapolated rare load. It is of interest to be able to estimate how many global maxima (or simulations) would be required to adequately define rare load fractiles. It is expected that with a larger number of simulations, additional useful information about rare large loads can be gained. This should better define short-term load distributions for two reasons. Firstly, additional simulations will realize more maxima and generally some very large loads that can help to fill out and better define the tail of the distribution; this will results in reduced uncertainty in the aggregated distribution and, ultimately, in the extrapolated result. Secondly, with additional simulations, allow estimation of lower and lower probabilities of exceedance leading to a reduction in the extent of extrapolation needed beyond the simulated loads data.

However, the real issue with running simulations is one of practicality. Carrying out a large number of simulations can be time-consuming; hence, it would be beneficial to know what minimum number of simulations is needed to limit the uncertainty in load predictions to some specified level. A proposal is to enforce a convergence criterion on the tail fractiles of the empirical short-term distributions, not on the long-term distributions nor on the extrapolated long-term (50-year) load itself; for instance, one could require that the uncertainty in the p-quantile load of the short-term load distribution must be no larger than some specified amount. This could be prescribed by requiring a limit on confidence intervals on the p-quantile load. We propose such a convergence criterion that may be expressed mathematically in the following manner:

$$\frac{\hat{L}_{\alpha,p} - \hat{L}_{(1-\alpha),p}}{\hat{L}_p} < \frac{q}{100} \tag{9}$$

where the denominator in Eq. (9) represents the empirical estimate of the p-quantile load (similar to $l(p)$ defined in Eq. (6) earlier) while the numerator represents the $(2\alpha\text{-}1)\%$ confidence interval on the p-quantile load. The right-hand side of the inequality contains the variable, q, which represents the maximum acceptable percent error

permitted on the normalized confidence interval (where normalization is with respect to the p-quantile load itself). The convergence criterion, as stated, implies that if the normalized confidence interval based on a certain number of simulations exceeds the specified maximum error, additional simulations will need to be run to reduce this normalized confidence interval. The rationale behind specifying the convergence criterion in this manner is that if the p-quantile is chosen to be reasonably far in the tail of the short-term distribution, uncertainty in its estimate can be controlled or limited. Since the tail quite directly influences the long-term aggregated distribution and the extrapolated load, the convergence criterion albeit indirectly aids in the overall purpose of statistical loads extrapolation. We need to next discuss how the confidence interval in the numerator of Eq. (9) can be estimated from data and how should the maximum percent error, q, be selected. We need to be cognizant of the excessive level of effort that may be needed if q is specified very small or if p is specified as a quantile level that is too far in the tail.

A. Confidence Intervals based on Bootstrapping

Estimates of the confidence interval of the p-quantile load are related to the number of simulations or data points used in estimating that load. The bootstrap technique proposed by Efron and Tibshirani[10,11] makes it possible to estimate such confidence intervals by a process that involves randomly resampling data, with replacement, many times. Efron and Tibshirani[11] point out that the bootstrapping is a computer-intensive procedure since the data set may need to be resampled thousands of times in order to accurately estimate statistics such as confidence intervals on quantities such as quantiles of a distribution. Henderson[12] provides an excellent review of the bootstrap method and offers many examples of its implementation and subtleties involved in its use.

Using the bootstrap procedure to form confidence intervals begins with taking the initial set of data on, say, n global maxima ($m_1, m_2, m_3, m_4, m_5 \ldots m_n$) and randomly resampling these data with replacement to form a new set ($m_1^*, m_2^*, m_3^*, m_4^*, m_5^* \ldots m_n^*$) or a bootstrap resampling of the same size as the original sample. Note that bootstrap resamplings will be composed of repeated values from the original sample since, for each resampling, data are sampled randomly with replacement. The process is repeated so as to form a large number, N_b, of bootstrap resamplings. From each of these sets of n data, individual estimates of the p-quantile can be obtained. From these N_b estimates, constituting the set, ($l_1, l_2, l_3, l_4, l_5 \ldots l_{N_b}$), confidence intervals can be found in the usual manner by ordering the data. These can then be used for the numerator of Eq. (9). The estimate of the p-quantile that is obtained from the original data represents the denominator of Eq. (9).

It is worthwhile to discuss the need to perform a sufficiently large number of bootstrap resamplings to obtain reliable estimates of confidence intervals. The literature suggests that sometimes as few as 25 bootstrap resamplings may be sufficient to estimate some statistics; however, it has also been pointed out that many more resamplings are required to estimate confidence intervals[10]. Lunneborg[13] has suggested that bootstrap resamplings be increased incrementally until some stability results in estimates of the standard error. However, as pointed out earlier, bootstrapping is a computer-intensive method and the question of how many bootstrap resamplings are required is perhaps not so important since it is just as easy to carry out 5,000 resamplings as it is carry out 25; the benefit of a the larger number is that it will lead to more reliable estimates of the desired statistic. See Chernik[14] for further discussion on this issue. In the present study, 5,000 bootstrap resamplings were used to form confidence interval estimates on load quantiles and these were found to be adequately stable. Figure 9 illustrates differences between confidence interval estimates on the 0.84-quantile OOPBM load when a small number of bootstrap resamplings is used (25 in the top figure) compared to when a large number is used (5,000 in the bottom figure). Variability in estimates of the normalized 90% confidence interval width is obvious with the smaller number as is evident by running the same bootstrap procedure 10 times. With the smaller number of bootstrap resamplings, 90% confidence intervals on a 0.84-quantile load from different runs can vary greatly from the stable estimates obtained with the larger number; deviations are smaller with the larger number of bootstrap resamplings.

