The determination of the extreme loads on wind turbines – some practical issues

Wim Bierbooms
Faculty of Aerospace Engineering
Delft University of Technology
Kluyverweg 1
2629 HS Delft
The Netherlands
w.a.a.m.bierbooms@tudelft.nl

Abstract

The probabilistic method commonly applied to arrive at the ultimate loading is as follows: for several different mean wind speeds load simulations are performed. For each mean wind speed a conditional distribution can be fitted to the load maxima for that particular wind speed. The overall distribution of the load response is obtained by a weighted average of these conditional distributions taking into account the probability of occurrence of the wind speed bins (Weibull).

Two practical issues are addressed:

- The plotting positions
- The averaging over the mean wind speeds

The plotting positions (for the m-th ranked value out of N) are unique and given by: $m/(N+1)$. This means that the plotting positions do not depend on the particular application and/or the anticipated distribution function.
The maximum of the 50 year estimates based on the exceedance probability $Q_{short}(L|U_i)$ is an upper bound of the long term 50 year load value L_{50}. A lower bound for L_{50} is given by the maximum of the estimates based on the relative exceedance probability $R_{short}(L|U_i)=Q_{short}(L|U_i)n_i$; with n_i the fraction of time for wind speed bin U_i. In the situation that load data of just 1 wind speed bin is available it is in general not possible to determine L_{50}. In case it is assumed that the considered wind speed bin governs the load, a good estimate (lower bound) of L_{50} is obtained by considering R_{short}. If it is assumed that the load distributions of the other wind speed bins are about the same, a good estimate is obtained by considering Q_{short} (upper bound).
Keywords: extreme value analysis, extreme external conditions

1 Introduction

During the design of a wind turbine the fatigue loads as well as the ultimate loads have to be considered. For the latter it is common to take the load level L_{50} with a return period of 50 years; i.e. this level is exceeded on average once every 50 years.

The challenge of an extreme value analysis is to obtain the 50 year value based on a data set which is much shorter than 50 years. Commonly the following steps are performed in order to estimate L_{50}:

- A number of N load maxima is considered, which are taken from either measurements or simulations. The maxima can be determined over any time period but usually an interval of 10-min. or 1 year is used. It is also possible to consider values above some threshold.
- The N load values are ranked to order m (m=1 the smallest and m=N the largest) and associated with a certain (non-exceedance) probability (the so-called plotting positions).

- The ranked values are fitted to some distribution function (e.g. GEV, GPD, Weibull, etc.). For this purpose some method should be used (e.g. LSQ, maximum likelihood, etc.)
- By extrapolation the requested load value with return period of 50 year is determined.

The above procedure involves several practical issues. In this paper the plotting position will be addressed in Section 2.

For wind energy applications the procedure outlined above is usually performed for each mean wind speed bin and next the results are combined (weighted averaged). Practical issues concerning this weighted average are discussed in Section 3.

2 Plotting position

2.1 Uniform histogram

As mentioned above, the ranked observations should be assigned a certain probability. This process is usually referred to as the determination of the plotting positions, since it is commonly performed graphically. In literature a whole range of different plotting positions can be found. In several recent papers [1, 2, 3] Lasse Makkonen has proven that there is only one correct plotting position (for the m-th ranked value out of N), namely:

$$P_m = \frac{m}{N+1} \tag{1}$$

Instead of repeating his derivation, an example will be given which substantiate Eq. (1). Say, 5000 observations (this can be either measurements or simulations) are available. These 5000 values can be depicted in a histogram, see Fig. 1. In the bottom graph just 5 (as example) bins are used. One may raise the question how to set the borders of these 5 bins in order to obtain a uniform histogram. In case the distribution *F* is known this is a straightforward task: the probability of a value inside each bin should equal 20% so the 4 borders are given by: $b_1 = F^{-1}(1/5)$, $b_2 = F^{-1}(2/5)$, $b_3 = F^{-1}(3/5)$ and $b_4 = F^{-1}(4/5)$; the left border of the first bin is minus infinity, the right border of the last bin equals infinity. For a Gumbel distribution (with location parameter 2 and scale parameter equal to 3) the result is shown in Fig. 2.

