# GumbelWind – a computer code for statistical extrapolation of ultimate loads on wind turbines

M. Hänler and U. Ritschel

Windrad Engineering GmbH, Querstr. 7, 18230 Zweedorf, Germany

Ultimate loads play an important role in the design of wind turbines. In this paper we describe the framework for the calculation of ultimate loads that are equipped with a well-defined return probability, e.g. a value F being exceeded on average once within 50-years. The exceedance probability depends on wind speed and other parameters and may be given in closed form by combining Weibull and Gumbel distributions for all relevant wind speeds (from cut-in to 50-years wind). To determine the distribution of extreme values we employ the peak-over-threshold (POT) method. A comparison with the block-maxima method is carried out. For the practical application in the design process we have developed a computer code, "GumbelWind", that determines ultimate loads and synchronous components from simulated or measured time series.

## 1. Introduction

The structural dynamics of wind turbines, especially of the large turbines installed in recent years, is determined by slow vibration modes that are excited by both deterministic forces due to the operation of the turbine and stochastic forces due to the turbulent wind. It is obvious that a "stochastic component" is always present in real loads, in particular in situations relevant for the dimensioning of the design, as for example operation at high wind speeds and 50-years wind.

As a consequence of the stochastic nature of loads one has to carry out additional statistical analyses on time series from simulations to finally obtain results with a well-defined return probability. The basis for the analyses is extreme value theory from which a number of methods for the practical application are derived. This fact was also taken into account by the standards. In the new edition of the IEC 61400 [1] the appendix "statistical extrapolation of loads" is added that deals with this subject and, thus, opens the way for a consequent evaluation of extreme loads on a stochastic basis.

In this paper we describe the peak-over-threshold (POT) method [2] for the statistical extrapolation of ultimate loads. To verify the method a comparison is carried out with the block maxima method [3]. Further, a probability distribution is defined that allows to include all load cases in a consistent probabilistic scheme.

On the basis of the POT method we have designed a computer code, "GumbelWind", that provides the automatic post-processing of time series (simulated or measured). It yields ultimate loads on a predefined probabilistic basis (e.g. 50 years value), gives results for synchronous component, and carries out various quality checks. Further it provides detailed information about the contributions of individual load cases to the ultimate load level and thus provides valuable information for design optimisation

## 2. Extreme value theory for the practitioner

The new edition of the IEC standard states "it is necessary to analyse the extreme values of the loading on a statistical basis in order to determine a suitable characteristic load" [1]. To this end let us consider a typical load for which we have a time series *M(t).* The local peak values of *M(t)*, let us call them $M_i$, can be viewed as the events of a stochastic process.

Of interest with regards to extreme loads is the number of events $M_i$ above a certain threshold $M_0$ within a certain reference time interval *T*. The basic probabilistic formula for this quantity is given by

$$\begin{aligned}&\text{Prob}(M_i > M_0 \mid T) \equiv \text{P}_e(M_0, T)\\&= \int_V dV\, \text{Prob}(M_i > M_0 \mid T, V)\rho(V)\end{aligned} \quad (1)$$

It contains the wind-speed distribution $\rho(V)$. The integration runs over all wind speeds that potentially contribute to the ultimate load. The probability $\text{Prob}(M_i > M_0 \mid T, V)$ is given by

$$\text{Prob}(M_i > M_0 \mid T, V) = 1 - \left[F_{\max}(M_0 \mid V)\right]^{n(V,T)} , \quad (2)$$

where $F_{\max}(M \mid V)$ is the (integral) probability function of local peaks and *n(V,T)* is the number of peaks in *T*. On the right hand side of (2), the time dependence comes in only via *n(V,T)*. The probability $F_{max}$ for the peak values itself is time-independent.

