The risks of extreme load extrapolation

Stefan F. van Eijk[1], René Bos[1], and Wim A. A. M. Bierbooms[1]

[1]Wind Energy Research Group, Faculty of Aerospace Engineering, Delft University of Technology, 2629 HS Delft, The Netherlands

Correspondence to: Stefan van Eijk (sfvaneijk@gmail.com)

Abstract. An important problem in wind turbine design is the prediction of the 50-year load, as set by the IEC 61400-1 Design Load Case 1.1. In most cases, designers work with limited simulation budgets and are forced into using extrapolation schemes to obtain the required return level. That this is no easy task is proven by the many studies dedicated to finding the best distribution and fitting method to capture the extreme load behavior as well as possible. However, the issue that is often
overlooked is the effect that the sheer uncertainty around the 50-year load has on a design process. In this paper, we use a collection of 96 years' worth of extreme loads to perform a large number of hypothetical design problems. The results show that, even with sample sizes exceeding $N = 10^3$ ten-minute extremes, designs are often falsely rejected or falsely accepted based on an over- or underpredicted 50-year load. Therefore, designers are advised to be critical of the outcome of DLC 1.1 and should be prepared to invest in large sample sizes.

1 Introduction

Wind turbine designers are confronted with the IEC 61400-1 Design Load Case 1.1 (IEC, 2005). This evaluates the structural integrity of the major load-carrying components on the basis of a 50-year return level, plus safety factors. As prescribed by Appendix F of the standards, a minimum of 300 minutes of time series—distributed over the relevant wind speeds—will have to be evaluated and followed by an extrapolation scheme to obtain the 50-year return level. Such extrapolations produce notoriously uncertain estimates, which is why the Design Load Case (DLC) 1.1 is often avoided or at least greatly simplified in early stages of the design. However, in cases where DLC 1.1 is design-driving (e.g., for foundations and controllers), dealing with this uncertainty is unavoidable.

Many past efforts to reduce this uncertainty have focused on trying out different sampling methods (Fogle et al., 2008; Agarwal and Manuel, 2009), new modeling techniques (Moriarty et al., 2004; Bos and Veldkamp, 2016), or finding the best distribution type to match the extreme load behavior (Pandey and Sutherland, 2003; Genz et al., 2006; Freudenreich and Argyriadis, 2007; Ragan and Manuel, 2007; Natarajan and Holley, 2008; Peeringa, 2009; Lott and Cheng, 2016). Yet, because most studies deal with relatively small sample sizes (e.g., $\ll 1$ year), the actual uncertainty that surrounds the 50-year return level is often underexposed. With a 63-year data set, Barone et al. (2012b) were able to establish the 90% confidence interval around the 50-year load for sample sizes up to $N = 2,000$ ten-minute maxima. This revealed not only that the 50-year levels are clouded by high uncertainty, but also that they suffer from a considerable bias. Inevitably, this has an effect on the choices made during design.

The aim of this paper is to demonstrate this with a simple exercise, using a collection of 96 years' worth of ten-minute load maxima released by Barone et al. (2012a). The uncertainty distribution is constructed by repeatedly sampling subsets of this data set and obtaining the 50-year loads through an automated extrapolation scheme. We then simulate a problem where a hypothetical designer has to choose between two or more concepts and record how often this uncertainty leads to wrong

5 choices. The results of this paper should help designers to estimate the required sample sizes for their problem, but also to form a critical attitude concerning the quality and reliability of extrapolated 50-year loads.

2 Methodology

Since the focus of this work is on the impact of uncertainty, rather than obtaining the highest possible quality result, the workflow is kept as simple as possible. Loads were extracted by drawing a random sample from a large set of crude Monte

10 Carlo results and the 50-year return period is found by a graphical fit.

2.1 Loads data set

The data set that was used for this study was generated by Barone et al. (2012a). It features the onshore version of the NREL 5 MW reference wind turbine, operating for 96 years in an IEC class 1B climate (IEC, 2005).[1] Ten-minute mean wind speeds were randomly drawn from a Rayleigh distribution, bounded by the cut-in and cut-out wind speeds of 3 and 25 m/s, respectively.