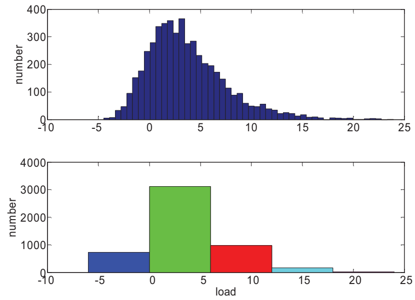

Figure 1: An example of a histogram of 5000 observations.

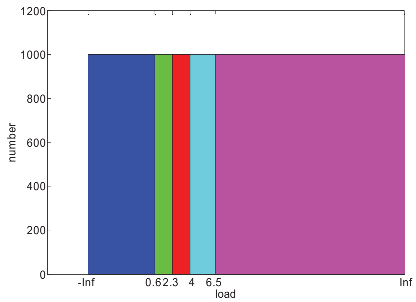

Figure 2: Uniform histogram of 5000 observations.

In case the distribution *F* is unknown, the 4 borders can be estimated by 4 random observations ranked to order: r_1, r_2, r_3 and r_4 with r_1 the smallest value (rank 1) and r_4 the largest value (rank 4). Next, 5000 (new) observations can be binned accordingly to these borders. Since it concerns a set of 4 random samples, the histogram will be different in case the procedure is repeated. In Fig. 3 results are shown for 4 different sets.

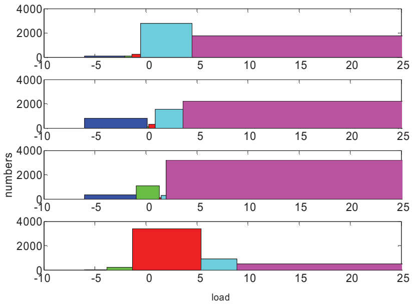

Figure 3: Four histograms of 5000 observations; the borders of the bins are based on ranked random observations.

One may average the results per bin. The results for the 5th bin (number of exceedances of r_4; magenta box in the graphs) are presented in Table 1:

Table 1: the number of exceedances of the 4th rank.

set	r_4	number of exceedances
1	4.50	1758
2	3.64	2221
3	1.97	3203
4	8.91	487
…	…	…
1000	5.48	1321
average	8.05	970
limit	7.88	1000

The averaged histogram of 1000 sets of the 4 borders is shown in Fig. 4. Please note that for each set the borders will be different (as shown in Fig. 3 and Table 1). So, it is not possible to mention load values on the horizontal axis. Instead, the order of the ranks are given.

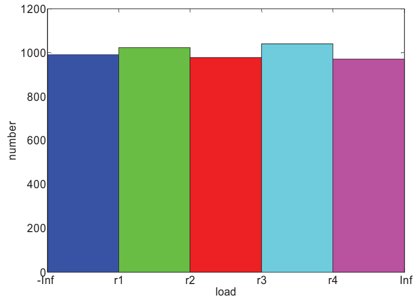

Figure 4: Averaged histogram of 5000 observations; the borders of the bins are based on ranked random observations.

In the limit that the number of sets goes to infinity, the averaged histogram will become uniform like Fig. 2. In that case the exceedance probability of r_4 equals exactly 1/5. By definition, the associated return period *R* of this event equals:

$$R = \frac{1}{P(x \geq r_4)} = 5 \qquad (2)$$

So, r_4 will be exceeded on average once every 5 (i.e. *not* 4) random samples; this makes sense since 5 bins are available.

2.2 Order statistics

We will now consider the statistics of ranked observations in more detail. As example the 4th rank (out of 4 random samples) will be dealt with. In order to distinguish the distributions it is common to indicate the ranked values by order statistics. In order that r_4 is less or equal a certain value *x* all 4 (independent) random samples should be less or equal *x*. Thus:

$$F_4(x) \equiv P(r_4 \leq x) =$$
$$P(x_1 \leq x)\,P(x_2 \leq x)\,P(x_3 \leq x)\,P(x_4 \leq x) =$$
$$F^4(x) \qquad (3)$$

with *F(x)* the parent distribution.