A concrete example for the design value would be the level that is exceeded on average once every 50 years. In that case the level $M_0$=$M_{50}$ is determined from the implicit equation

$$\text{P}_e(M_{50}, T) = \frac{T}{50\,years} = 3.8\times10^{-7} \quad \text{for } T = 10\,\text{min} \quad (3)$$

The problem with the above procedure is that the probability function $F_{max}$ is in general not known. Thus even if the number of peak values is counted carefully, it is hardly possible to derive a reliable estimate for $M_{50}$. As a consequence, the formulae in the IEC standard [1] can not be utilized directly for the actual work.

This is where the extreme-value theory (EVT) comes in. For ultimate state analyses only the extreme events are of interest. EVT states that for a large class of underlying probability distributions, the distribution of the extreme values takes a simple form, a function with a small number of parameters that can be easily fitted to the sample data obtained from the analysis of the time series. The precise form of this function depends on the way the extreme values are chosen. As mentioned above there are two methods in EVT to extract extreme values from a given set of data. We will describe them here briefly.

The block-maxima method [3] was historically the first to be developed. The idea is to decompose the observation period into blocks of equal length and choose the extreme value of every single block. In the limit of large blocks the distribution of the chosen extreme values will approximate

$$G_{\mu,\sigma,\xi}(M_i) = \begin{cases} e^{-(1+\xi\frac{M_i-\mu}{\sigma})^{-\frac{1}{\xi}}} & \xi \neq 0,\ 1+\xi\dfrac{M_i-\mu}{\sigma} > 0 \\ e^{-e^{-\frac{M_i-\mu}{\sigma}}} & \xi = 0 \end{cases} \quad (4)$$

the so called generalized extreme value distribution, where $\xi$, $\sigma$ and $\mu$ are parameters of the distribution that may be used to fit the data.

A very powerful alternative method based on the theorems of EVT is the POT method [2]. In the POT scheme a threshold $M_t$ is set and the conditional probability that $M_i < M_0$ subject to the condition that $M_i > M_t$ is considered. It can be shown that the conditional probability approaches the function

$$P_{\sigma,\xi}(x) = \begin{cases} 1-(1+\dfrac{\xi}{\sigma}x)^{-\frac{1}{\xi}} & \xi \neq 0,\ 1+\dfrac{\xi}{\sigma}x > 0 \\ 1-e^{-\frac{x}{\sigma}} & \xi = 0 \end{cases} \quad (5)$$

called the generalized Pareto distribution, where $x = M_i - M_t$ *and* $\zeta$ and $\sigma$ are parameters that determine the form of the function and may again be obtained by fitting to the data.

The $M_i$, used for the analysis must be statistically independent realizations of a stochastic process. Now, in the case of the wind turbine the time dependence of the loads is a result of stochastic forces acting on an oscillating system. As a result, a particular high peak event tends to be surrounded by other high values that are deterministically related to each other. As will be shown in the next section, the POT count has to be supplemented by a procedure that assures statistical independence of the data, whereas with the block maxima method the statistical independence is automatically implemented. On the other hand, the POT method uses the given data much more efficiently such that POT is clearly our preferred method.

The virtue of the POT method is that the simple form (5) can be fitted to data obtained by counting the peaks over a certain threshold. Since the functional form is given, standard methods from statistics like the maximum-likelihood method can be used to determine the unknown parameters, and tests of the goodness of fit like the $\chi^2$-test can be employed straightforwardly.

## 3. Application in load calculations

In this section the application of the POT method to ultimate loads is described. It may be used to determine ultimate levels for loads, i.e. sectional forces and moments, and other quantities as for example the blade-tip-to-tower distance. Further, it may be used to analyze both measured and simulated time series.

