Reply to comments of “Anonymous Referee #2”:

First, the authors wish to thank “Referee #2” for the detailed and thoughtful comments. The reviewer’s comments will serve to improve the paper considerably. Point by point replies follow:

General comments

The authors have presented a very interesting idea for efficiently improving the esti- mates of 50-year extreme loads for wind turbines. Most attractive about the method is its overall simplicity, using the well-known concept of importance sampling and deriving a simple gradient for the variance of the extrapolation estimate that can be used to add the required number of additional samples for each wind bin, in order to reduce the variance of the estimate to an acceptable level. With a few exceptions, to be noted below, the presentation of both background details and the method is clearly written. The results seem to show that the method works quite well, without having to run the optimization for a prohibitively long time. However, there is one aspect of the setup of the problem that obscures the true meaning of the results, again to be expanded on below. Unless this can be clarified, interpretation of the results is difficult. As a preliminary demonstration of the approach, and assuming they hold up to scrutiny, these results are promising and the method could be of use both in conventional assessments of the 50-year extreme load and also in other settings like reliability assessments. It is also nice to see that the authors are open and critical about the overall power of the method and they make some interesting points about the limits of the current analysis framework and suggestions for other ways to improve extreme value analysis for wind turbines.

Specific comments

Introduction:

On page 2, lines 14-15 you write: “... reliable extrapolation of nonlinear physics under uncertain forcing is extremely problematic, especially without knowledge of the form (e.g. quadratic) of the nonlinearity." This is certainly true, but an equally important reason why the specific type of long-term extrapolation usually done for the evaluation of 50-year extreme loads is problematic is precisely the large differences in timescale between the data and the extrapolated estimate. This impacts the problem in many ways, certainly also through the nonlinearity you mention, but in a more practical sense this large difference in timescales means that any uncertainty in the data is necessarily magnified by the extrapolation. Small errors in the short-term data set could potentially lead a designer to significantly over- or underestimate the long-term extreme loads. Later in the paper you show that there can be a large variance in the extrapolation, which is in turn reduced by your proposed method, so this overall point should be mentioned here.

This is an excellent point about the difficulties due to time scales and we can certainly add some language to that effect to the problem introduction.

Section 2.2, Extrapolation:

On page 4, line 2 you write: "In this paper we use a 3 parameter Wiebull distribution." Why this distribution? This choice may for the demonstration of the method be seemingly irrelevant, but some reason for the choice would be instructive for the reader. One might wonder, for example, if this is indeed the distribution that overall gives the best fit for your data and is therefore the easiest to use for illustrating the method. Certainly, the Gumbel distribution, for instance, can be easier to work with (since it has only 2 parameters), so it seems there must be some motivation for choosing the Weibull. For reasons of clarity and reproducibility, it would also be of interest to know what method you used to fit the distributions to your data. Maximum likelihood estimation perhaps? Please state this.

We agree there is a lack of a full explanation here. We used the 3-parameter Weibull because it has worked well in previous work. There are many excellent papers concerning extrapolation and efforts to justify and/or distinguish between different extreme value distributions (e.g., Ragan and Manuel,"Statistical Extrapolation Methods for Estimating Wind Turbine Extreme Loads", Toft, Sorenson, and Veldkamp, "Assessment of Load Extrapolation Methods for Wind Turbines"). It is not the purpose of the current paper to assess the different distributions.

Regarding the fitting procedure, in light of the interest in particular in extrapolation, rather than just fitting, we have fit the empirical cumulative distribution function of the data directly to the theoretical cumulative distribution function (CDF) of the distribution by nonlinear least squares. We have done this separately for the data from each wind speed bin. In order to emphasize the largest peaks (i.e. the lowest probability values) we do not use all the data, just the M_{pks} largest peaks, where M_{pks} is an algorithmic parameter. As an exercise, we experimented with using different values of M_{pks} (e.g., see Fig 2).