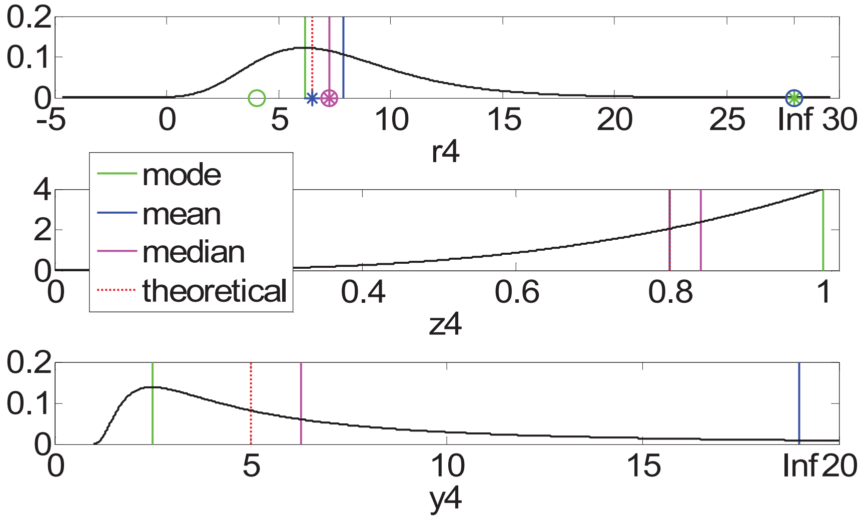

Figure 5: Top: density function of 4th order statistics r_4; as parent distribution the Gumbel distribution is taken (μ=2, σ=3)
Middle: density function of transformed value $z_4=F(r_4)$; non-exceedance probability
Bottom: density function of transformed value $y_4=R(z_4)$; return period

The density function f_4 is the derivative of F_4, so:

$$f_4(x) = 4F^3(x)\,f(x) \tag{4}$$

This density is shown in Fig. 5 (top) taking the Gumbel distribution again as parent distribution. Also the mode, mean and median is indicated in the graph. The mean value equals 7.88 as already mentioned in Table 1.

The distribution value F (non-exceedance probability) associated with a return period R is (see also Eq. (2)):

$$F = 1 - \frac{1}{R} \tag{5}$$

Consequently, the load with a return period of 5 is given by:

$$l_5 = F^{-1}(\frac{4}{5}) \tag{6}$$

For the example distribution l_5=6.5 (see also Fig. 2); in Fig. 5 this load level is indicated by a red line. It can be seen that it *not* coincides with the mode, mean or median.

Inspired by the analysis in Section 2.1 one may consider the transformation of r_4 to a new random variable z_4 through $z_4=F(r_4)$. z_4 can be considered as the normalized number of non-exceedances (in between 0 and 1) of the 4th rank. So, it is comparable with Table 1 (with the difference that in this table the number of exceedances is given). In general, a transformation to a new random variable is expressed as $z_4=g(r_4)$ with $g(x)$ some given function. The density of z_4 follows from:

$$f_4(z) = \frac{f_4(x)}{g'(x)} \tag{7}$$

So, for our case we obtain (using Eq. (4)):

$$f_4(z) = \frac{f_4(x)}{f(x)} = 4F^3(x) = 4z^3 \tag{8}$$

This density is shown in the middle graph of Fig. 5, together with the mode, mean and median. The mean value of the transformation of the 4th order statistics can be determined analytically:

$$\begin{aligned} \bar{z}_4 &= \int_0^1 z\, f_4(z)\, dz = 4\int_0^1 z^4\, dz \\ &= 4\frac{1}{5}z^5\Big|_0^1 = \frac{4}{5} \end{aligned} \tag{9}$$

The probability connected with a return period of 5 is 0.8, Eq. (5). In Fig. 5 this probability is indicated by a red dotted line; it happens to be equal to the mean, Eq. (9)

. For completeness, the back transformed values of the mode, mean and median of z_4 are shown by colored stars on the x-axis of the top graph. Since the transformation function F(x) is an increasing function the median values of r_4 and z_4 coincide.