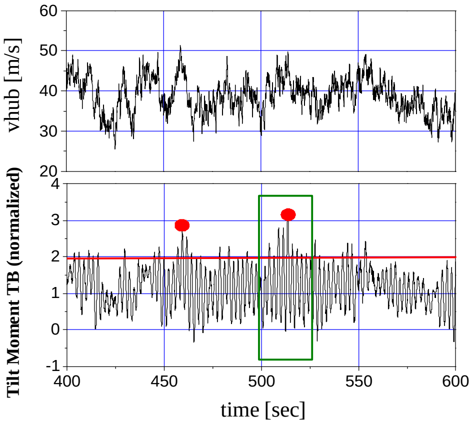


**Fig. 1:** *Procedure to determine statistically independent peak values. Further explanations are given in the text.*

Fig. 1 shows part of a simulated time series for the bending moment at the tower base for a typical horizontal axis wind turbine. The bending moment (lower graph) is normalized by the average bending moment at rated wind speed. The upper graph displays the hub height wind speed in sufficiently large distance in front of the rotor. In this particular example the load case with

50-years wind with the turbine idling has been considered. The method, of course, is not restricted to 50 years wind.

To apply the POT method the following steps have to be taken:

- Generate time series of loads (or other parameters as for example tip-to-tower distance) with statistical independent wind turbulence time series. About 15 time series of 10 minutes length (per operational state) are needed to obtain reliable results for loads.
- Choose a threshold and carry out the POT count. In Fig. 1 the threshold (depicted by the red line) was chosen at the load level 2 for example.
- Make sure that the peaks used to determine the Pareto parameters are statistically independent. Here one has to find a compromise between long-time spans between the selected peaks and a large enough sample from a limited number of time series. In the example all peeks in the neighbourhood of the marked peek within the green rectangle were discarded.
- Fit the Pareto distribution to the selected POT events, for example with the maximum likelihood method.
- Carry out quality checks to assure statistical independence and quality of the fit on the basis of the mean excess function and the $\chi^2$-test.
- Obtain the desired ultimate value, for example the 50 years load, on the basis of equation (3)

If the quantity is a component of a load vector, the synchronous components have to be determined as well. It turns out that the synchronous components can be described as a sum of a term that is linearly correlated to the analyzed quantity and a normally distributed contribution. This opens the possibility to determine the synchronous components equipped with a certain probability.

The procedure described above determines the ultimate value of a certain load component and the synchronous components on a well defined probabilistic basis. A concrete example for the application is given in the next section.

## 4. Example: 50-years tower loads

In order to define the tilt moment and synchronous components, the local coordinate system at the tower base is defined as follows: x-direction along the tower axis downwards, y-direction horizontal and perpendicular to the nacelle axis, and z-direction horizontal along the axis of the nacelle.

To determine the 50-years value for the tilt moment My, we have analyzed 80 time series according to the rules described in the previous section. The result of the POT analysis is that the 50-years value of the load is 3.28. The highest observed level lies about 10 % above the 50-years load. The reason for this is simply that with 80 time series we have simulated much more time with extreme wind conditions than would occur within 50 years.

Quality checks for the obtained distribution are the $\chi^2$-test and the mean excess function. In the example considered here the condition was imposed that the $\chi^2$ value was within the 95% quantile.

Concerning synchronous components, we find that Fx and Fz are correlated with My. Fy and Mz are uncorrelated. Subtracting from each component the correlated part, one finds that the remainder can be described well by a normal distribution centred at zero. Fitting the width of the normal distribution one finally can determine synchronous load components with a certain probability.

## 5. POT versus block-maxima method

In order to check whether the critical 50-years value obtained above is independent of the method chosen, we implemented the block maxima method essentially in the same way as described above for the POT method. The main difference is that we have a 3D parameter space now. On the other hand it is not necessary to take additional measures to guarantee statistical independence of data and how to choose a threshold.

The largest difference between the results is about 2.5%. This uncertainty or error is in agreement with our expectations taking into account the sample sizes used in the analyses.

The main difference of both methods lies in their efficiency. According to our experience a factor 4 of time series is needed for the block maxima method compared to POT in order to give stable results, such that the POT method is clearly preferred.