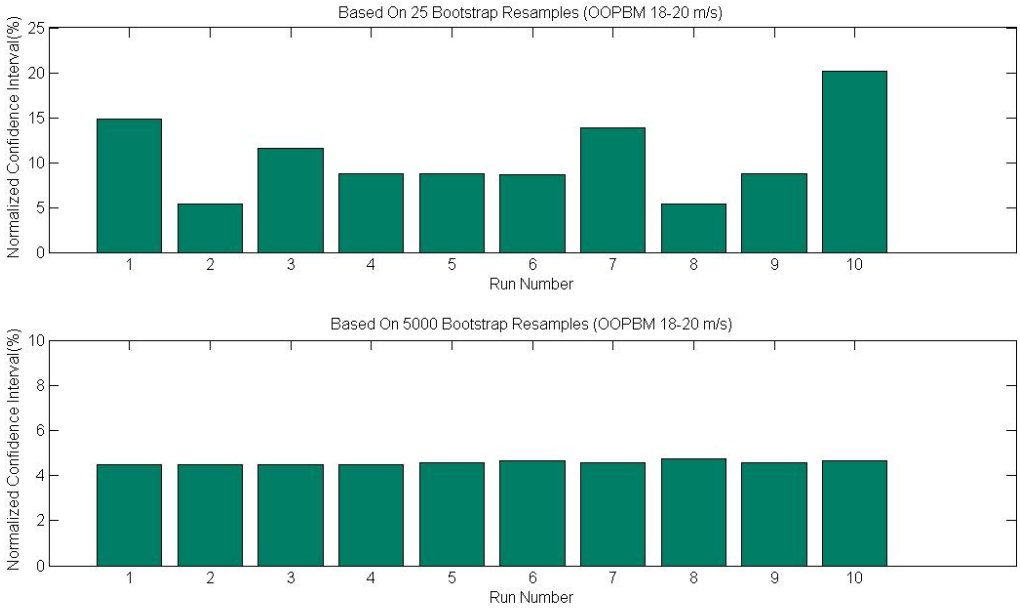


Figure 9: Normalized 90% confidence interval estimates on the 0.84-quantile global maximum OOPBM load in the 18-20 m/s wind speed bin based on 30 simulations followed by 25 bootstrap resamplings (top figure) and 5,000 resamplings (bottom figure).

B. Confidence Intervals based on the Binomial Distribution

As an alternative to the bootstrap procedure discussed above, the binomial distribution may also be used to obtain confidence interval estimates on the p-quantile load[15]. This can limit the computational effort necessary when evaluating the numerator in Eq. (9). The theoretical development is presented here.

We start by writing the formula for the binomial probability mass function, $B(j;\ m,\ p)$, which expresses the probability of i occurrences of an event of interest in m Bernoulli trials when the probability of occurrence of the event in any single trial is p.

$$B(i;m,p)=\frac{m!}{i!(m-i)!}p^{i}(1-p)^{m-i};\quad i=0,1,2\ldots m \tag{10}$$

In the present context, the event refers to the non-exceedance of the estimated p-quantile load while m refers to the number of aeroelastic simulations run as well as to the number of global maxima extracted. From the definition of the probability mass function in Eq. (10), a cumulative distribution function, $C(j;\ m,\ p)$, may be written as follows:

$$C(j;m,p)=\sum_{i=0}^{j}B(i;m,p);\quad j=0,1,2\ldots m \tag{11}$$

To form the confidence interval on the p-quantile load needed for the numerator of Eq. (9), two load levels, x_k and x_l, need to be found such that the following is true (where X refers to the p-quantile load):

$$P(x_k<X<x_l)=2\alpha-1 \tag{12}$$

To simplify notation, the quantities, $\hat{L}_{(1-\alpha),p}$ and $\hat{L}_{\alpha,p}$, in Eq. (9) have been replaced by x_k and x_l, respectively, in Eq. (12). The load levels, x_k and x_l, need to be obtained by searching the rank-ordered extremes, x_1^*, x_2^*, … x_m^*, from the m simulations and then interpolating. It is generally possible to find two integer values, k^* and l^*, where k^* is the largest integer such that

$$C(k^*;m,p)\le(1-\alpha) \tag{13}$$

and where l^* is the largest integer such that

$$C\left(l^*;m,p\right)\le \alpha \tag{14}$$

and $1 \le k^* < l^* \le m$-1.

The integers, k^* and l^*, are such that x_{k*} and x_{l*} bound the desired load levels, x_k and x_l, from below. This in turn means that $x_{k*} \le x_k \le x_{(k+1)*}$ and $x_{l*} \le x_l \le x_{(l+1)*}$. Once the integers, k^* and l^*, are found using Eqs. (13) and (14), the load levels, x_k and x_l, may be found by interpolation as follows:

$$x_k = \frac{x_{(k+1)^*} - x_{k^*}}{C_{(k+1)^*} - C_{k^*}}\left[\left(1-\alpha\right) - C_{k^*}\right] + x_{k^*} \tag{15}$$

$$x_l = \frac{x_{(l+1)^*} - x_{l^*}}{C_{(l+1)^*} - C_{l^*}}\left[\alpha - C_{l^*}\right] + x_{l^*} \tag{16}$$

where, to simplify notation, we have set $C(k^*;m,p) = C_{k*}$, $C((k+1)^*;m,p) = C_{(k+1)*}$, $C(l^*;m,p) = C_{l*}$, and $C((l+1)^*;m,p) = C_{(l+1)*}$. The desired $(2\alpha$-1)% confidence interval required for Eq. (9) is simply equal to $(x_l - x_k)$.