By means of a next transformation $y_4=g(z_4)$ with $g(x)=1/(1-x)$, see Eq.(2), the accompanying return period of the 4[th] order statistics can be studied. Similar as for $f_4(z)$ the density $f_4(y)$ is obtained, see the bottom graph of Fig. 5:

$$f_4(y) = 4(1-\frac{1}{y})^3 \frac{1}{y^2} \qquad (10)$$

The mean value of y_4 equals infinity. The back transformed values of the mode, mean and median of y_4 are shown by colored circles on the x-axis of the top graph.

Although the order statistics are dependent on the parent distribution, Eq. (4) and top graph of Fig. 5, the distribution of z_4 (and y_4) is not! So, the obtained result of the mean value, Eq. (9) (as well as the middle and bottom graphs of Fig. 5) is universal.

2.3 Correct plotting positions

In the previous section the specific case N=4 and m=N is considered. The distribution of any m-th order statistics r_m (out of N) is well known and can be found in text books on statistics or e.g. [4]. The mean value of the random variable $z_m=F(r_m)$ is given by (exact):

$$\overline{z}_m = \overline{F}(r_m) = \frac{m}{N+1} \qquad (11)$$

Again, this result is independent from the parent distribution.

During an extreme value analysis a number of N load maxima are taken and ranked to order. This can be considered as taking N random samples from an unknown distribution which are subsequently ranked. The accompanying non-exceedance probability of the m-th rank is given by Eq. (11). This implies that the only correct plotting position of the m-th rank (out of N samples) is indeed given by Eq. (1).

For wind energy application, the above can be described as follows. Say, for some particular wind turbine at some site 100 10-min. load maxima (e.g. blade root flaping moment or tower base moment) are available. Applying the correct plotting positions, Eq. (1), a researcher will put the 100-th rank $r_{100,1}$ (i.e. the largest of the 100 maxima) at level F=100/101 (which is equivalent with a return period of 101 times 10-min.). The real average number of exceedances (unknown to the researcher) in 101 10-min. periods will in general not be equal to 1; we will denote it as N_1. Based on some other set of 100 10-min. load maxima (of the same turbine at the same site) another researcher will obtain some other 100-th rank value $r_{100,2}$; the real average number of exceedances of this load level equals $N_{50,2}$. Say, in total K sets of load time series are available and K different reseachers obtain K values $r_{100,k}$ with real average number of exceedances N_k. No preference whatsoever can be expressed for any of these $r_{100,k}$ estimates. The average of N_k will approach 1 in case K goes to infinity. In fact, the above is demonstrated in the example of Section 2.1.

2.4 Conclusion

In [1, 2, 3] it is proven by Lasse Makkonen that the plotting positions are unique and given by Eq. (1). This means that the plotting positions do not depend on the particular application and/or the anticipated distribution function. So, all other formulations of plotting positions than Eq. (1), which may be based on the mode, mean or median of the m-th order statistics, are incorrect and should not be applied. The same holds for software which has implemented these incorrect plotting positions.

3 Averaging over mean wind speeds

3.1 Background

The long term distribution $F_{long}(L)$ of 10-min. load maxima *L* can be obtained through a weighted average (convolution) of the conditional short term distributions $F_{short}(L|U)$ per (10-min.) mean wind speed *U:*

$$F_{long}(L) = \int_0^{\infty} F_{short}(L\,|\,U)\, f(U) dU$$
(12)

with $f(U)$ the probability density function of the mean wind speeds (commonly described by the Weibull distribution).