## 6. Extension to general load situations

Using eqn. (1) (also contained in Ref. [1]) as a starting point it is possible to include any load case equipped with a certain probability in the extrapolation scheme. The remaining task is to calculate the short term load distribution for this case.

To this end we have combined the loads from energy production (DLC 1) and from extreme wind situations (DLC 6). The wind as the source of stochastic behaviour in simulations consists essentially of three different contributions distinguished by their recurrence period. All components of the wind with recurrence periods of less the 10 minutes are called turbulence and are covered by power spectra (von Karman, Kaimal, etc.). Wind events with recurrence periods from 10 minutes to some months are covered by the annual distribution (Weibull, Rayleigh). Extreme situations are taken into account by extreme wind velocities ($v_1$, $v_{50}$) which are defined just by there recurrence period.

In order to combine normal and extreme wind situations one has to find a combined distribution of the wind velocity which is on one hand valid for all velocities and on the other hand in good agreement with the IEC standard on both ends of the speed range. In extreme wind situations, the square of the wind speed follows a Gumbel distribution. As suggested in Ref. [4], we fit the square of the speed to a Gumbel distribution, utilizing the fact that at $v_1$ and $v_{50}$ the exceedance probability is known. From that we obtain a consistent velocity distribution at the right end of the speed range (high speeds). At the left end we start with the Rayleigh distribution according to IEC standard. Finally normalisation and the requirements of continuity lead us to the point where both distributions can be fitted together.

## 7. Features of GumbelWind

The newly designed computer code GumbelWind determines ultimate loads and synchronous components utilizing the POT-method of extreme value theory. In its present configuration the program is designed to handle both production load cases (DLC 1.1) and extreme wind cases (DLC 6). The user has to supply time series separated into wind bins. The bins are defined by mean velocity and direction of the wind.

Given the annual mean wind velocity GumbelWind calculates the ultimate loads and synchronous components according to the corresponding IEC class. Internally the best fit of the distribution of POT-values to a generalized Pareto distribution is computed for every load sensor and every wind bin utilizing a maximum likelihood procedure. This optimal distribution is cross-checked against the data with the help of a $\chi^2$ adaptation test to a level of significance of 95%.

The result of GumbelWind is the matrix of extreme loads and synchronous components for the extreme loads. Furthermore the contribution of the individual load cases – several situations might contribute now to the ultimate load – is provided, which may yield valuable information for the design optimisation.

## 8. Summary and Outlook

In this paper we have described a procedure to determine ultimate loads on wind turbines. More details on the method and its application are contained in Ref. [5]. Utilising the peak-over-threshold method, we were able to design a computer code that allows:

- to determine an ultimate load level with a certain probability, as for example that the level is exceeded on average once within 50 years,
- to carry out quality checks in order to guarantee that the conditions for the application of extreme value theory are satisfied.
- to determine the values of the synchronous load components.

The method can be viewed as a post-processing tool for load data comparable for example with the rain-flow-count procedure used for fatigue calculations. It eliminates the statistical uncertainty of absolute ultimate values from a limited sample. As a consequence it justifies a reduction of the safety factor, as suggested also in the IEC standard [1]. The method can be straightforwardly applied to other quantities like for example the tip-to-tower distance. It can also be used for the evaluation of measured data.

## References

[1] Int. Standard IEC 61400-1, Wind turbines – Part 1: design requirement, 3. ed. 2005.

[2] J. I. Pickands, Statistical inference using extreme value order statistics, Annals of Statistics, (1975):119–131

[3] E. J. Gumbel, Statistics of Extreme, Columbia Univ. Press, New York, 1958.

[4] N. J. Cook, The designers guide to wind loading of building structures, Part 1, Butterworth (1985)

[5] M. Hänler et al., Systematic modelling of wind turbine dynamics and earthquake loads on wind turbines, Proceedings of the EWEC, Athens (2006)