15 Turbulent wind fields were generated by TurbSim on a 20×20 grid with a width and height of 137 m and were fed to the FAST v7 aeroelastic code. Every simulation ran for eleven minutes, of which the first minute was discarded to avoid any start-up transients. More details can be found in the original paper.

Each output channel contains over 5 million ten-minute extremes. In this paper, we will use the tower base overturning moment, which plays a major role in the design of foundations. Figure 1 shows the entire set of loads at the respective wind

20 speeds.

2.2 Extreme load distribution

The wind speeds were drawn directly from the Rayleigh distribution. Therefore, the cumulative distribution of extreme loads follows naturally from ranking a set of N loads and assigning a plotting position:

$$\hat{F}(M_i) = \frac{i}{N+1}. \tag{1}$$

25 In this case, however, the wind speeds outside of the operating range will have to be accounted for:

$$\hat{F}(M_i) = 1 - \left(1 - \frac{i}{N+1}\right) \int\limits_{\text{cut-in}}^{\text{cut-out}} f(\bar{U}) \, \mathrm{d}\bar{U}, \tag{2}$$

where $f(\bar{U})$ is the mean wind speed distribution. Then, plotting the entire data set yields the return level plot shown in Figure 2. With 96 years worth of load data, the 50-year return value can be interpolated, which yields 115 MN·m.

---

[1]The original paper specifies a class 2B site, but this has been corrected with the release of the data set (see http://energy.sandia.gov/?page_id=13173).
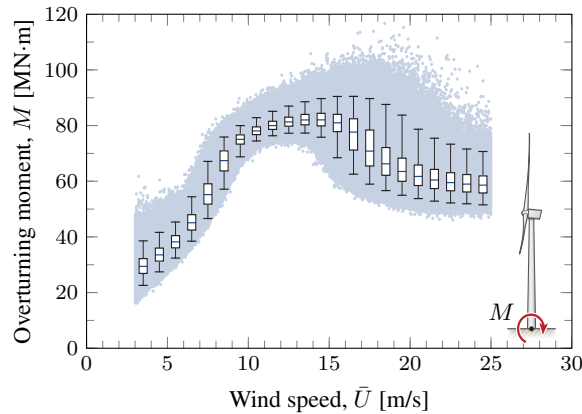
Figure 1. The data set, containing over 5 million ten-minute extreme overturning moments between the cut-in and cut-out wind speeds. The box plots indicate the scatter per 1-m/s bin, where the boxes mark the 25th and 75th percentiles, the whiskers mark the 2.5th and 97.5th percentiles, and the bar is the median.
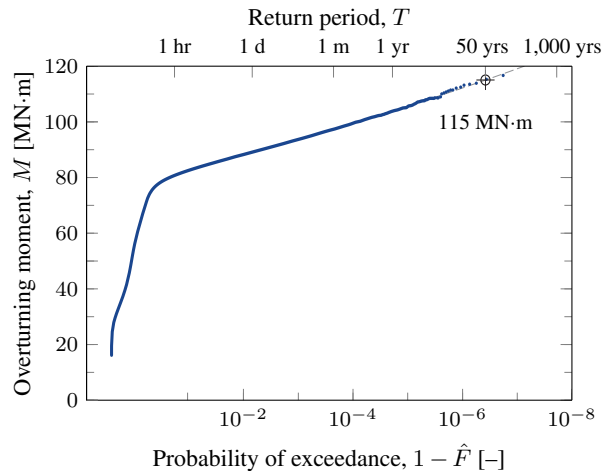
Figure 2. Return level plot of the tower base overturning moment, with a 96-year sample size, yielding a 50-year value of 115 MN·m.