Eq. (12) can be discretised as follows:

$$F_{long}(L) \approx \sum_{i=1}^{N} F_{short}(L\,|\,U_i)\, n_i = 1 - n_{tot} + \sum_{i=1}^{N^*} F_{short}(L\,|\,U_i)\, n_i \quad (13)$$

with

$$n_{tot} = \sum_{i=1}^{N^*} n_i \quad (14)$$

$n_i = f(U_i)dU$ indicates the fraction of time for each wind speed bin. Normally, the considered bins are restricted to the range in between the cut-in and cut-out wind speed (so, assuming that the load outside this range is negligible. This is indicated by the summation limit N^*; the total fraction of time between cut-in and cut-out is n_{tot}, Eq.(14). Please note that it is necessary to consider the fraction of time outside this range (e.g. 1-n_{tot}) in order to have a correct normalization of the long term distribution (i.e. for very large values of the load L the non-exceedance probability should go to 1). In order to avoid this normalization issue one can consider the exceedance probability Q instead.

$$\begin{aligned}
&Q_{long}(L) = \\
&1 - F_{long}(L) \approx \\
&1 - (1 - n_{tot}) - \sum_{i=1}^{N^*} F_{short}(L\,|\,U_i)\, n_i = \\
&n_{tot} - \sum_{i=1}^{N^*} (1 - Q_{short}(L\,|\,U_i))\, n_i = \\
&n_{tot} + \sum_{i=1}^{N^*} Q_{short}(L\,|\,U_i)\, n_i - \sum_{i=1}^{N^*} n_i = \\
&\sum_{i=1}^{N^*} Q_{short}(L\,|\,U_i)\, n_i \equiv \sum_{i=1}^{N^*} R_{short}(L\,|\,U_i)
\end{aligned}$$
(15)

where we have introduced a new variable, the relative exceedance probability R:

$$R_{short}(L\,|\,U_i) = Q_{short}(L\,|\,U_i)\, n_i \quad (16)$$

The load L_{50} with a return period of 50 years is given by, Eq. (2):

$$Q_{long}(L_{50}) = \frac{1}{50*365*24*6} = 3.8\,10^{-7}$$
(17)

Due to the convolution (weighted average) according to Eq. (15) it is in general not known which exceedance values of the short term distributions $Q_{short}(L\,|\,U_i)$ are of relevance. We will investigate this in more detail by treating several examples.

3.2 Examples

A Weibull distribution with scale parameter 8 and shape parameter 2 is assumed. As cut-in and cut-out wind speed 4 m/s and 25 m/s resp. are taken. The short term distributions $F_{short}(L\,|\,U_i)$ should be obtained from load measurements or simulations. In order to study the averaging process here some distribution is simply assumed: a normal distribution. For each of the 5 examples considered some particular mean values μ_i are taken (the standard deviations are always set to 1). Any other distribution could have been taken as well, but this will not change the conclusions.

Example 1

In example 1 two wind speeds bins are considered: 5 m/s (2.5 m/s – 7.5 m/s) and 10 m/s (7.5 m/s – 12.5 m/s). The fraction of time of these wind speed bins is 49% and 33% resp. As mean values μ_i of the short term distributions, which are assumed to be normal, 0.5 and 1.0 are taken. For clarity, in this example $Q_{long}(L_{50}) = 1/10 = 0.1$ is taken instead of Eq. (17). The short term distributions $Q_{short}(L\,|\,U_i)$ are indicated by blue lines, the long term distribution $Q_{long}(L)$, Eq. (15), by a red line. Furthermore, the relative short term distributions $R_{short}(L\,|\,U_i)$, Eq.(16), are given by dashed green lines. (see Appendix for the graphs).

Due to Eq. (13) a lower bound of $F_{long}(L)$ can be determined:

$$F_{long}(L) \geq 1 - n_{tot} + F_{\min}\sum_{i=1}^{N^*} n_i \geq (1-n_{tot})F_{\min} + F_{\min}n_{tot} = F_{\min} \quad (18)$$

with $F_{\min}$ the minimum, for load level L, of the short term distributions $F_{short}(L \mid U_i)$

Similarly, based on Eq. (15), an upper bound of $Q_{long}(L)$ can be established:

$$Q_{long}(L) \leq Q_{\max}\sum_{i=1}^{N^*} n_i = Q_{\max}n_{tot} \leq Q_{\max} \quad (19)$$

with $Q_{\max}$ the maximum, for load level L, of the short term distributions $Q_{short}(L \mid U_i)$

To put the above in other words: in the graphs there should be at least one blue line on the right (or on top) of the red line.