C. Normal Approximation to the Binomial Distribution

Note that the binomial cumulative distribution function is needed to find the integer values of k^* and l^* as well as to obtain the values of x_k and x_l using Eqs. (15) and (16). It is possible to replace the binomial distributed integer random variable representing the number of occurrences in m trials (where the probability of event occurrence in a single trial is p) by a normally distributed real random variable with mean equal to mp and variance equal to $mp(1-p)$. If this is done, the confidence interval developed following the steps indicated by Eqs. (11) to (16) remains valid; the only difference is a simplification of Eq. (11) which may be approximated as follows:

$$C\left(j;m,p\right) \cong \Phi\left(z_j\right) \quad \text{where } z_j = \frac{(j+0.5)-mp}{\sqrt{mp(1-p)}} \tag{17}$$

In Eq. (17), $\Phi(\ )$ refers to the cumulative distribution function of a standard normal random variable. The steps involving determination of the values of k^* and l^* in Eqs. (13) and (14) represent the most significant advantage since it is much easier to find k^* and l^* by using the inverse cumulative distribution of a standard normal random variable than it is to evaluate the binomial cumulative distribution function using Eqs. (10) and (11). Note that the normal approximation to the binomial involves a continuity correction of 0.5 in the definition of z_j in Eq. (17) since we are replacing an integer random variable by a real one. Although the normal approximation to the binomial is most accurate when the values of mp and $m(1-p)$ are greater than 5, in other cases as well the approximation is reasonably accurate as we shall show.

D. Binomial Confidence Bounds Simplified for Wind Turbine Applications

The confidence intervals based on the binomial distribution as developed in Eqs. (10) to (16) are less computationally than those computed using the bootstrap procedure. It is possible, though, to simplify the binomial-based confidence intervals to an even greater extent for applications to statistical extrapolation of wind turbine extreme loads. Essentially, this is done by tabulating values of k^* and l^* that will result for most common situations where the number of simulations is on the order of 15 to 35 for each wind speed bin. This number of simulations will be shown to be reasonable if, in Eq. (9), the 90% confidence interval on the 84th percentile load is computed and the maximum error on the normalized confidence interval of Eq. (9) is to be less than 15% (i.e., q = 15). Note that the simplified approach to the binomial-based confidence interval can only be reasonably tabulated for specific values of p and α. For instance, for p equal to 0.84 and α equal to 0.95, Table 5 provides values of k^* and l^* as well as two other values, A and B, needed for interpolating as is done in Eqs. (15) and (16).

Table 5: Parameters needed to establish binomial-based confidence intervals (for α=0.95 and p=0.84).

For 90% Confidence Interval on the 84th Percentile Load	No. of sims.	k^*	l^*	A	B
	15	9	14	0.50	0.32
	16	10	15	0.27	0.19
	17	11	16	0.10	0.03
	18	11	16	0.87	0.96
	19	12	17	0.58	0.90
	20	13	18	0.35	0.83
	21	14	19	0.16	0.76
	22	14	20	1.00	0.69
	23	15	21	0.69	0.60
	24	16	22	0.45	0.50
	25	17	23	0.25	0.39
	26	18	24	0.08	0.26
	27	18	25	0.85	0.12
	28	19	25	0.58	0.98
	29	20	26	0.36	0.91
	30	21	27	0.18	0.83
	31	22	28	0.02	0.75
	32	22	29	0.75	0.66
	33	23	30	0.51	0.56
	34	24	31	0.31	0.44
	35	25	32	0.13	0.32

Table 5 works in conjunction with a design equation that is tailored to be used along with it. This design equation and Table 5 give the 90% confidence interval for the 84th percentile ten-minute maximum (i.e., a = 0.95 and p = 0.84 in Eq. (9)). The design equation can be written as follows:

$$\left(x_l - x_k\right) = \left(x_{l^*} - x_{k^*}\right) + B\left(x_{(l+1)^*} - x_{l^*}\right) - A\left(x_{(k+1)^*} - x_{k^*}\right) \tag{18}$$

where l^*, k^*, A, and B are given in Table 5 as a function of the number of simulations run and x_{l*}, $x_{(l+1)*}$, x_k^*, and $x_{(k+1)*}$ are obtained from the rank-ordered simulated extremes.

As an illustration of how Eq. (18) and Table 5 can be used, consider a situation where 20 simulations have been carried out. Then, $l^* = 18$, $k^* = 13$, $A = 0.35$, and $B = 0.83$ according to Table 5, and if the rank-ordered extremes, x_{13*}, x_{14*}, x_{18*}, and x_{19*} are found, Eq. (18) can be used to compute the confidence interval needed for Eq. (9). If the convergence criterion is not met according to Eq. (9), additional simulations may be run and new values of l^*, k^*, A, and B obtained again from Table 5. This may be repeated until the convergence criterion is met.

