Statistical extrapolation: prediction of the 50-year maximum wind speed.

Author

Mazzini, Gonzalo Eduardo - s202683

December 28, 2020

1 Introduction

In the context of extreme winds estimation, one data set of wind measurements from the DTU Høvsøre test station in Jutland, Denmark, was analysed. It contains 10-minute averaged wind speeds and the correspondent wind direction (which is not taking into account for this analysis) for almost 16 years. Figure 1 shows the wind speed time series.

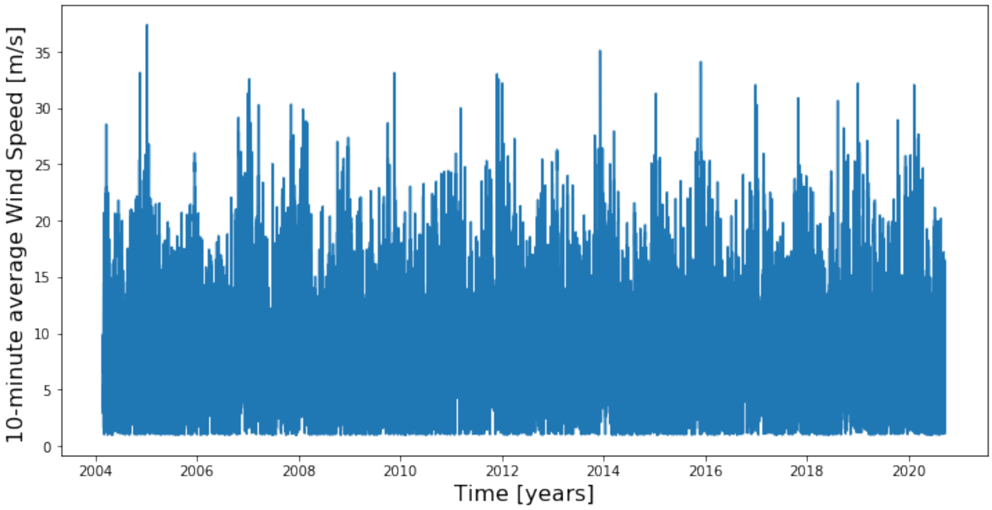

Figure 1: 10-minute averaged wind speed

In Table 1 basic statistic parameters of the wind speed measurements are shown:

Wind speed series (statistic parameters)	
Mean	9.21 [m/s]
Std.	4.42 [m/s]
Min	0.51 [m/s]
25 %	5.99 [m/s]
50 %	8.79 [m/s]
75 %	11.94 [m/s]
Max	40.59 [m/s]

Table 1: Mean, Standard deviation, Min., Max., Median, 25th and 75th percentile of the wind speed time series

The main goals for this assignment are the following:

- Estimation of the 50-year max. wind speed using measurements from the DTU Høvsøre test station fitting a Gumbel distribution.
- Compare two different methods of fitting (Probability Weighted Moments and Maximum Likelihood Estimation).

- Evaluate the uncertainty of the results when using segments of different length (between 2 and 50 years) for the calculation.
- Calculate an expression for the standard deviation of the 50-year max. wind speed as function of the segment length.

2 Calculation of the 50-year maximum wind.

The Gumbel distribution is used to model the distribution of the maximum (or minimum) of a number of samples of certain distributions whose tails decay relatively fast. It is commonly used for predicting occurrences of natural extreme events such as flood water levels and high winds. The expression for the Gumbel probability density function (pdf) is the following:

$$f(z) = \frac{\exp(-z - \exp(-z))}{\alpha} \tag{1}$$

where z is:

$$z = \frac{x - \beta}{\alpha} \tag{2}$$

Notice that α is the scale parameter and β the location parameter.