For this example, the 50 years load L_{50} equals 1.90; in the graphs this is indicated by a red star. The values corresponding to this load level are shown in the bar graph as well as in a table (see Appendix). Notice that the relative exceedance probabilities $R_{short}(L_{50} \mid U_i)$, 4th column in the table, equals the product of the exceedance probabilities $Q_{short}(L_{50} \mid U_i)$, 2nd column, and the fraction of time n_i of the wind speed bin, 3rd column. Furthermore, the summation of the 4th column equals $Q_{long}(L_{50})$. In the graphs $R_{short}(L_{50} \mid U_i)$ are indicated by green stars; $Q_{short}(L_{50} \mid U_i)$ by blue stars. Note: in the bar graph the contribution of each mean wind speed bin is given: this is the ratio of $R_{short}(L_{50} \mid U_i)$ and $Q_{long}(L_{50})$, see Eq. (16) and (17). Per definition, the sum of the contributions equals 100%.

The 50 years load L_{50} based upon the short term distributions $Q_{short}(L \mid U_i)$ equals 1.78 and 2.28 resp. (blue circles in the graphs; 5th column in the table). This means that if 100% of the time a mean wind speed of 5 m/s (10 m/s) occurs the 50 year load would be 1.78 (2.28).

The 50 years load L_{50} based upon the relative short term distributions $R_{short}(L \mid U_i)$ equals 1.33 and 1.51 resp. (green circles in the graphs; 6th column in the table). This means that 1.33 would be the 50 year load if 49% of the time (i.e. according to the Weibull distribution) the mean wind speed equals 5 m/s; in this calculation it is also assumed that during the remaining 51% of the time the load is negligible. A similar explanation holds for the result 1.51 for a mean wind speed of 10 m/s.

From Eq. (15) it is clear that $Q_{long}(L)$ is larger than (or equal to) each of the relative short term distributions $R_{short}(L \mid U_i)$, so the maximum of the 50 years load values based upon the relative exceedance probabilities is a lower bound of the overall 50 years load L_{50}.

As stated above, at least one blue line is on the right of the red line. This implies that at least one of the 50 years load values based upon the exceedance probabilities is higher than L_{50}. So, the maximum of the 50 years load values based upon the exceedance probabilities can serve as an upper bound of the overall 50 years load L_{50}. Indeen, L_{50}=1.90 is in between 1.51 and 2.28.

In this example the value corresponding to U=10 m/s, i.e. 2.28, is higher than L_{50}=1.90. The other 50 years load value (i.e. 1.78 for U=5 m/s) based upon the exceedance probability happens to be lower than L_{50}. In general, it is not necessary that one of the 50 years load values based upon the short term exceedance probabilities is lower than L_{50} (see example 2).

Example 2

In example 2 the wind speeds from 4 m/s to 25 m/s (with a bin width of 2 m/s) are considered. As mean values μ_i of the short term distributions 1 is taken for all bins. The convolution is almost equal to the short term distribution; the small difference is due to the fact that the fraction of time of all wind speed bins is 87% instead of 100%. The contribution of each bin to the 50 year load L_{50} equals the fraction of time according to the Weibull distribution (see bar graph). A very good estimate of L_{50} = 5.92 is obtained by the 50 year value corresponding to the short term distribution $Q_{short}(L \mid U)$: 5.95 (upper bound).

Example 3

Example 3 is almost similar to example 2 except the mean value of the short term distribution $F_{short}(L \mid 14)$ for U=14 m/s is taken as 2. As a consequence L_{50} = 6.31 is governed by this wind speed bin (see also bar graph). In other words, a very good estimate of L_{50} is obtained by the 50 year value corresponding to the relative short term distribution $R_{short}(L\,|\,14)$ for this wind speed bin: 6.28 (lower bound).