On page 4, line 29 you write: "To gather peaks, we take the maximum of each 1 minute segment in our simulations." Yet, you already stated on lines 17-20: "Finally, there is the length of time between independent peaks ... values as low as 4 seconds have been justified in previous studies (Ragan and Manuel, 2008), and 10 seconds seems to be more than 20 adequate." So why this choice of 1 minute separations? Based on what you write later in the paper it seems to be motivated by a desire to only use the largest peaks, hence using smaller separations would yield maxima that might not give a good description of the extreme behavior of the system. If this is the case, or if there is some other motivation, it would be instructive to have it stated clearly here. To be very precise, given that one wants a number of peaks that exactly divides the total simulation length, why not a 30 second separation, which would give twice the number of maxima and hence more data per simulation with which to fit the distributions while maintaining 3 times the required separation to maintain independence of the peaks?

The tradeoff between gathering more peaks (at the risk of sacrificing statistical independence) versus less peaks (at the risk of not having enough data) is another interesting detail that we could study. In this vein, our use of 1 minute intervals and your proposal to use 30 second separations has a similar motivation, which is simply to pick a reasonable point along that tradeoff that avoids the pitfalls of either end point.

Section 3.2, ASIS as stochastic optimization:

As a more general point, it is clear from the fact that the variance is what is being minimized, as *well as studying equation (21) and from the algorithm summary on page 10, that the procedure is only ever going to add samples to the various bins, never remove samples. However, viewed as a more general optimization problem, an unattentive reader could believe that such an algorithm might in fact reduce the number of samples in a bin. Perhaps this "uni-directionality" of the algorithm should be stated more clearly to avoid any possible confusion over what its purpose is.*

Agreed. In principle we can imagine "solving" the bin-optimization problem once and for all, which would give us an optimal distribution of bins. Then, given a certain computational budget, we could apportion the samples to bins proportionally, which would indeed possibly reduce the number in some bin from the original search. Our orientation was more from an "on-line" perspective: why not use data from bins you have already run simulations on? It would be easy to add a sentence clarifying this point.

On page 10, lines 11-12 you write: "Also, the N_i need to be integers, which is accomplished by rounding. The resulting error is likely subsumed into the general convergence of the stochastic optimization procedure." This statement needs a more convincing justification. For a k-dimensional optimization problem, which for a discrete solution set induces a k-dimensional lattice of discrete points, it is not immediately clear that a local minimum in the continuous case is also a local minimum in the discrete case. That is, when going from some continuous set of N_i (a point in k-space) to a discrete set of N_i (a point on the k-lattice) by rounding each N_i to the nearest integer, the corresponding function value is not necessarily lower than at neighboring points on the lattice. If such a correspondence between minima in continuous and discrete space can be established for this particular case, it needs to be justified by specific arguments or at least by reference to another work.

We have no such formal justification for this procedure. The thinking is simply that we assume the variance of the estimate as a function of bin distribution is a smooth enough function that it is a reasonable *approximation* to round to the nearest bin counts. Given that we do have to round to integer bin counts, this is the best we can do. Furthermore, this is a stochastic problem, so it is likely that the continuous-valued bin counts have some error associated with them anyway. We claim (without proof, but believe it is a reasonable assertion) that the error induced by the unconverged sampling procedure is just as much or more than that induced by the rounding to integers of the bin counts. We are certainly not claiming that the rounded minimum is lower than the "relaxed" (i.e. continuous-valued) minimum; this is patently false, because the possible discrete bin-counts are a subset of the continuous ones.

On page 10, line 18 you write: "In fact we have to decide some number of "large peaks" we will use to evolve N." Why? Presumably you already have all the information from the extrapolation that has already been performed before estimating the gradient. So why must the gradient only be estimated from some limited number of large peaks? The motivation for this is certainly not clear from the text.