2.1 Using all the data-set:

The expected 50-year maximum wind was calculated using the Probability Weighted Moments (PWM) method to fit a Gumbel distribution. The following equations for calculate α and β were used:

$$\begin{aligned} \alpha &= \frac{2b_1 - b_0}{ln(2)} \\ \beta &= b_0 - \gamma\alpha \end{aligned} \tag{3}$$

where γ is the Euler constant (=0.577), and b_0 and b_1 are respectively:

$$\begin{aligned} b_0 &= \frac{1}{N}\sum_{i=1}^{N} Umax \\ b_1 &= \frac{1}{N}\sum_{i=1}^{N} \frac{i-1}{N-1} Umax. \end{aligned} \tag{4}$$

$Umax$ refers to the yearly maximum wind speed (sorted in ascending order) and N the number of years (14 in this case) used for the calculation. Finally, the expected 50-year maximum can be computed using:

$$V_{50max} = \beta - \alpha.ln[-ln(1 - \frac{1}{50})] \tag{5}$$

In Table 2 the results using all the data are presented;

$V(50)$	α	β
41.9 [m/s]	2.8	31.1

Table 2: Expected 50-year maximum w.s, α and β for the PWM method using all the data.

2.2 Gumbel fitting using only two years:

Instead of using the fourteen years of measurements, now two years periods are used for fitting the Gumbel with the PWM method, i.e. just two maximum wind speeds. Table 3 shows the years and the corresponding values for expected 50-year maximum, α and β.

Period	$V(50)$	α	β
2005/06	66.5 [m/s]	8.2	34.3
2007/08	38.2 [m/s]	1.9	30.6
2009/10	52.6 [m/s]	6.2	28.2
2011/12	34.4 [m/s]	0.6	32.0
2013/14	66.1 [m/s]	8.4	33.2
2015/16	42.1 [m/s]	2.3	33.1
2017/18	30.1 [m/s]	0.2	30.2

Table 3: Expected 50-year maximum, α and β for two years chunks using the PWM method.

From Table 3 the Coefficient of Variation (CV) was calculated, in order to obtain a standardized measure of dispersion for V_{50}, α and β. As shown in Table 4, the value of α is the most sensitive to variations when using just two years chunks.

	V_{50}	α	β
CV	0.29	0.82	0.49

Table 4: Coefficient of Variation ($= \sigma/\mu$).

2.3 Fitting with the Maximum Likelihood method:

Another method for distribution fitting is the Maximum Likelihood Estimation (MLE). The MLE is one of the most widely used methods for fitting statistical distributions. In this method, the parameter values are found such that they maximise the likelihood that the process described by the model produced the data that were actually observed. The following table compare the results between both methods:

Fitting method	V_{50}	α	β
PWM	41.9 [m/s]	2.8	31.1
MLE	42.1 [m/s]	3.1	30.0

Table 5: Expected 50-year maximum, α and β for the both PWM and MLE method.

2.4 How well does it follow a Gumbel distribution?

The Kolmogorov-Smirnov test (KS-test) can be used to compare a sample with a reference probability distribution, which tests whether one empirical distribution differs substantially from the theoretical. For the Kolmogorov test, the null hypothesis is that the two samples have the same distribution. In this case, the One-way Kolmogorov Test (comparing a data sample to a theoretical reference distribution) results are the following:

Komolgorov-Smirnov test	PWM	MLE
statistic D	0.19	0.17
p-value	0.501	0.681

Table 6: Results for the One-way K-S test.

The statistic D is the maximum absolute difference between the two cumulative distributions functions (empirical for the sample and the cdf of the reference distribution), while the $p - value$ represents the probability of the null hypothesis. Thus, the null hypothesis cannot be rejected for neither PWM and MLE, as it is in both cases higher than 0.05 (most common level of significance used). Moreover, from the results it is observed a better fit for the Maximum Likelihood Method (MLE) because it has a lower statistic D and thus a higher $p - value$.

3 Uncertainty in the 50-year wind speed.

3.1 Gumbel distributed simulation for 1.000 maximum yearly wind speed.

With the final aim of analyzing the uncertainty of the predicted V_{50} value, a sample of 1.000 random values of the uniform distribution (between 0 and 1) was generated using the Monte Carlo algorithm.