The tail of the distribution shows a characteristic bend, or "knee", that hints that more than one process is at work. Indeed, tracing back the wind speeds belonging to the 10% highest loads points towards a region well above the rated wind speed (see Figure 3). It turns out that this is due to a particular controller response to negative gust amplitudes (e.g., Bos et al., 2015; Bos and Veldkamp, 2016), which also explains the shape of the scatter plot in Figure 1.
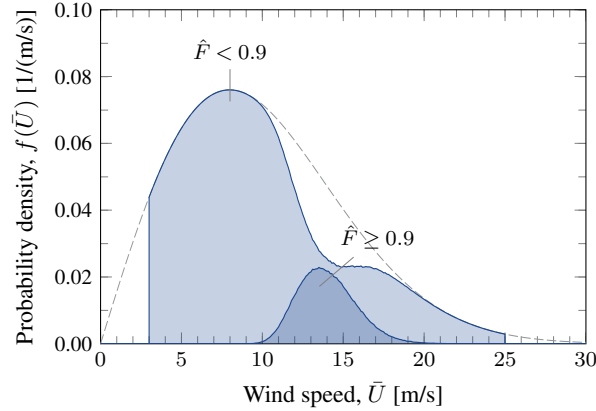
Figure 3. Histogram of the sampled wind speeds, where the dashed line marks the Rayleigh mean wind speed distribution belonging to an IEC class 1B climate. The light and dark filled areas correspond to the lowest 90% and highest 10% loads, respectively.

2.3 Extrapolation scheme

In many practical situations, a designer does not have the computational resources available to simulate several decades of operation. That is when the 50-year load has to be found by extrapolating. The objective is then to match the tail behavior with a distribution function, for example by least-squares fitting. A good candidate for this is the generalized extreme value distribution (GEV):

$$G(M; \mu, \sigma, \xi) = \exp\left[-\left(1 + \xi \frac{M - \mu}{\sigma}\right)^{-1/\xi}\right], \tag{3}$$

where μ is the location parameter, σ the scale parameter, and ξ the shape parameter.

To estimate the uncertainty that comes from repeatedly extrapolating different sets of loads, this process has to be automated. However, the difficult part is then to decide where the tail exactly starts under varying sample size. A simple solution that seems to work in most cases is to assume that the tail covers the second half of the distribution when drawn on Gumbel paper; i.e.,

$$-\ln\left[-\ln\left[\hat{F}\right]\right] > -\tfrac{1}{2}\ln\left[-\ln\left[\hat{F}(M_1)\right]\right] - \tfrac{1}{2}\ln\left[-\ln\left[\hat{F}(M_N)\right]\right]. \tag{4}$$

For the full data set, this means that the GEV distribution is fitted to the upper 0.07% of the data, which results in the Q-Q plot shown in Figure 4.
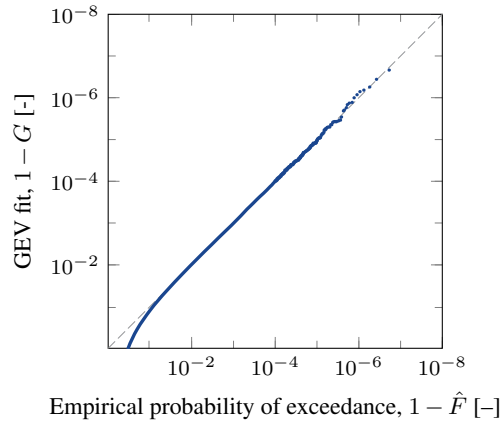
Figure 4. Q-Q plot, showing how well the empirical tail behavior matches with an generalized extreme value distribution ($\mu = 78.2$ MN·m, $\sigma = 2.05$ MN·m, $\xi = 0.026$).

This procedure is automated for $k = 10^4$ sets of loads, which yields a collection of 50-year return levels, distributed according to $f(\hat{M}_{50\,\text{yrs}})$:

$$
\begin{array}{ccccccccc}
 & & & rank \rightarrow & & & & & \\
M_{1,1} & \leq & M_{2,1} & \leq & \ldots & \leq & M_{N,1} & \rightarrow & \hat{M}_{50\,\text{yrs},1} \\
M_{1,2} & \leq & M_{2,2} & \leq & \ldots & \leq & M_{N,2} & \rightarrow & \hat{M}_{50\,\text{yrs},2} \\
\vdots & & \vdots & & & & & & \vdots \\
M_{1,j} & \leq & M_{2,j} & \leq & \ldots & \leq & M_{N,j} & \rightarrow & \hat{M}_{50\,\text{yrs},j} \\
\vdots & & \vdots & & & & & & \vdots \\
M_{1,k} & \leq & M_{2,k} & \leq & \ldots & \leq & M_{N,k} & \rightarrow & \hat{M}_{50\,\text{yrs},k}
\end{array}
$$