E. Application of the Convergence Criteria to the LE^3 Loads Data Set

Converegence criteria for four load measures, OOPBM, FATBM, OOPTD, and IPBM were studied using the LE^3 loads database. Based on Eq. (9), we are interested in computing the percent error in terms of the normalized 90% confidence interval of the 84th percentile ten-minute global maximum for each load type for different numbers of simulations. If the maximum allowable percent error, q, is specified, the appropriate number of simulations can be run. Tables 6(a)-(d) show computed percent errors when 30 simulations are run for each wind speed bin and for all of the four loads. The results presented in these tables suggest that the convergence criterion is adequately met if the maximum error permitted is 15% (i.e., if q is equal to 15). For IPBM loads, even a 10% maximum error criterion would be met when 30 simulations are run. For the OOPBM and OOPTD, slowest convergence is seen in the 16-18 m/s wind speed bin but even there, 30 simulations lead to normalized 90% confidence intervals on the84th percentile load that are smaller than 15%.

Table 6(a): Estimates of the 84th percentile load, x_k, x_l, and the binomial-based normalized 90% confidence interval for OOPBM based on 30 simulations.

OOPBM	x_{84} (MN-m)	x_k (MN-m)	x_l (MN-m)	$(x_l - x_k)/x_{84}$ (%)
2<*V*<4	3.50	3.36	3.55	5.5
4<*V*<6	5.82	5.54	5.90	6.2
6<*V*<8	9.09	8.52	9.47	10.5
8<*V*<10	12.45	12.28	12.63	2.8
10<*V*<12	13.52	13.38	13.60	1.6
12<*V*<14	13.84	13.59	13.90	2.3
14<*V*<16	13.87	13.57	14.02	3.2
16<*V*<18	13.33	12.17	13.84	12.5
18<*V*<20	11.58	11.08	11.66	5.0
20<*V*<22	10.83	10.33	10.95	5.8
22<*V*<24	10.05	9.78	10.57	7.9
24<*V*<26	9.80	9.54	10.12	5.9

Table 6(b): Estimates of the 84th percentile load, x_k, x_l, and the binomial-based normalized 90% confidence interval for FATBM based on 30 simulations.

FATBM	x_{84} (MN-m)	x_k (MN-m)	x_l (MN-m)	$(x_l - x_k)/x_{84}$ (%)
2<*V*<4	24.75	23.56	25.52	7.9
4<*V*<6	35.82	35.52	36.00	1.3
6<*V*<8	51.51	48.94	53.59	9.0
8<*V*<10	73.30	72.18	74.97	3.8
10<*V*<12	80.51	79.27	80.72	1.8
12<*V*<14	85.04	83.20	86.78	4.2
14<*V*<16	84.06	82.82	85.49	3.2
16<*V*<18	81.16	78.20	83.07	6.0
18<*V*<20	71.89	67.24	72.69	7.6
20<*V*<22	62.05	60.55	64.95	7.1
22<*V*<24	63.45	59.23	64.80	8.8
24<*V*<26	61.84	59.09	62.59	5.7

Table 6(c): Estimates of the 84th percentile load, x_k, x_l, and the binomial-based normalized 90% confidence interval for OOPTD based on 30 simulations.

OOPTD	x_{84} (m)	x_k (m)	x_l (m)	$(x_l - x_k)/x_{84}$ (%)
2<*V*<4	2.08	1.93	2.14	10.2
4<*V*<6	3.35	3.24	3.47	6.8
6<*V*<8	5.05	4.80	5.25	9.0
8<*V*<10	6.93	6.87	7.04	2.5
10<*V*<12	7.73	7.48	7.78	3.9
12<*V*<14	7.74	7.65	7.77	1.6
14<*V*<16	7.69	7.45	7.73	3.7
16<*V*<18	7.52	6.56	7.69	15.0
18<*V*<20	6.04	5.71	6.36	10.7
20<*V*<22	5.52	5.17	5.65	8.8
22<*V*<24	4.94	4.52	4.95	8.6
24<*V*<26	4.45	4.18	4.78	13.5

Table 6(d): Estimates of the 84th percentile load, x_k, x_l, and the binomial-based normalized 90% confidence interval for IPBM based on 30 simulations.

IPBM	x_{84} (MN-m)	x_k (MN-m)	x_l (MN-m)	$(x_l - x_k)/x_{84}$ (%)
2<*V*<4	3.79	3.75	3.81	1.7
4<*V*<6	4.25	4.21	4.28	1.8
6<*V*<8	4.96	4.90	5.05	2.9
8<*V*<10	5.67	5.60	5.82	3.9
10<*V*<12	6.03	5.96	6.06	1.6
12<*V*<14	6.35	6.28	6.45	2.6
14<*V*<16	6.45	6.33	6.54	3.2
16<*V*<18	6.70	6.61	6.75	2.0
18<*V*<20	7.03	6.72	7.26	7.6
20<*V*<22	7.37	7.16	7.42	3.6
22<*V*<24	7.62	7.43	7.90	6.2
24<*V*<26	7.69	7.44	7.85	5.2

Short-term load distributions for the four load types are summarized in Fig. 10 for the single wind speed bin in each case that showed slowest convergence based upon the normalized 90% confidence interval on the 84th percentile load. The 90% confidence intervals are showed for each empirical distribution at the (1 – 0.84) exceedance probability level. The 84th percentile load is shown along with confidence interval based on the bootstrap method as well as the binomial method.