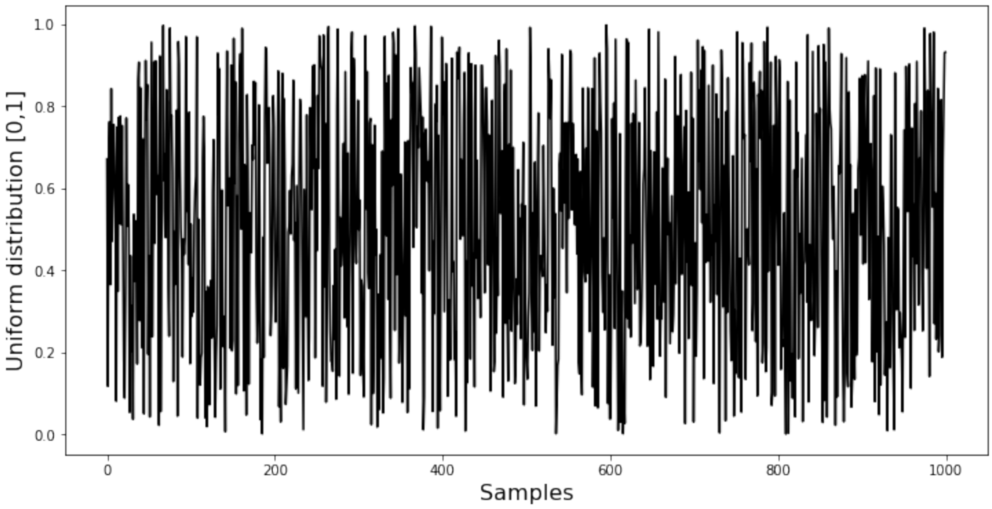

Figure 2: Uniform (0=lower, 1=upper) sample using Monte Carlo

Next, applying the inverse CDF of the Gumbel distribution (using α and β from table 1) results in a sample of 1.000 simulated yearly maximum wind speed.

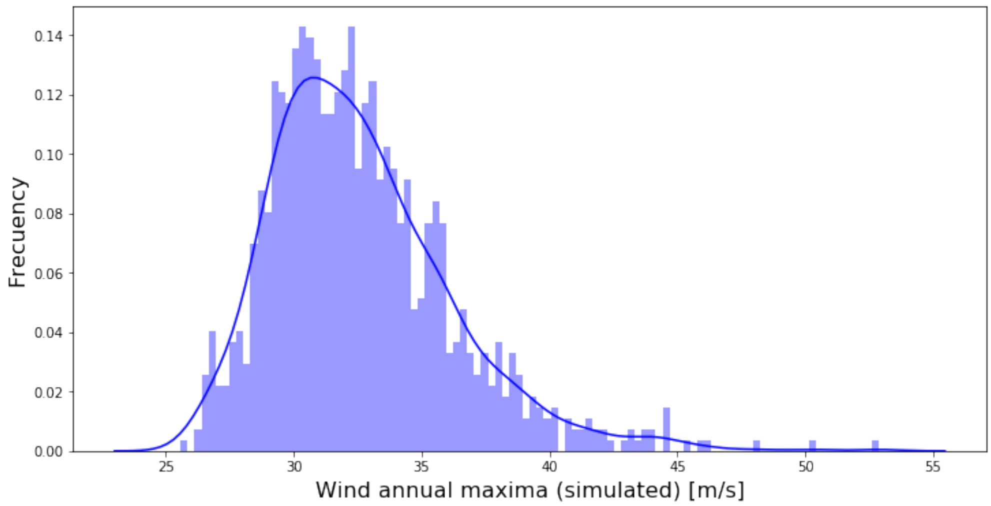

Figure 3: Simulated yearly maximum

3.2 Distributions of V_{50} and α.

Making use of the simulated maximum wind speed values and taking bootstrap-sampling of two years periods (T_{es}=2), i.e. two maximum wind speed, 1.000 values of V_{50} and α were calculated. The distribution for α can be seen in the following plot:

