Reply to comments of “Referee #1”, Lance Manuel:

First, the authors wish to thank Professor Manuel for his detailed and thoughtful comments. The reviewer’s comments point out some lack of precision in our discussion, especially that of IFORM, and will serve to improve the paper considerably. Point by point replies follow:

Overview and General Comments

This is a most interesting presentation of a vexing problem that has proven to be a challenge to wind turbine loads analysts for many years. The ideas developed by the authors and the narrative discussing the desire to “bridge” more conventional extrapolation methods and variance reduction techniques that go beyond brute-force Monte Carlo simulations are welcome. Casting the problem as an optimization problem, albeit without the usual formalisms, so as to adaptively improve estimates of long-term loads is done most effectively. Throughout, there are interesting insights and discussions that make for an illuminating reading and exposure to the essential issues.

Specific Section-by-Section Comments

Introduction

The description about IFORM on Page 2, Lines 12–13 should be clarified. More correctly, only with the environmental contour (EC) method which is the most commonly employed version of IFORM, one uncouples the “environment” from the “response,” and global extremes of interest associated with a target return period are approximated by using the maximum response from among all response levels derived only for candidate environmental variables consistent with that return period (in other words, response variability given environmental conditions is neglected). That said, in cases such as the one described in this article, where the environment is described using only one random variable (wind speed), the EC method has limited use. The EC method is better suited when a pair of random variables (say, wind speed and turbulence intensity, or wind speed and wave height for offshore turbines) are included. When only one random variable defines the environment, the “environmental contour” for a target return period is reduced to a single wind speed. This has limited use because, for instance, in the present example, for both side-to-side and fore-aft tower bending moments, the method will suggest that only rare and high wind speeds around cut-out that are associated with a 50-year return period need to be considered in turbine aeroelastic simulations. This will clearly lead to inaccurate 50-year fore-aft tower moments.

The applicability of EC does not depend on the *number* of random inputs as much as their joint probability. Even with a large number of random inputs, we can envision response variability being the governing influence, i.e. a situation where the extreme load occurs at rather common environmental conditions. This issue is just especially manifest in the 1-variable case, where indeed the 50-year EC is just a single point.

Now, despite the preceding comment that the EC method reduces the conventional environmental contour to a degenerate point or single-valued wind speed to consider for turbine response simulations, in fact, in the present case, one could instead use IFORM in its more general form and use wind speed and response as two random variables and formally derive estimates of 50-year side-to-side and fore-aft tower bending moments. This is discussed later, along with comments offered in the context of Section 2.4 (IFORM).

This is true, however, such estimates reintroduces extrapolation, which ASIS is designed to avoid (see below).

Extrapolation

Page 4, Line 14: Strictly speaking, the 50-year return period event is that event that is exceeded "on average" once in 50 years. Even though it is not the same in general, sometimes the event is defined as one that is exceeded "on average" with a probability of 1/50 in one year.

Thank you for this clarification; we should base our definition on the underlying assumption that this is a Poisson process and define our terms precisely from it.

Page 4, Line 26: 3.8^{-7} *should be* 3.8×10^{-7}.

Thank you, will be corrected.

Monte Carlo importance sampling for extreme loads

Page 6, Line 23: The comment that some form of accept-reject sampling can be used with importance sampling is an intriguing one. It is unclear how exactly this would be done given that $Y(x)$ *is not known in closed form; any additional notes, even if included very briefly, regarding such sampling would help.*

The comment is meant to address the problem that the normalization constant is not known, not that Y(x) is an expensive function. As long as we can *evaluate* Y(x), even if by simulation (we assume we can also evaluate f(x)), we can sample from any distribution proportional to Y(x)f(x) by the accept-reject algorithm (see, e.g. https://en.wikipedia.org/wiki/Rejection_sampling), which involves sampling uniformly in a 2D region containing the function Y(x)*f(x). The probability of x w.r.t. the Y*f distribution is just the proportion of these uniform samples below Y*f in this 2D "box". This procedure does require assumptions on the bounds of Y and the support of Y and f, but in principle these can be made large enough to "cover" any meaningful probability for Y*f.