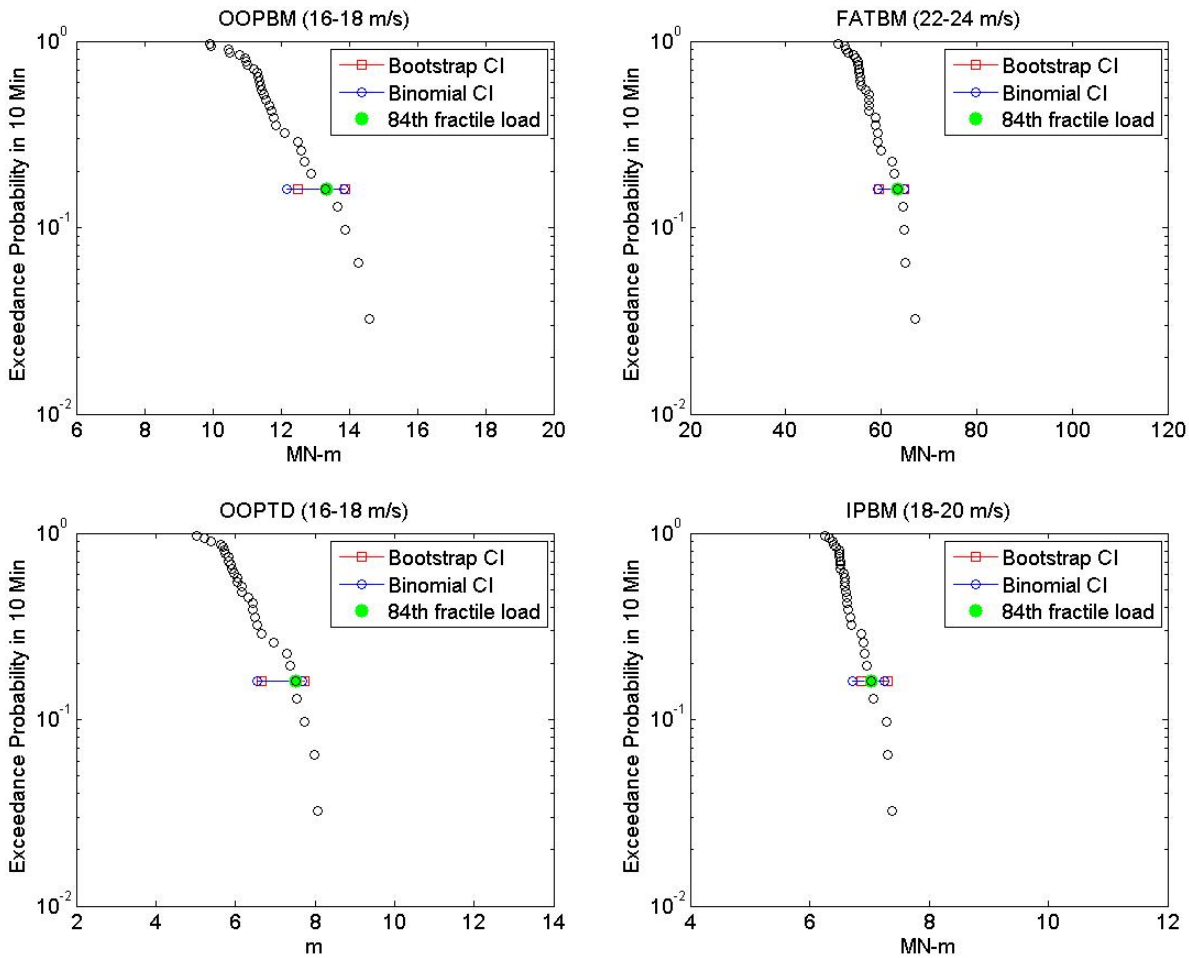


Figure 10: Short-term distributions for OOPBM, FATBM, OOPTD, and IPBM based on 30 simulations. Bins selected have largest 90% relative confidence bounds (RCBs) on the 84th percentile load. The 84th percentile load is shown as are binomial- and bootstrap-based confidence intervals.

The preceding discussion suggests that, for the LE^3 data set, adequately stable tails for the short-term distributions in all wind speed bins can be obtained if 30 simulations are run. Enforcing the convergence criterion of a maximum percent error of 15% on the normalized 90% confidence interval on the 84th percentile load in the short-term distributions leads us to state that these distributions are reasonably well estimated. The next and more important issue to study with regard to extrapolation is whether or not the controlled uncertainty in short-term distributions propagates to stable aggregated long-term distributions as well. Tail stability of aggregate long-term distributions for each load type can be evaluated using bootstrap procedures in a similar manner to that employed for the short-term distributions. However, since the aggregated distribution involves weighting based on the Rayleigh wind speed distribution, it is not possible to bootstrap the data in this long-term distribution directly. Instead, the short-term distributions are each bootstrapped separately and then for each wind speed bin's bootstrap resampling, the aggregation is carried out leading to multiple long-term distribution curves. From these multiple long-term probability distributions, it is possible to evaluate confidence intervals on any desired long-term load quantile.

Figure 11 shows bootstrap-based aggregated long-term distribution curves for four load types using the number of simulations required to optimally satisfy convergence with q = 15, α = 0.95, and p = 0.84 in Eq. (9). Representative 90% confidence bounds reflecting the range of bootstrap-based loads predicted at a given fractile are shown where these bounds are large. These confidence bounds suggest that the aggregated long-term distribution has very low uncertainty compared to the short-term distributions. This result might be expected since not all wind speed bins' short-term distributions exhibit large tail uncertainty (as can be verified from Table 6(a) to (d)). All the wind speed bins are aggregated in the long-term distribution but are weighted to different degrees according to the Rayleigh distribution. The worst wind speed for convergence at the short-term level is diminished (in relative

terms) in importance at the long-term level. This is likely why long-term distributions appear fairly stable and have low uncertainty.