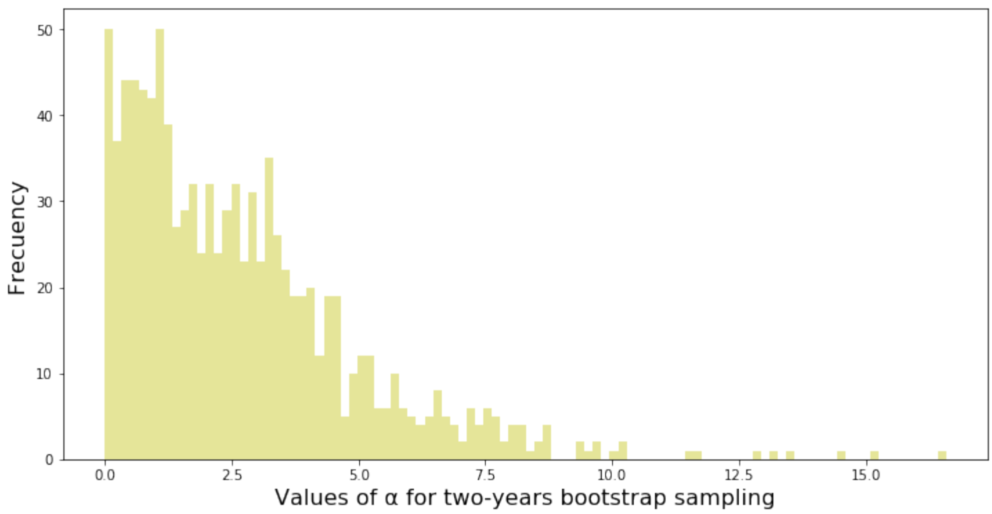

Figure 4: Values of α using two-years bootstrap sampling

For the expected maximum wind speed (V_{50}), the resulted distribution is shown in Figure 5. The standard deviation value for the V_{50} wind speed using two-years bootstrap is $\sigma = 10,4$ [m/s].

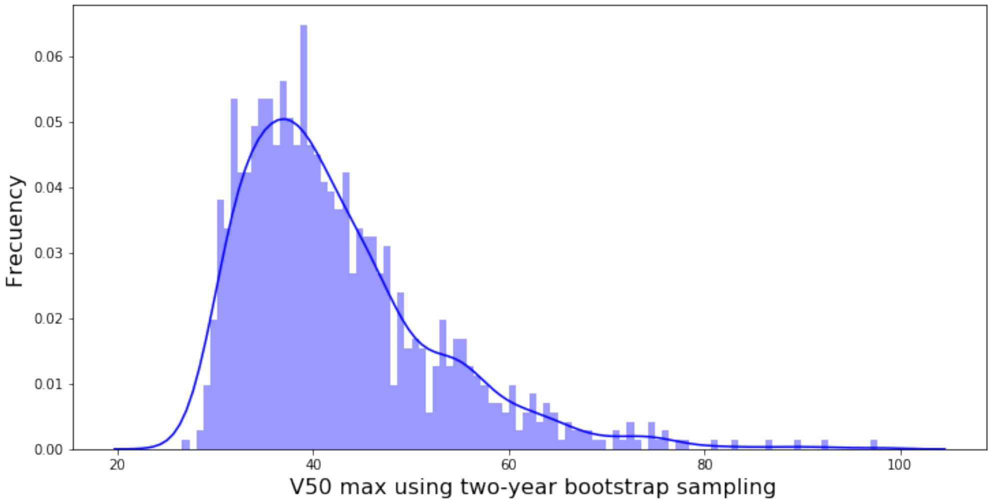

Figure 5: V_{50} using two-years bootstrap sampling

3.3 Getting an expression of σ_{v50} as a function of T_{es}

In order to get an expression for σ as a function of T_{es}, bootstrap-sampling with different values of T_{es} was performed. With the obtained values of σ, different polynomial fits has been made. Finally, after analyzing the error of the different polynomial expressions, the 5th grade was chosen, due to the small difference in the error rate between the 5th and 6th grade polynomial fit. Figure 6 shows the polynomial fits for the 4th and 5th grade, and Table 7 displays the calculated coefficients for the 5th grade.