IFORM

As stated earlier, the EC method doesn't apply here as there is no environmental contour corresponding to the authors' example—such a contour is a degenerate single wind speed value obtained as $F^{-1}(3.8\times10^{-7})$ *and as such has limited value for, say, the fore-aft tower bending moment where the derived 50-year load will certainly be under-predicted. Indeed, in this single environmental random variable case, the degenerate single simulation needed for any response or load of interest would require simulations to be run for a single wind speed above cut-out, i.e., for V equal to 43.3 m/s. This would be meaningless.*

The authors correctly point to the deficiencies of IFORM (on Page 7) but, in light of comments in the preceding paragraph, since there is no environmental contour at all that can be defined to describe their example study, much of the extended discussion regarding the EC method and environmental contours as presented in Section 2.4 is not relevant.

The discussion of EC is in part simply for reasons of completeness, but more importantly it is for conceptual aid. For us to reach low probability events *without extrapolation* we need to correlate the extreme events directly to environmental conditions. Otherwise, we must *model* the response variability, which puts us back in the modeling and extrapolation context we are seeking to overcome.

Now, in a most interesting way, the very issue that doesn't allow for a critique of the EC method—namely, that the authors choose only V as an environmental random variable—actually allows the more general IFORM procedure to be used with the authors' own simulation results and will lead to reasonable results (how this can be achieved is presented here very briefly). The idea is as follows: Consider that there are two random variables—wind speed, V, and the response or load, Y , whose statistics are derived from 10-min simulations . We will assume that we know the probability distribution for V (for instance, here, the authors use a Weibull V with shape and scale parameters equal to 2 and 11.28 m/s, respectively); we establish conditional distributions for Y given V based on simulations. We can use IFORM, though not the EC method, to find the required quantiles of Y |V for any V of interest. This is purely a geometry problem (involving mapping of V and Y to two independent standard normal random variables). To illustrate this, because the results presented in Figure 1 are the easiest to read off and learn from without great effort, one would find using IFORM that for TwrBsMxt, the 24 m/s bin would require that the desired 50-year response must have a probability of exceedance in

10 minutes of 5.95×10^{-6}. Given the 5th, 25th, 50th, 75th, and 95th percentile loads in Figure 1, a 2-parameter Weibull fit to these data leads to a 50-year TwrBsMxt value of 29,700 kN-m. Other (lower) wind speeds occur more often and associated load levels to be checked for those wind speeds using the IFORM procedure must be rarer, i.e., with exceedance probabilities in 10 minutes that are smaller than 5.95×10^{-6}. Given the data, these wind speeds do not lead to TwrBsMxt values at the desired probability levels that exceed what was found for V equal to 24 m/s. In a similar manner, for TwrBsMyt, selecting V equal to 16 m/s, the desired 50-year response for IFORM must have a probability of exceedance in 10 minutes of 7.20×10^{-7}. Again, from the data in Figure 1, a 2-parameter Weibull fit to these data read off easily, leads to a 50-year TwrBsMyt value of 94,300 kN-m. Again, note that other wind speeds and associated (different) response quantiles need to be checked, as part of the IFORM procedure, to ensure that the largest load quantile across all the wind speed bins is then claimed as the 50-year load. Details regarding all the calculations are not presented here but IFORM computations are based on the Weibull V and loads data from Figure 1.

This is wonderful! You are the most inspired (and inspiring reviewer) ever. We have no objection with your procedure, and it is certainly interesting that your results largely agree with ours. We would point out, however, that in carrying out the IFORM procedure described above, you have taken our data and *fit* it to a Weibull distribution. The resulting low probability estimations are then made possible by *extrapolation* of this fitted model. It would be interesting to investigate how this differs from simply fitting an extreme value distribution to the empirical data directly (as in the traditional bin-based IEC-recommended method). It is quite possible that the IFORM, even though still based on fitting and extrapolation, is fundamentally more accurate because in some sense it “factors out” the environmental probabilities. If some form of extrapolation is inevitable to get to 50-year loads in a tractable amount of computing time, maybe the combination of ASIS’s variance-minimization-sampling and general IFORM (as you describe) is a promising approach.

In sum, the authors’ comment regarding searching on the environmental contour (or just inside it) is not pertinent here. There is also no need to discuss above-median response levels in this context. Both the preceding comments would have been appropriate if, in addition to wind speed, another environmental random variable were included such as turbulence intensity. As illustrated, IFORM in its general form can be easily employed here and response variability can be directly accounted for—as the authors state correctly, this variability is ignored by the EC method. It is not ignored by IFORM in general and, as shown in the previous paragraph, even with the limited data presented in Figure 1, the method works quite efficiently in deriving 50-year loads. By running additional simulations at critical wind speeds, the resulting loads data and subsequent distribution fits to the same will lead to reduced uncertainty in derived 50-year loads.