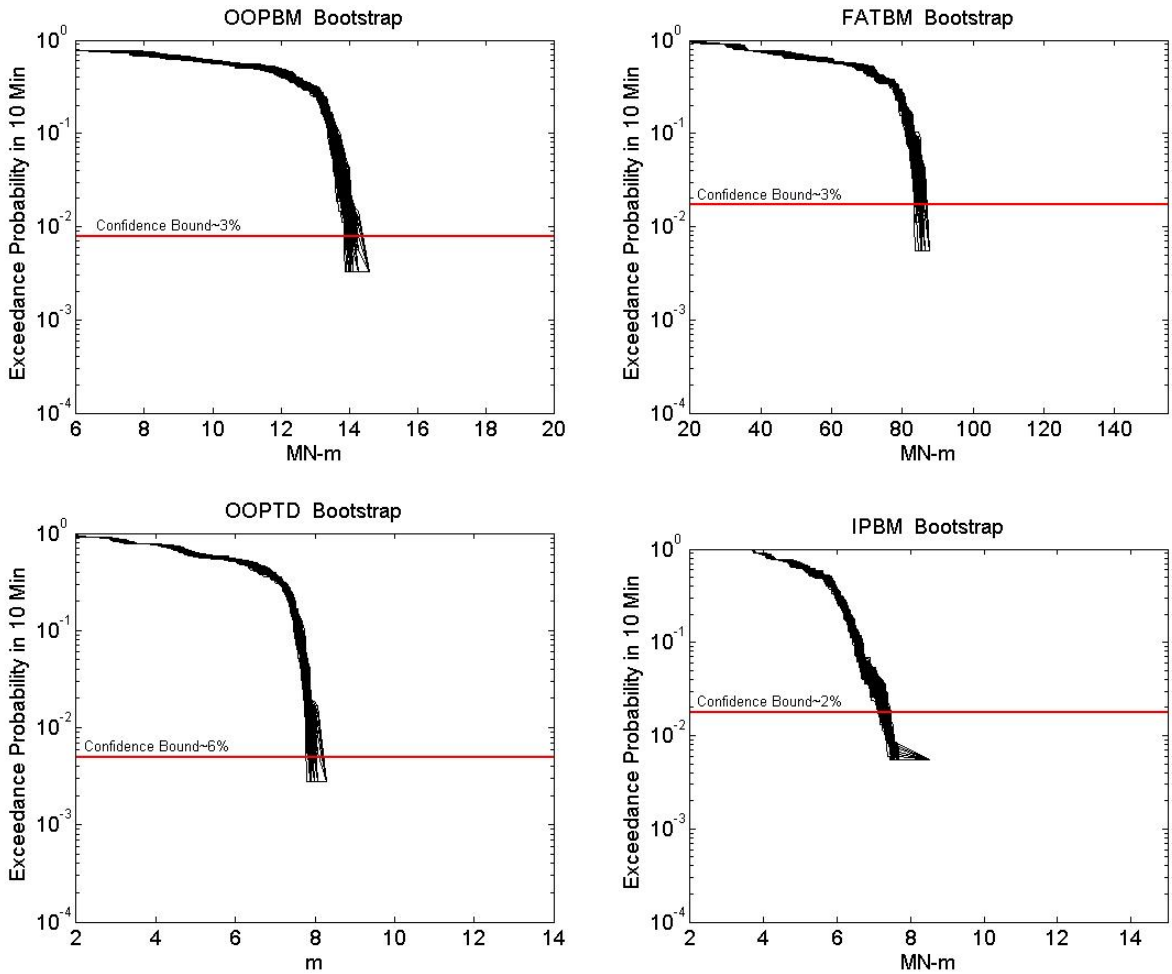


Figure 11: Bootstrap-based aggregated long-term distribution curves for OOPBM, FATBM, OOPTD, and IPBM. Representative error bounds reflecting the spread in the bootstrap curves are shown.

In closing our discussion on convergence criteria, we note that when a maximum error of 15% was imposed for the short-term distributions (in conjunction with 90% confidence intervals of the 84th percentile load), 90% confidence bounds of 6% or smaller resulted in the long-term distributions. It is important to note, though, that these results were based on studies of four loads and came from one wind turbine model alone. If simulated loads data from other turbines are considered, results might be different; therefore, it is suggested that the value of q in Eq. (9) be adjusted as appropriate so as to not cause excessive amount of simulations. The convergence criterion triad ($\alpha = 0.95$, $p = 0.84$, $q = 15$) for short-term distributions has been demonstrated to yield stable tails of the short-term distributions as well as stable long-term distributions. It was found that if the short-term distributions are verified through convergence checks, uncertainty in aggregated long-term distributions is significantly lower.