This is, operationally, a separate step from the extrapolation (in fact, the ASIS variance minimization can be performed without ever performing extrapolation). In retrospect, there may be an interesting connection between the number of peaks used for extrapolation (M_{pks}) and the number of peaks that enter into the gradient-of-variance calculation, but this is beyond the scope of the current paper (though if you are confused, we will clarify this distinction in the text). The choice of the "5 largest peaks" is rather arbitrary, enough so that more than the single largest peak contributes, but not so many as to deemphasize the goal of finding *large* peaks.

Section 4, Results:

To demonstrate the method, you set up a problem where you seek the ideal number of peaks, M_{pks}, to use for the fitting. However, it is unclear how this variable relates to the extrapolation procedure already described. Initially, you have used 10 maxima, 1 per minute of simulation time, for each wind speed bin. Is it this number of 10 maxima which is now a variable? If so, how is this actually accommodated in the extrapolation? For example, in Figure 2 you show the extrapolation for many different values of M_{pks}. How do you get the additional peaks? Do you perform another T minutes of simulation corresponding to the number of additional peaks needed? Or do you decrease the separation time between maxima and hence extract more maxima from the simulation data you already have? If the former, do you perform specifically the required amount of simulation time for each case or do you perform enough simulations for all the different values of M_{pks} you have used and then simply use the 5, 10, 20, 40 etc largest peaks from this expanded set of potential maxima? These different ways of solving the problem have very different statistics and therefore different implications for the extrapolation itself. Using the largest 20 maxima from 160 minutes of simulation time is very different from using 20 maxima from 20 minutes, which is very different from using the 20 largest maxima from 10 minutes and so on. In fact, depending on which of these approaches is used, it is not clear whether the results are truly meaningful. Nor is it clear how the number of maxima used might or might not interact with the number of samples in each bin as dictated by the ASIS optimization procedure.

Clearly the description of the choice of peaks to use for fitting is inadequate. For Figure 2, for example, we have run FAST for 10 minutes 20 separate times for each bin (the figure shows only the results for the 20 m/s bin, but the others are similar). This results in 200 minutes of total simulation time, thus according to our 1-peak-per-minute convention (which is fixed throughout the paper), 2000 peaks. Each line on the figure (e.g. "5 peaks", "10 peaks", etc.) is the result of fitting the 3 parameter Weibull CDF to just the largest M_{pks} (e.g. 5, 10, etc) of these 2000. The line labelled "empiricial CDF" is the empirical CDF of the 2000 peaks. We were hoping to *avoid* having to argue about the distinctions you mention above by choosing very "vanilla", uncontroversial values for these types of parameters, but our lack of clarity has unfortunately brought them all back into play. We apologize and will clarify in the next draft. We do not think the choices we made effected the meaningfulness of the results. The number of maxima used is similar to the choice of threshold in peak-over-threshold method, or the simulation time in methods just taking the single maximum from each simulation. In this context it is well-studied in the literature, but we agree that its role in ASIS is a new twist. Figures 3, 4,and 5 suggest that it is not a completely critical issue, however. Quite a range of values for M_{pks} are effective, and the effectiveness (e.g., see Fig 3) is a relatively (for stochastic optimization) smooth function of M_{pks}.

More details about the optimization are needed in order to ensure that the results are clear and reproducible. For example, the criteria for termination of the algorithm are unclear.

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. In short, monitor the (left and middle for ASIS and ASIS+extrapolation, respectively) curves in Figure 2 until the variance is below the desired level.

Technical corrections

- *Section 2.2, page 4, line 2: "In this paper we use a 3 parameter Wiebull distribution." Should be "Weibull"*
- *Section 3.1, page 8, equation (12): $Y(x)$ should be $Y(x_i)$*
- *Section 3.1, page 8, lines 14-15: "Here we denote the dataset as $\{Y_{i,k}\}$ where i indexes over wind speed bins and k indexes over the peaks we have extracted at that wind speed." This repeats almost exactly information already given in Section 2.2, page 3, lines 30-31. Consider removing (since we are already aware of the notation) or rephrasing.*

Thank you for these.

And thank you again for your careful reading and consideration of the paper.

On behalf of the co-authors,

Sincerely,

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