$\downarrow set$

$$\downarrow$$

$$f(\hat{M}_{50\,\text{yrs}})$$

The medians and other quantiles are then estimated by sorting. In order to catch very bad fits, any extrapolated 50-year loads that are more than 50% higher than the "real" value are discarded and resampled (as they would by an experienced engineer). The end result is the situation depicted in Figure 5, which can be repeated for different sample sizes.
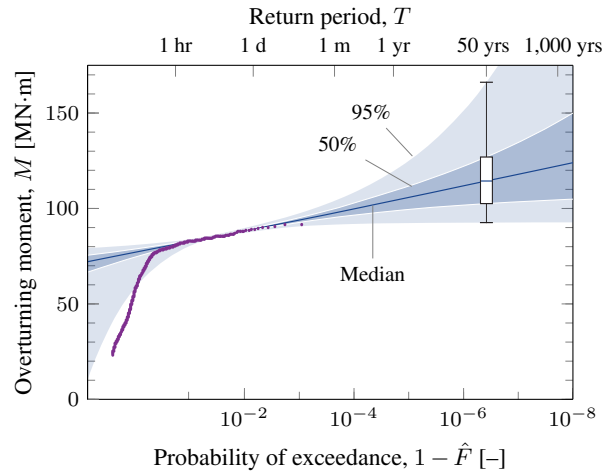
Figure 5. Return level plot of the tower base overturning moment, showing all the $k = 10^4$ distribution fits with a sample size of $N = 10^3$.

3 Results

Based on the load set and the extrapolation scheme, we can estimate how far a single 50-year load prediction would be off from the true value.

3.1 Uncertainty surrounding the 50-year overturning moment

5 First, Figure 6 shows how the median and confidence intervals around the 50-year level vary as a function of sample size, N. Evidently, the larger the sample size, the smaller the error. Overall, the GEV distribution has a tendency to overpredict the 50-year level, which is also found by fitting a straight line (i.e., a Gumbel distribution with $\xi = 0$) as done by Barone et al. (2012b). Even with $N = 10^4$—which is already more than two months of simulated time—the median prediction still lies about 5% above the ture 50-year level.

10 For sample sizes smaller than $N = 10^4$, it can be very hard to establish an appropriate tail behavior. The GEV fit will often return a 50-year level that is more than 50% higher than the true value, partially owing to large values of the shape parameter, ξ. In such cases, a simple straight line will produce better predictions.

In addition, the root-mean-squared (RMS) error provides a single measure for the quality of the result:

$$\varepsilon_{\mathrm{RMS}} = \sqrt{\frac{1}{k}\sum_{j=1}^{k}\left(\hat{M}_{50\,\mathrm{yrs},j} - M_{50\,\mathrm{yrs}}\right)^2}, \tag{5}$$