VII. Real Uncertainty in Short-Term Loads

We have shown that confidence intervals on any load quantile at the short-term level may be estimated using bootstrap and binomial-based methods. These estimates rely on a limited single set of simulated data. It is possible, though, to examine confidence intervals and evaluate convergence criteria by using the entire LE^3 loads data set. If all the 1,200 global maxima for each wind speed bin are employed, real uncertainty can be estimated directly without resorting to methods such as bootstrapping. For example, this may be done by subdividing the set of 1,200

global maxima per bin into 40 individual sets of 30 maxima. Then, real confidence intervals of the *p*-quantile load may be found by extracting 40 estimates of the *p*-quantile load. The convergence criterion based on Eq. (9) can then be checked using this confidence interval. Since all the data here are real wind turbine maxima, not statistical estimates, the 90% confidence interval was normalized by the mean of these values. Figure 12 shows a comparison of confidence intervals for four loads (OOPBM, FATBM, OOPTD< and IPBM) based on 30 simulations except in the "Real" case which was based on 1,200 simulations. The figure shows no consistent trend or difference between confidence intervals based on real data versus those based on bootstrapping and the binomial approach. Note that the figure also shows that the normal approximation to the binomial is generally quite good for estimation of confidence intervals.

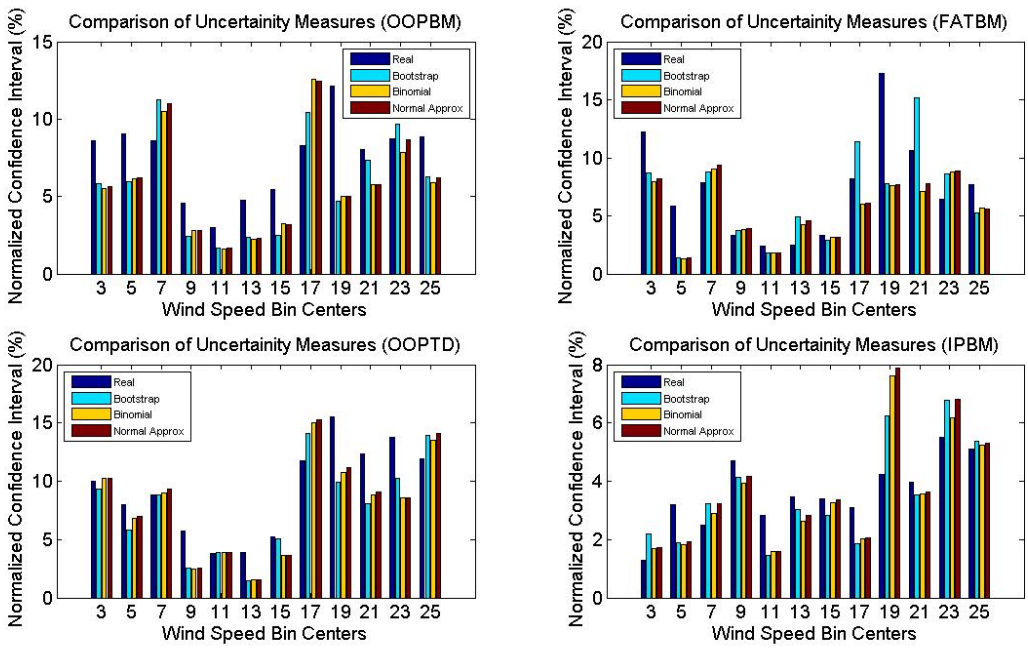


Figure 12: Normalized 90% confidence intervals on the 84th percentile load based on 1,200 simulations (Real Data) as well as based on 30 simulations followed by bootstrapping, a binomial-based method, and a normal approximation to the binomial. Results are shows for four load types (OOPBM, FATBM, OOPTD, and IPBM) and for all wind speed bins.

The 40 sets of 30 real global maxima can be studied in comparison with the bootstrap estimates of the error bounds on the aggregated long-term distributions. This is done by aggregating one full set of real maxima at a time across all the wind speed bins and then repeating this procedure for all the 40 sets. In all, 40 aggregated long-term distribution curves and error bounds on each can be computed at any desired load quantile. Figure 13 shows that the bootstrap estimates of uncertainty in the aggregated long-term distribution differ from the true variability only very slightly (by less than 2-3% in terms of the normalized 90% confidence interval on the same rare fractile levels indicated in Fig. 11). Even using real data, however, the normalized confidence intervals are still significantly lower than those found in the short-term distributions, reaffirming our conclusion that meeting the short-term convergence criteria leads to small error bounds in aggregated long-term load distributions.

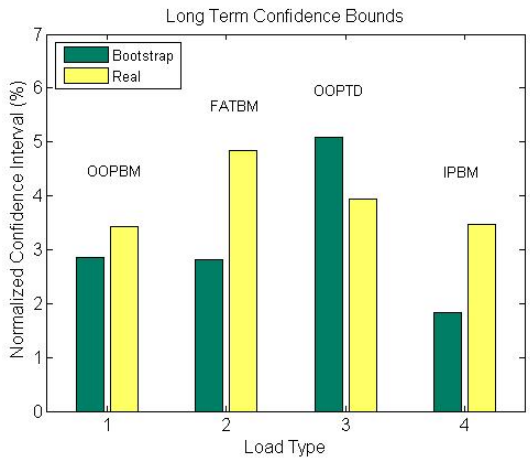


Figure 13: Normalized (relative) 90% confidence bounds on the aggregated long-term distribution for four load types (OOPBM, FATBM, OOPTD, and IPBM) based on real data as well as bootstrapping.

VIII. Conclusions

Using the LE^3 loads data, several questions related to the theory and practical implementation of statistical loads extrapolation have been addressed. First, we have emphasized the need to understand which wind speeds tend to cause largest loads of different types. This information is useful to have when determining where greater simulation effort is needed.