Again, this is perhaps an important “intermediate step” between the traditional bin-based extrapolation method and the fully model-free ASIS method, i.e. more accurate than direct extrapolation but more computationally tractable than ASIS. But we feel it is important to point out that it does still indeed rely on fitting and extrapolation to achieve the desired 50-year return periods. Filling this gap precisely would be a very interesting area for future study.

Adaptive stratified importance sampling (ASIS)

Reference to POE, on Page 8, Line 20, should really be to the probability of non-exceedance.

Indeed correct; we can easily correct this.

In Step 2 of the algorithm on Page 10, why not simply obtain new samples (with rounding) in proportion to $\Delta_N J(N)$? It is not clear but it seems that the most important bin (where $\Delta_N J(N)$ is largest) is allocated some samples and the remaining (of 20) are then randomly allocated to other bins. Why randomly?

Randomly because we cannot strictly rely on information gained from the limited number of samples gathered so far. Due to random variation, the first N samples might *not* be leading us toward the correct bins, so "following" them could lead to a *local* solution to the minimal variance problem. Choosing the rest of the samples randomly is a crude (but common) strategy in global optimization.

The algorithm, as presented, appears not to state what is the criterion for stopping or convergence. The discussion regarding the "umbrella" concept that suggests a minimal superset of sample distributions across bins is exactly what is needed. It is the only way to guarantee adequate samples of response extremes to meet very distinct response characteristics such as between TwrBsMxt and TxrBsMyt. The ASIS algorithm as presented doesn't explicitly state this but assumes convergence when all the response measures are adequately sampled in all bins so as to yield unbiased 50-year response values, presumably with some specified confidence level on these predictions. If it helps, in an offshore wind turbine application, Sultania and Manuel [3], employed bootstrap-based confidence intervals for specific sea states (akin to bins here) to arrive at the appropriate number of simulations for accuracy in response probability distributions, conditional on the environment. Given the results presented in this article, it appears that convergence on long-term loads for each bin is indeed achieved by the authors in their examples by examining the variance of loads associated with low POE levels.

The lack of precise convergence criteria is indeed a shortcoming of the present paper, especially because in practice this is critical: the user is seeking an estimate of the 50-year load with some acceptable measure of its accuracy. We have pointed at the way one would achieve this in practice. In particular, we would recommend (as in the "Results" section) to uses ASIS iteratively in conjunction with extrapolation and bootstrapping: For each ASIS iteration, subsample via bootstrapping, form a large number of extrapolations, thus estimate the 50-year load *and its variance*. The stopping criteria is then a user-specified threshold for the variance.

It was not clear, upon reading, what was the reason for using the 5 largest loads. The largest loads will automatically drive the tail of long-term loads distributions and, as such, the 50-year load, when these largest loads are included along with all smaller loads; so, why retain only the 5 largest? Some clarification would help here.

This is indeed another part of the algorithm that would need further study before committing it to "production" use. Calculating the gradient of the variance using the 5 largest peaks was just an intuitive guess as to the number of loads that would drive the sampling in an effective way. A similar parameter is the number of peaks used for extrapolation M_{pks}, which we studied (at least graphically) in Figures 2, 4, and 5. Future work would be warranted to tune these parameters more systematically; hopefully it is convincing that the exact values of them does not undermine the principle of the method.

Results

The results in Figures 3–5 are very interesting and suggest that the methodology proposed by the authors offers a robust and efficient means of deriving long-term turbine loads for design. The rapid reduction in variance in estimates of long-term loads by using the ASIS algorithm is convincing.

A few observations are offered. 1. It would be very useful and insightful to see how g(x) or the bin-wise importance sampling changes with iteration for the the two load measures, TwrBsMxt and TwrBsMyt, separately, and what the ultimate umbrella sampling ends up being, after convergence, or as the number of iterations changes—given the contrasting characteristics of the two load measures, one might expect a bi-modal sampling with one mode around or slightly above rated and another closer to cut-out will result.

In our earlier paper ("Advances in the Assessment of Wind Turbine Operating Extreme Loads via More Efficient Calculation Approaches" in: AIAA SciTech 2017) we have (see Figure 6 therein) plotted the distributions that ASIS led to. In that paper, the ASIS method was applied separately to each load, and they do result in bin distribution peaks at different wind speed values. So it would indeed be predicted that a bimodal distribution would emerge.