15 where $M_{50\,\mathrm{yrs}}$ is the "true" 50-year level. As shown in Figure 7, the RMS error stays fairly constant up until a sample size of $N = 10^4$ when predictions above +50% are discarded. It is not until sample sizes of close to 10^5 that the tail shape of the empirical extreme load distribution is reliable and most of the fitting problems disappear. Ultimately, the RMS error falls into the classic $1/\sqrt{N}$ rule that is often found with Monte Carlo methods.
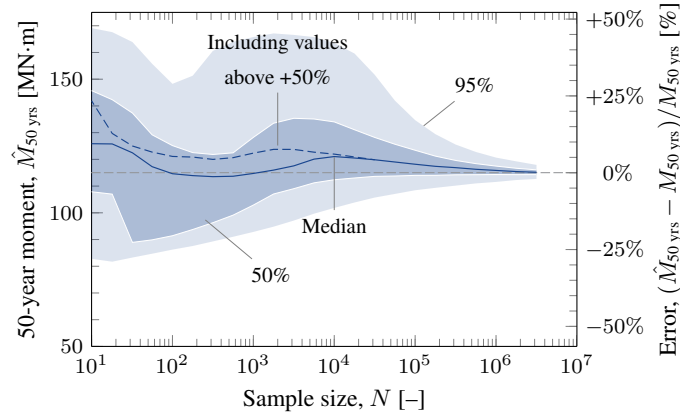
Figure 6. Error in the extrapolated 50-year overturning moment, as a function of sample size.
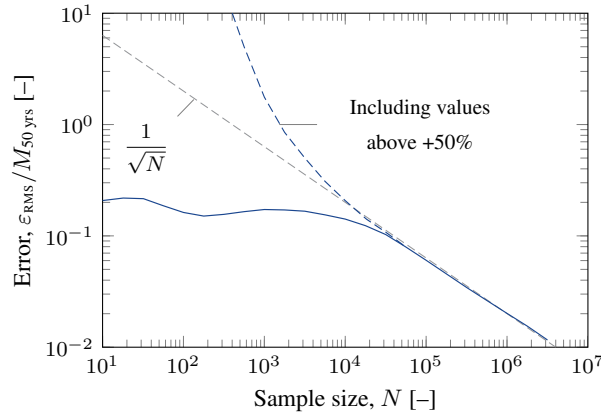
Figure 7. Root-mean-squared error in the extrapolated 50-year overturning moment, as a function of sample size.

3.2 Effect on conceptual design

How this uncertainty affects the design process is demonstrated here with a very simple example. A new concept is proposed that is an exact copy of the NREL 5 MW machine, but with a different wall thickness at the base of the tower. The second moment of area then changes according to

$$I_{yy} = \frac{\pi}{4} \left[r^4 - (r - t)^4 \right], \tag{6}$$

where $r = 3$ m is the base radius and $t = 35$ mm is the original wall thickness (Jonkman et al., 2009). An extreme overturning moment would cause a compressive stress of

$$\sigma_z = \frac{Mr}{I_{yy}} + \frac{mg}{A}, \tag{7}$$
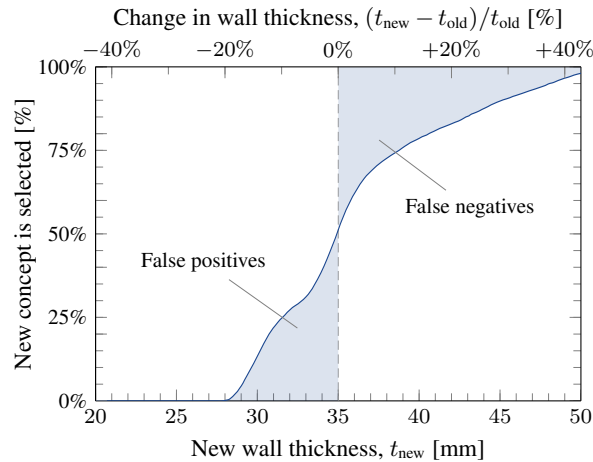
Figure 8. Outcome of a concept selection, where a new wall thickness if either accepted or rejected on the basis of a 50-year stress level with a sample size of $N = 10^3$.

where $mg = 6.82$ MN is the total weight of the wind turbine and

$$A = \pi \left[r^2 - (r - t)^2 \right], \tag{8}$$

is the cross-sectional area of the tower base section.

The objective of the exercise is to find a new wall thickness to reduce the 50-year stress levels; i.e.,

5 $\sigma_{z,\text{new}} < \sigma_{z,\text{old}}.$ (9)

This might seem trivial at this point—any thicker wall is guaranteed to reduce the stresses—but the actual difficulty is to determine the 50-year moment. Whereas the original design has already gone through an extensive load analysis from which the 50-year load level is known, any new concept has to go through this process again.[2]

Due to the uncertainty that surrounds this 50-year level, the new design can be falsely rejected or falsely accepted. Figure 8
10 shows how often this happens when the load analysis is carried out 10^4 times with a sample size of $N = 10^3$. When the wall thickness is reduced by 10% to 31.5 mm, the new design will appear to have lower stresses in 25% of the cases (i.e., the *false positives*). On the other hand, even when the wall thickness is increased by 20% to 42 mm, the new design has a 17% chance to still be rejected (i.e., the *false negatives*). The closer the new design is to the original, the harder it will be to make the right choice.