Example 4

In example 4 the mean values of the short term distributions $F_{short}(L\,|\,U)$ are assumed to be linear with the mean wind speeds: $\mu = U/10$. The combination of this relation and the Weibull distribution leads to a mean wind speed of 16 m/s contributing most to the 50 year load value (see bar graph). The short term probabilities of the wind speed bins, for L=L_{50}, range from $1.1\,10^{-4}$ to $6.3\,10^{-9}$ (see 2[nd] column in the table). These probabilities are (far) less or larger than the long term exceedance probability $Q_{long}(L_{50}) = 3.8\,10^{-7}$. Since L_{50} is not governed by just one wind speed bin, it is not possible to estimate in advance these short term probabilities. To put it into other words: for the overall 50-year load value the 2-month value van be relevant for one wind speed and for another wind speed the 3000-year value.

Example 5

Finally, in example 5 the mean values of the short term distributions are assumed to have the following relation:

$$\begin{aligned} U \le 12 &: \mu = \frac{U}{10} \\ U > 12 &: \mu = \frac{36-U}{20} \end{aligned} \qquad (20)$$

The above should represent a crude estimation of a thrust versus mean wind speed relation of pitch regulated wind turbines (with rated wind speed: U=12 m/s). The most contributing wind speed is now shifted to 12 m/s. A fair estimate (lower bound) of L_{50} = 5.82 is obtained by the 50 year value corresponding to the relative short term distribution $R_{short}(L\,|\,12)$ for this wind speed bin: 5.63.

The approximation of Eq. (12) by Eq. (13) improves by decreasing the bin width. By taking a wind speed bin width as small as 0.01 m/s the calculation of the 50 year value does not change (L_{50} = 5.82). So, it is sufficient to take a bin width of 1 to 2 m/s.

3.3 Conclusions

Above, we have introduced the short term exceedance probability $Q_{short}(L\,|\,U_i)$ as well as the relative exceedance probability $R_{short}(L\,|\,U_i) = Q_{short}(L\,|\,U_i)\,n_i$; with n_i the fraction of time for wind speed bin U_i. The maximum of the 50 year estimates based on $Q_{short}(L\,|\,U_i)$ is an upper bound of the (overall) long term 50 year load value L_{50}. A lower bound for L_{50} is given by the maximum of the 50 year estimates based on $R_{short}(L\,|\,U_i)$.

In the situation that the short term distribution $F_{short}(L\,|\,U)$ of just 1 wind speed bin is available it is in general not possible to determine the overall 50 year value L_{50}. Based on the given examples one may consider the following three cases:

1) It is assumed that the considered wind speed bin governs the load
2) It is assumed that the load distributions of the other wind speed bins are about the same
3) It is assumed that another wind speed bin (or a whole range of bins) governs the load

For case 1) the situation is similar to the examples 3 and 5. A good/fair estimate (lower bound) of L_{50} is obtained by considering the relative exceedance probability R_{short}. Since this relative exceedance probability depends on the fraction of time the width dU of the wind speed bin is of importance. A range dU should be taken over which the load is assumed to be about the same level as the considered mean wind speed.

Case 2) resembles example 2. A good estimate of L_{50} is obtained by the 50 year value corresponding to the short term distribution $Q_{short}(L\,|\,U)$. In case the load distributions of the other wind speed bins are less or equal to the one of the considered wind bin, this estimate is an upper bound for L_{50}.

For case 3) no sensible statement can be made, except that the 50 year value corresponding to the relative short term distribution $R_{short}(L\,|\,U)$ is a lower bound for L_{50}. The case that a whole range of wind speed bins governs the load is like example 4. Even in the best case (U=16 m/s) the estimates 5.70 and 6.55 (based on *R* and *Q* resp.) are poor with respect to the real value for L_{50} =6.09.

References

[1] Makkonen Lasse, Plotting positions in extreme value analysis. Journal of Applied Meteorology and Climatology; 45: 334-340, 2006.

[2] Makkonen Lasse, Bringing closure to the plotting position controversy, Communications in Statistics – Theory and Methods, 37:3, 460-467, 2008.

[3] Makkonen, Lasse, Problems in the extreme value analysis, Structural Safety 30, 405-419, 2008.

[4] R.I. Harris, Gumbel re-visited – a new look at extreme value statistics applied to wind speeds, Journal of Wind Engineering and Industrial Aerodynamics 59, 1-22, 1996.