$a0$	$a1$	$a2$	$a3$	$a4$	$a5$
16.320	-3.834	0.459	-0.0253	6.194e-04	-5.360e-06

Table 7: Coefficients of the 5th grade polynomial expression for σ

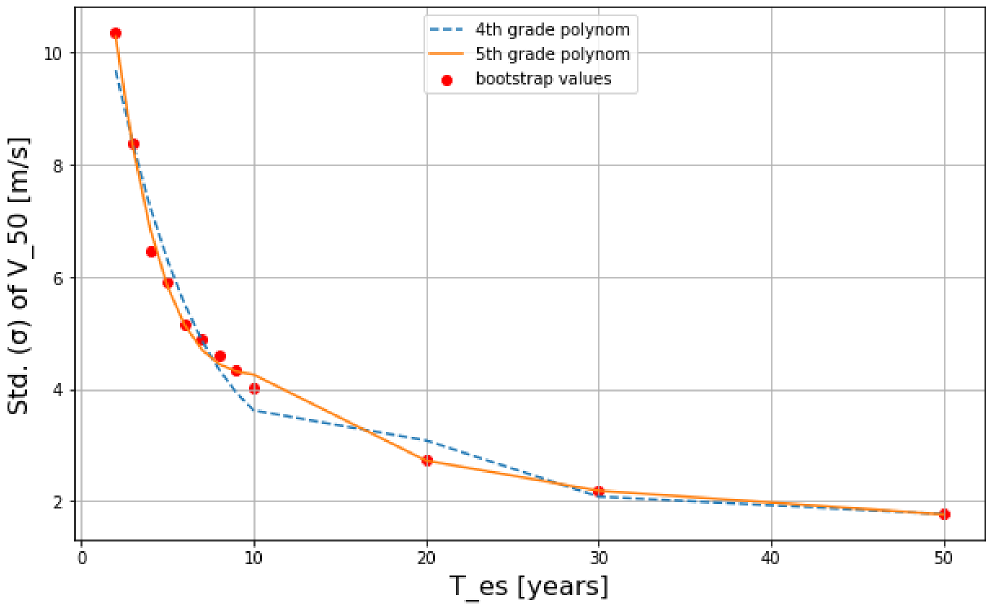

Figure 6: Polynomial fit (4th and 5th grade) for the Std. of V_{50}

3.4 Empirical expression (Ott 2011)

The following expression for the variance of the extreme wind estimate was obtained via Monte Carlo simulations (Ott 2011):

$$VarUt = \frac{\alpha^2\pi^2}{6N}(1 + 0.584q_t + \frac{0.234q_t^2}{1 - 0.823/N}) \tag{6}$$

where q_t is:

$$q_t = \frac{ln(T/T0) - \gamma}{ln2} \tag{7}$$

In Figure 7 the above expression is plotted for different values of α.