We have formulated the entire sequence of steps involved in predicting a long-term load of interest (such as the 50-year return period load) by first defining short-term distributions and then aggregating these to long-term distributions.

We have compared the use of global and block maxima and their differences in the context of statistical loads extrapolation. Issues related to independence of block maxima have been addressed and statistical tests of independence formulated. Ignoring unimportant wind speed bins, it was found that block sizes of around 10-15 seconds for four loads (OOPBM, FATBM, OOPTD, and IPBM) led to independent block maxima when checked at the 1% significant level based on a statistical test. Finally, though, with the LE^3 data, it was demonstrated that there is no advantage gained from using block maxima over global maxima when short-term loads are estimated. Moreover, even if block sizes are very small so as to exhibit dependence, ignoring this again leads to small error in estimation of short-term loads.

In order to assure stable or robust short-term distributions especially in the tails of these distributions, a convergence criterion was developed that relies on computing confidence intervals on rare load quantiles. Based on studies with the LE^3 data, the convergence criterion that is proposed is that the normalized 90% confidence interval on the 84th percentile load (normalization is with respect to the 84th percentile load estimate from the simulations) may not exceed 15%. The 90% confidence interval that forms part of the convergence criterion may be estimated based on bootstrap methods as well as the binomial distribution. Both procedures have been developed here. Additionally, for the binomial method, an approximation using the normal distribution was presented. Finally, a design equation and additional tabulated parameters were presented that enable quick computation of the normalized 90% confidence interval on the 84th percentile load as long as the number of simulations run is between 15 and 35. The convergence criteria were applied for four loads (OOPBM, FATBM, OOPTD, and IPBM) and with 30 simulations, the normalized 90% confidence interval on the 84th percentile load never exceeded 15%. Bootstrap and binomial-based confidence intervals were reasonably similar.

The convergence criteria applied to the short-term loads distributions were verified to lead to smaller uncertainty in aggregated long-term load distributions. Small uncertainty in long-term distributions is expected to lead to good, robust predictions of extrapolated rare loads although that was not a focus of this study.

In closing, it is important to note that all of these conclusions were derived based on findings from the LE^3 data set and the simulated loads data are for the LE^3 5MW turbine model alone. It is possible that the study of loads from other turbines will lead to different conclusions. Nevertheless, the focus of this study was to develop several new ideas that, it is hoped, will aid in statistical loads extrapolation for practicing engineers.

Acknowledgments

The authors are pleased to acknowledge useful discussions with members of the LE^3 working group over the last year. In particular, they thank Dr. Pat Moriarty of the Natural Renewable Energy Laboratory for the simulated loads data and Dr. Paul Veers of Sandia National Laboratories for continued encouragement and advice on various aspects of this work. Finally, the authors wish to express their gratitude for the financial support received from Sandia National Laboratories (by way of Contract No: 743378) and from the National Science Foundation (by way of Career Award Grant No: CMMI-0449128).

References

[1]International Electrotechnical Commission, "IEC 61400-1: Wind Turbines – Part 1: Design Requirements," Edition 3, 2005.

[2]Jonkman, J. M., Butterfield, S., Musial, W. and Scott, G., "Definition of a 5MW Reference Wind Turbine for Offshore System Development," NREL/TP-500-38060, National Renewable Energy Laboratory, Golden, Colorado, (to be published).

[3]Moriarty, P., "Database for Validation of IEC Extrapolation Techniques," Unpublished manuscript submitted to Wind Energy, 2007.

[4]Jonkman, B. J. and Buhl, M. L. Jr, "TurbSim User's Guide," NREL/TP- 500-41136, National Renewable Energy Laboratory, Golden, Colorado, 2007.

[5]Jonkman, J. M. and Buhl, M. L. Jr, "FAST User's Guide," NREL/EL- 500-38230, National Renewable Energy Laboratory, Golden, Colorado, 2005.

[6]Ragan and L. Manuel, "Statistical Extrapolation Methods for Estimating Wind Turbine Extreme Loads," Proceedings of the 45th AIAA Aerospace Sciences Meeting and Exhibit, AIAA-2007-1221, Reno, Nevada, January 2007.

[7]Blum, J.R., Kiefer, J. and Rosenblatt, M., "Distribution free tests of independence based on the sample distribution function," Ann. Math. Statist. Vol. 32, pp. 485-498, 1961.

[8]Hollander, M. and Wolfe, D.A., "Nonparametric Statistical Methods," Wiley, New York, 1999.

[9]Skaug, H.J. and Tjøstheim, D., "A nonparametric test of serial independence based on the empirical distribution function," Biometrika 80, pp. 591–602, 1993.

[10]Efron, and Tibshirani, "An Introduction to the Bootstrap," Chapman and Hall, New York, 1993.

[11]Efron, and Tibshirani, "Statistical Data Analysis in the Computer Age," Science, Vol. 253, No. 5018, pp. 390-395, 1991.

[12]Henderson, A.R., "The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data," Clinca Chimica Acta, Vol. 359, pp. 1-26, 2005.

[13]Lunneborg, C.E., "Data Analysis by Resampling: Concepts and Applications," Duxbury Press, Pacific Grove, California, 2000.

[14]Chernik, M. R., "Bootstrap Methods: A Practitioners Guide," John Wiley and Sons, New York, 1999.

[15]Hogg, R.V. and Craig, A., "Introduction to Mathematical Statistics," 5th Edition, Prentice Hall, 1995.