15 The positive bias that is often present in extreme load extrapolations (e.g., Barone et al., 2012b; Van Eijk, 2016) make it particularly difficult to prove that new designs are capable to reduce 50-year levels. With a crude Monte Carlo method, the only solution is to further increase the sample size, as shown in Figure 9. However, it will take an immense computational effort to completely remove the uncertainty from the design process.

---

[2]Maybe not for something like a new wall thickness, but more so for different control schemes or for rigorous changes to the blade design.
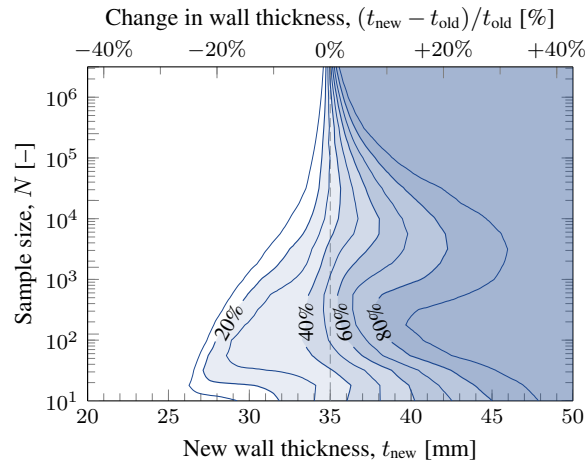
Figure 9. Relative number of times a new wall thickness is accepted on the basis of a 50-year stress level with different sample sizes.
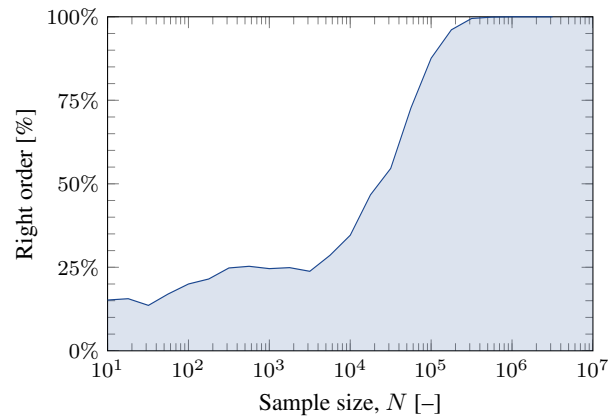
Figure 10. How often five concepts, ranging from 25- to 45-mm wall thickness, are ranked in the right order from lowest stress to highest.

Another case is a comparison between several concepts, where the 50-year stress levels contain the same degree of uncertainty. Five concepts, from 25- to 45-mm wall thickness, are ranked among each other, such that

$$\sigma_{z,1} \leq \sigma_{z,2} \leq \sigma_{z,3} \leq \sigma_{z,4} \leq \sigma_{z,5}. \tag{10}$$