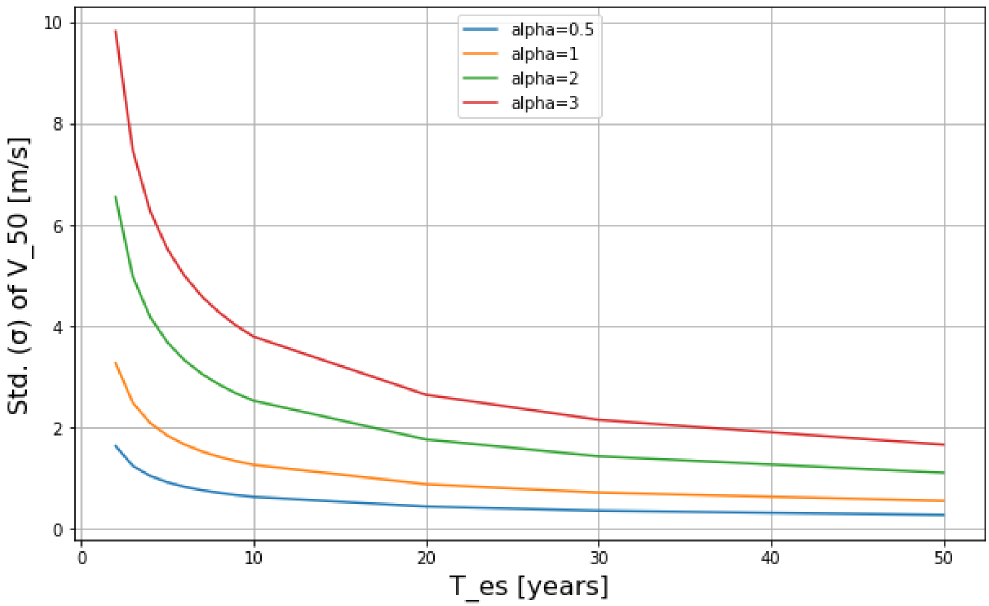

Figure 7: Corrected uncertainty via Monte Carlo (Ott 2011) for different α values.

Finally, the above expression was used to compute which value of α should be used to obtain the best fit to the bootstrap values obtained in section 3.2. The result was an α of 3.1825.

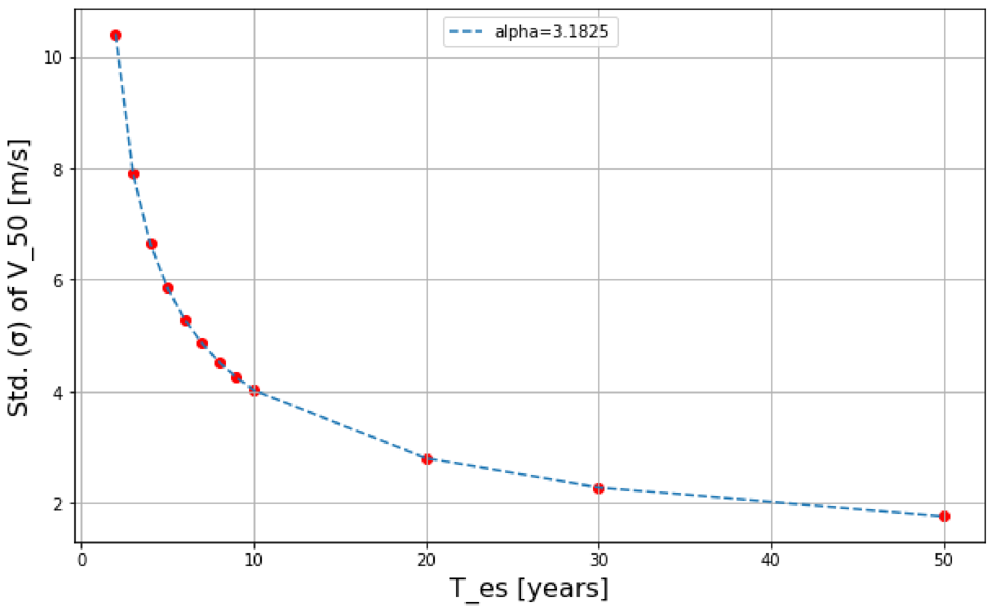

Figure 8: Ott expression for α value of 3.1825 showing the best fitting to the data obtained in section 3.2

Conclusions

- Given a time series measurements of wind speed, a prediction of the 50-year max. wind speed can be estimated, using statistical extrapolation.
- The Maximum Likelihood Estimation shows a slightly better fit (for α and β) to the Gumbel distribution, according to the KS-test in section 2.4.
- According to the Coefficient of Variation (Table 4), the value of α is the most sensitive when using different two years chunks.
- From Figure 7 can be seen that the higher the value from α, the more uncertainty in the 50-year Wind speed.
- Even though the 50-year Wind speed can be estimated with just two years of information (i.e two maximum wind speed), the uncertainty of that estimation will be high. As Figure 7 shows, the uncertainty decreases significantly, especially between 2 and 10 years chunks.
- The value of α equal to 3.1825 gives the best fit for the Ott expression to the obtained data via bootstrapping.