2. It is not completely clear what M_{pks} refers to. For instance, does $M_{pks} = 40$ imply that, *after accounting for samples from all bins, only the largest 40 are used or does it mean that for each bin, the largest 40 loads are retained and then combined with 40 from the other bins before extrapolation?*

The M_{pks} largest loads from each bin were used to separately fit the extreme value distributions in a bin-wise fashion.

3. The non-monotonic upward trending cooefficient of variation on the 0.05 POE load seen in Figure 3 for a very large number of peaks might be caused by dependence among the peaks that results when too many peaks are extracted from each 10-min simulation (see Fogle et al [2]).

Indeed, this is likely, especially because the effect is most pronounced for the lowest iterations where there are fewer total runs to choose from.

4. Is it possible that the x axis units in Figure 4 are 10,000's of KN-m. rather than 1,000's of kN-m as stated? This would be consistent with Figures 1 and 2 which appear to show TwrBsMxt loads an order of magnitude higher. It would also be consistent with the IFORM-based estimates that were computed (above) as 29.7×10^3 kN-m and 94.3×10^3 kN-m, respectively, for TwrBsMxt and TwrBsMyt. The caption for Figure 5 is correct; the one for Figure 4 might need to be corrected.

Agreed. Again, thank you for reading carefully! Will correct.

5. As presented, Figures 4 and 5 do not show loads for the low POE level of 3.8×10^{-7} associated with the target 50-year return period. It would have been useful to see those results. Indeed, ASIS-based convergence at the lowest POE levels that are presented, suggest that 25 iterations and the use of 40 peaks is very good.

It would of course be possible to extrapolate to the 50-year levels in practice. Though ASIS appears to work well, we see there is no free lunch; it would still require many iterations to achieve empirical 50-year estimates.

In the caption for Figure 3, 5^{-2} should be 5×10^{-2}.

Thank you for point that out; we will correct this.

Conclusions

The closing discussion is most helpful in setting this work in the context of other studies regarding the derivation of long-term loads for wind turbine design. The ASIS algorithm, as presented, will prove useful to loads analysts. While the authors have demonstrated its effectiveness using 100 independent tests in their numerical studies, in practice, it may be useful to suggest use of the ASIS algorithm followed by bootstrapping, after each *iteration, and confirmation that the coefficient of variation on some load quantile (as in Figure 3, for example) is acceptably low—for example, 5%. It is clear that ASIS can achieve this target with a far smaller number of aeroelastic simulations than extrapolation based on ordinary or conventional sampling methods. The authors do, in fact, recognize in their closing discussions, that bootstrapping with a single data set could be used with ASIS and thus reduce computational effort.*

Only to provide a contrast with the authors' work, it should be noted that IFORM-based approaches, referred to in this article, can be very efficient. This has been demonstrated in this discussion using the authors' own data (crudely derived from their Figure 1); note that in the illustration presented response variability is not ignored as is done with the environmental contour method. Accounting for response variability in IFORM is not difficult—it is conceivable that an ASIS-like formulation for sampling could prove far more efficient than even IFORM but this needs to be demonstrated for situations where more than one random variable defines the "environment."

It is possible, as noted above, that though IFORM in this context still involves extrapolation, that it is fundamentally more accurate than direct fitting to extreme value distributions, because the environmental contour has already been accounted for. Thus, an improvement on the recommend ASIS+extrapolation+bootstrap approach might instead be ASIS+IFORM+bootstrap. This would be a subject of future study.

Finally, related to this last point, the ASIS algorithm would be especially important to employ along with the introduction of stochastic turbulence and with the treatment of turbulence intensity or turbulence standard deviation explicitly as a random variable. The role of gusts and turbulence in extreme loads is known to be significant and the need then for bivariate importance sampling distributions could present challenges (Bos et al. [1], van Eijk et al. [4]); at the same time, the benefits to be derived from approaches such as ASIS, if its efficiency and accuracy is demonstrated in such cases, would then add to its appeal.

This would be another interesting area of future study/application. In principle it is a simple extension of the method to apply the stochastic variance minimization procedure to more than one variable.

On Page 15, Line 3, "stratified adaptive" should be "adaptive stratified" to be consistent with the ASIS acronym.

Yes.

Finally, thank you very much once again for the incredibly careful and thoughtful reading of the paper.

On behalf of the co-authors,

Sincerely,

Peter Graf,
Computational Science Center and National Wind Technology Center, National Renewable Energy Laboratory