In the ideal case, the 45-mm wall thickness should end up at rank 1, the 40-mm one at rank 2, etc. However, how often this ideal ranking happens in practice is shown in Figure 10. For sample sizes smaller than, say, $N = 5 \cdot 10^3$, the right order is only found in 25% or less of the cases. It is not until $N = 3 \cdot 10^5$ that the uncertainty is small enough for the order to be right roughly 100% of the time. From $N = 10^4$ to $N = 10^5$, the quality of the predictions increases sharply, which is when the tail of the distribution starts to take its proper shape.
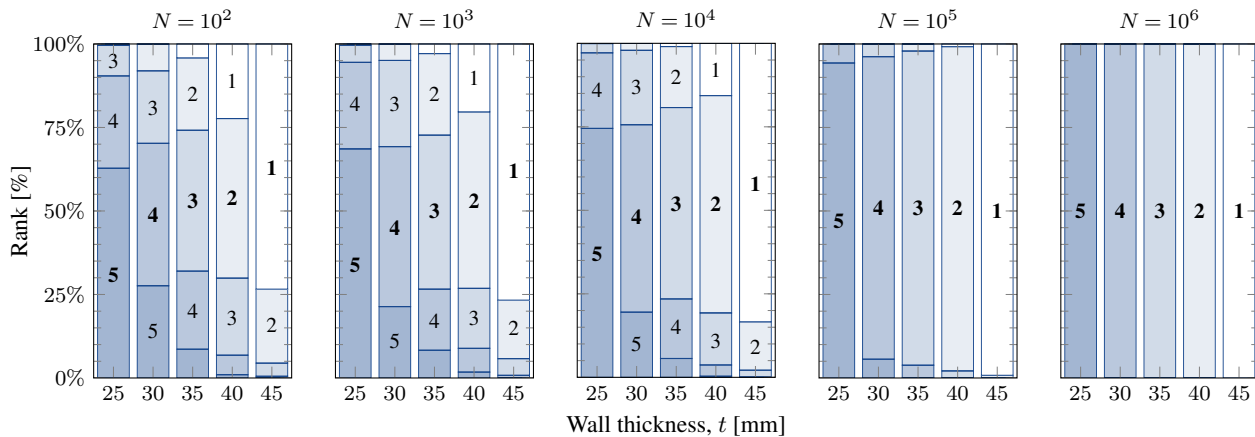
Figure 11. Ranking of five wall thicknesses on the basis of a 50-year stress level with different sample sizes.

How often each rank is assigned to each concept is shown in Figure 11. Clearly, the 45-mm wall thickness does not always appear the best and the 25-mm wall thickness does not always appear the worst. The closer the concepts are to each other, the harder it becomes to distinguish them by their 50-year stress levels. In fact, the chances of selecting a concept other than the thickest, 45-mm one are higher than 15% for $N \leq 10^4$.

4   Discussion

The uncertainty around the 50-year level clearly has a very large impact on the design process. In this paper, we have focused on wall thickness in order to produce results that are counter-intuitive. This is to demonstrate that extrapolated 50-year values can be misleading and can easily trick the designer into making bad choices. It is therefore very important that the designer is skeptical enough of their own results.

In this case, those bad choices are obvious. However, in many other cases, for example when choosing between different foundation types or when tuning the gain settings for a controller, they can be very difficult to spot. Considering this, one may wish to avoid having to deal with DLC 1.1 altogether. Yet, if DLC 1.1 is design-driving or if certain extreme load behavior is only found in very large sample sizes—which was the case with the baseline NREL 5 MW controller—doing a large number of load cases cannot be avoided.

Situations where good designs are wrongly discarded or where bad designs are wrongly accepted have a high chance of occurring when the sample sizes are small, especially during the initial design phases. 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 $N = 10^5$. The effects of changes in wall thickness that in the order of more than 10% are then easily recognizable.

The most obvious solution to reduce the uncertainty is to use high-performance computers in order to run extensive simulation campaigns (e.g., Barone et al., 2012a, b). An alternative remedy is to rely on importance sampling, which is a well-known variance reduction method that allows the user to allocate the computational resources for the most severe conditions (e.g., Bos et al., 2015; Bos and Veldkamp, 2016).

5 ## 5   Conclusions

The goal of this paper was to demonstrate the effects of the uncertainty around extrapolated 50-year loads. It showed that, unless very large sample sizes are used, DLC 1.1 is a very unreliable measure for the performance of a design. This uncertainty has a pronounced effect on early phases of the design, when computational resources are often scarce.

One should always take into account that it is very time-consuming to prove that concepts are able to reduce the 50-year
10 load, unless the design changes are very radical. In an example where the bottom tower wall thickness of the NREL 5 MW reference turbine was varied, a 10% increase in wall thickness was identified as a way to reduce the stress in only 75% of the cases with a sample size of $N = 10^3$. In fact, more than 10^5 simulations were required to decrease the probability of a false rejection to 10%. Another example where five wall thicknesses, ranging from 25 to 45 mm, were ranked in order from the lowest to highest stress showed a similar trend. With $N = 10^3$, the correct order appeared in only 25% of the cases, which
15 improved to roughly 90% for $N = 10^5$.

