The risks of extreme load extrapolation

The authors present a study on the uncertainty inherent to statistical load extrapolation due to the limited sample sizes. The article is well written and easy to follow. However, there are several issues which are outlined in the following comments:

General comments

1) The distribution fit to a random process is representative only if the process is stationary and ergodic (i.e., its statistical properties can be sufficiently well determined over a given sampling period). Subsequently stationarity and ergodicity are central requirements for the validity of load extrapolations. In the present paper, these conditions are not taken into account and I am afraid that they are violated in many of the cases. As an example, for samples drawn from a Rayleigh distribution with mean μ, the standard deviation of the sample mean is defined as $\sigma_{\mu_s} = \mu\sqrt{(4-\pi)/(N\pi)}$, where N is the sample size. For $\mu = 8.5 m/s$ as in the present paper and for sample size $N = 1000$, the standard deviation of the sample mean wind speed will be 0.14m/s, which means that the 95% confidence interval will be 0.55m/s wide. In effect, samples of this size may be quite different from each other and will result in different extrapolations. This sampling uncertainty arises due to the sample size being too small to properly represent the statistical properties of the parent distribution, and not due to the quality of any subsequent distribution fit. The quality of the fit should relate instead to the realization-to-realization uncertainty for samples drawn from the same stationary process. To summarize – the uncertainty sources should be distinguished and their effect accounted for separately, or the sampling statistical uncertainty should be eliminated by drawing samples with the same wind statistics.

2) The extrapolation procedure analysed by the authors is not necessarily consistent with the actual design approach used by manufacturers. According to the IEC 61400-1 standard, it is acceptable to define the long-term distribution of loads according to two approaches:
1) first carry out extrapolations for simulations binned according to wind speed, and then aggregate the long-term load distribution based on the extrapolated functions (so-called "extrapolate, then aggregate" approach).
2) "First aggregate, then fit on the aggregated distribution" – i.e. the approach used by the authors where realizations are drawn directly from the long-term wind distribution.
In practice, approach 1) is much more common because it requires fewer simulations (approximately 300-400 10-minute periods), it allows using the same simulations used in calculating the fatigue limit state (DLC1.2), and the simulation sample is not conditional on the wind speed distribution. In fact, approach 2) is not even included in the IEC61400-1, ed.3 version cited by the authors – it first appears in the amendment to the standard published in 2010.

3) There is no discussion on the effect of the variation of the wind speed in the original MC sample, and in the random subsets used to test the uncertainty in extrapolation techniques. This variation means that in each sample there are a lot more simulations at wind speeds around the mean, but only few at high wind speeds – e.g., for a sample of size 1000, only 13 simulations on average will represent wind speeds above 20m/s.

4) What about other load cases? Very often, if the design loads from one load case (e.g. DLC1.1) are low, another load case (e.g. emergency shutdown with gust, DLC4.2) will become design-driving. So a lower design load prediction in DLC1.1 will not necessarily lead to lower material thickness, it will simply eliminate the load case as a potential design driver.

5) Page 10, lines 16-20, the authors conclude: "In any case, we can conclude that the bare minimum of 300 minutes of time series, as prescribed in Appendix F of the standards (IEC, 2005), is not sufficient to produce any reasonable 50-year estimate. Based on this load set and this extrapolation produce, one should instead aim for sample sizes larger than N = 10^5." I do not agree with this conclusion. Due to the issues outlined in my comments above, some additional uncertainty is present in the extrapolation. If these uncertainties are eliminated, or if another extrapolation procedure (e.g. aggregation after fitting) is used, the accuracy of the extrapolation may improve and the required sample size will become smaller.

Specific comments:

6) Page 3, last paragraph: "the wind speeds belonging to the 10% highest loads points towards a region well above the rated wind speed (see Figure 3)". I don't think this statement agrees with Figure 3, where the mode of the distribution $f\left(\overline{U}\middle|\hat{F} \geq 0.9\right)$ is actually at wind speeds just below rated.

7) Page 4, line 5: Why is the GEV distribution a good candidate? In my experience, the GEV, or other 3-parameter representation as the 3-parameter Weibull, are good candidates for extrapolation in situations where only few data points are available and not all of them belong to the upper tail of the extremes distribution. However, if the amount of data and the threshold selection result in a data set which is predominantly from the upper tail which has a characteristic log-linear behaviour, I would think that a 2-parameter Gumbel distribution would provide more robust and accurate fit.

8) Page 5, line 5: "any extrapolated 50-year loads that are more than 50% higher than the "real" value are discarded and resampled": - This data censoring approach is quite crude. It also relies on the known true value which will not be known in practice. Given the large number of samples, it is possible to establish a confidence interval for each load level and discard the outliers, see e.g. Naess and Gaidai, Structural Safety 31 (2009), pp. 325-334.

9) Figure 9: It seems that the uncertainty for sample size N = 100 is smaller than the uncertainty for N = 1000. This is against the logic and indicates a possible sampling bias. Could the authors find an explanation for this?

Technical comments:

10) Page 8, lines 5-10: It does not get entirely clear how the stress criteria for accepting or rejecting a given design relate to the extrapolated loads. Please improve the explanation.