These results show that a critical attitude is required when judging extrapolated extreme loads. When DLC 1.1 is not the design-driver, it might be best to avoid it altogether. Otherwise, using high-performance computing or importance sampling methods will be the best approach.

References

Agarwal, P., Manuel, L.: Simulation of offshore wind turbine response for long-term extreme load prediction, Eng. Struct., 31(10), 2236–2246, 2009, doi:10.1016/j.engstruct.2009.04.002.

Barone, M., Paquette, J., Resor, B., and Manuel, L.: Decades of wind turbine load simulation, 50th AIAA Aerosp. Sci. Meet. Incl. New
5   Horizons Forum Aerosp. Expo., Nashville, TN, United States, 9–12 January, 2012, doi:10.2514/6.2012-1288.

Barone, M., Paquette, J., Resor, B., Manuel, L., and Nguyen, H.: Simulating the entire life of an offshore wind turbine, EWEA Annual Event, Copenhagen, Denmark, 16–19 April, 2012.

Bos, R., Bierbooms, W. A. A. M. , van Bussel, G. J. W.: Importance sampling of severe wind gusts, 11th PhD Semin. Wind Energy Eur., Stuttgart, Germany, 23–25 September, 2015.

10 Bos, R., Veldkamp, H. F.: A method to find the 50-year extreme load during production, J. Phys. Conf. Ser., 753, 42021, 2016, doi:10.1088/1742-6596/753/4/042021.

van Eijk, S. F.: Wind turbine load extrapolation, Master's thesis, Delft University of Technology, 2016.

Fogle, J., Agarwal, P., Manuel, L.: Towards an improved understanding of statistical extrapolation for wind turbine extreme loads, Wind Energy, 11, 613–635, 2008, doi:10.1002/we.303.

15 Freudenreich, K., Argyriadis, K.: The load level of modern wind turbines according to IEC 61400-1, J. Phys. Conf. Ser., 75, 12075, 2007, doi:10.1088/1742-6596/75/1/012075.

Genz, R., Nielsen, K. B., Madsen, P. H.: An investigation of load extrapolation according to IEC 61400-1 Ed. 3, Eur. Wind Energy Conf., Athens, Greece, 27 February–2 March, 2006.

IEC: 61400-1 Wind turbines – Part 1: Design requirements, 3rd edn., International Electrotechnical Commission, Geneva, Switzerland, 2005.

20 Jonkman, J. M., Butterfield, S., Musial, W., and Scott, G.: Definition of a 5-MW reference wind turbine for offshore system development, Technical Report NREL/TP-500-38060, National Renewable Energy Laboratory, Golden, CO, United States.

Lott, S., Cheng, P.-W.: Load extrapolations based on measurements from an offshore wind turbine at alpha ventus, J. Phys. Conf. Ser., 753, 072004, 2016, doi:10.1088/1742-6596/753/7/072004.

Moriarty, P. J., Holley, W. E., Butterfield, S. P.: Extrapolation of extreme and fatigue loads using probabilistic methods, Tech. rep. NREL/TP-
25   500-34421, National Renewable Energy Laboratory, Golden, CO, United States, 2004.

Natarajan, A., Holley, W. E.: Statistical extreme load extrapolation with quadratic distortions for wind turbines, J. Sol. Energy Eng., 130(3), 031017, 2008, doi:10.1115/1.2931513.

Pandey, M. D., Sutherland, H.J.: Probabilistic analysis of LIST data for the estimation of extreme design loads for wind turbine components, J. Sol. Energy. Eng., 125(4) 2003;125:531. doi:10.1115/1.1626128.

30 Peeringa, J.: Comparison of extreme load extrapolations using measured and calculated loads of a MW wind turbine, Eur. Wind Energy Conf., Marseille, France, 16–19 March, 2009.

Ragan, P., Manuel, L.: Statistical extrapolation methods for estimating wind turbine extreme loads, 45th Aerosp. Sci. Meet. Exhib., Reno, NV, United States, 8–11 January, 2007.