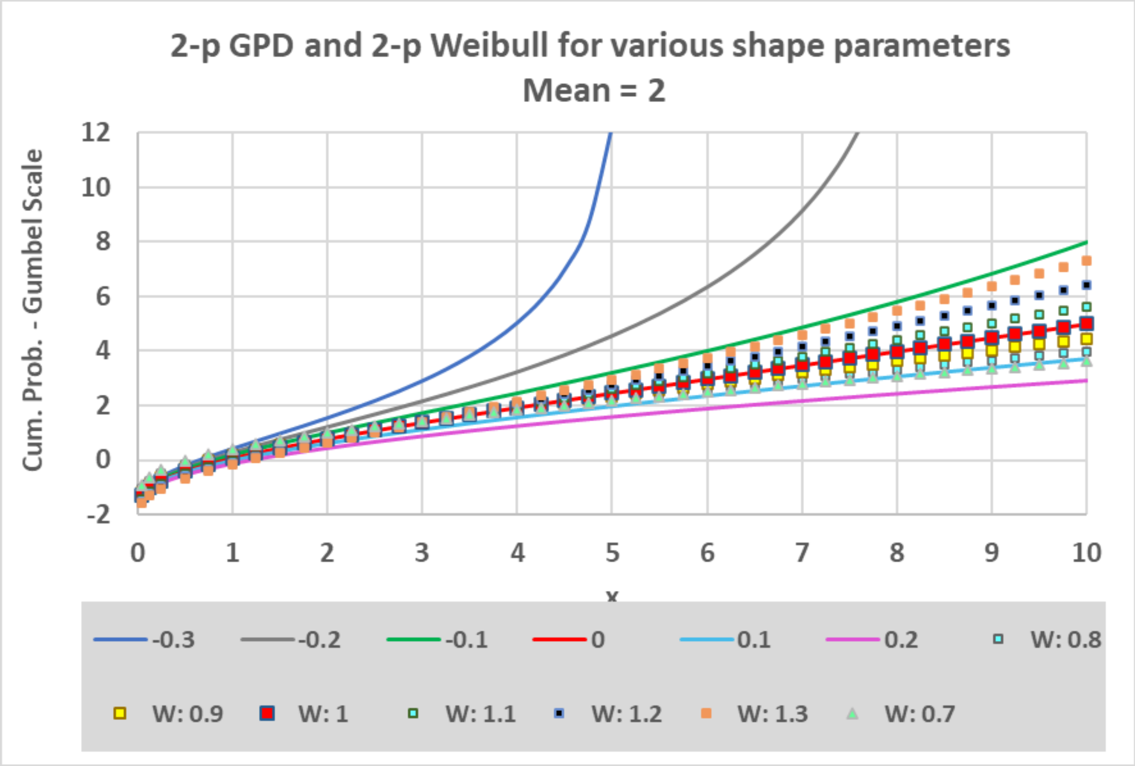
Why shall we use Generalized Pareto in practice:

Do a sufficient amount of data belong to the domain of asymptotic behaviour?



Sverre Kristian Haver, University of Stavanger, Haver & havet

LRFD design practice:

A Semi-Deterministic Design Recipe

- ULS (External action effects: 10-2 per year)

$$\gamma_p x_p + \gamma_v x_v + \gamma_e x_e \leq \frac{y_c}{\gamma_m}$$

ULSa: $$1.3x_p + 1.3x_v + 0.7x_e \leq \frac{y_c}{1.15}$$

ULSb: $$1.0x_p + 1.0x_v + 1.3x_e \leq \frac{y_c}{1.15}$$

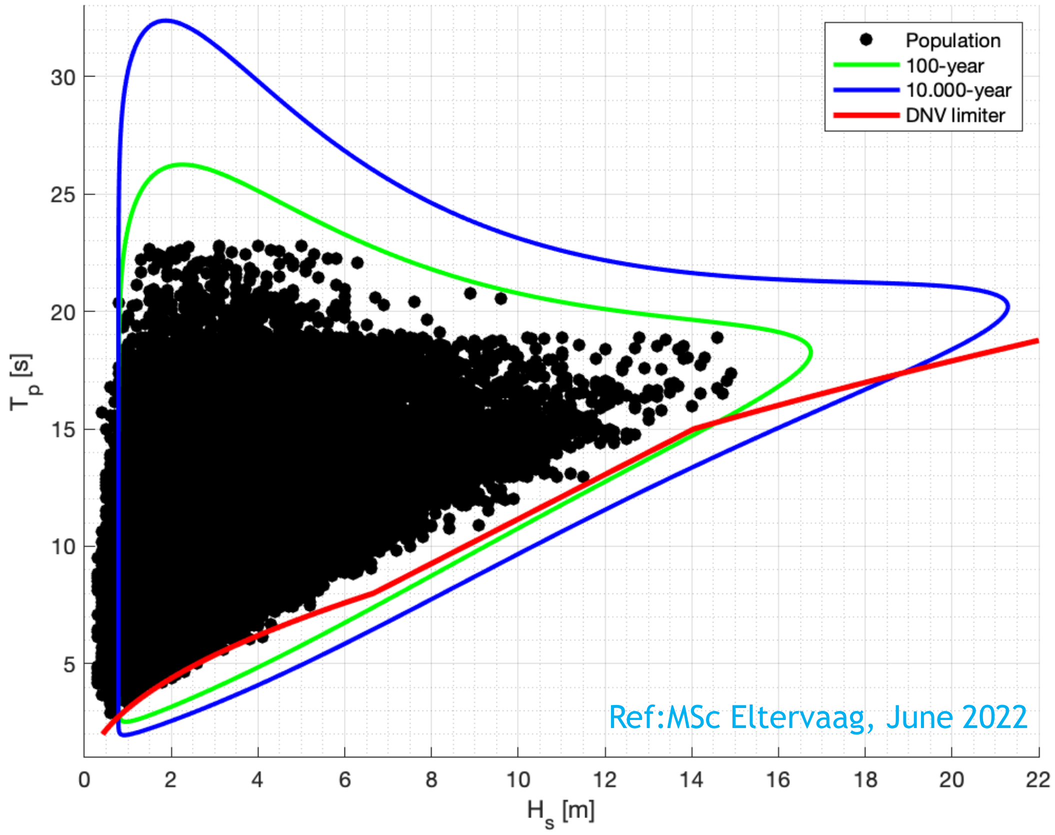
100-year action effect

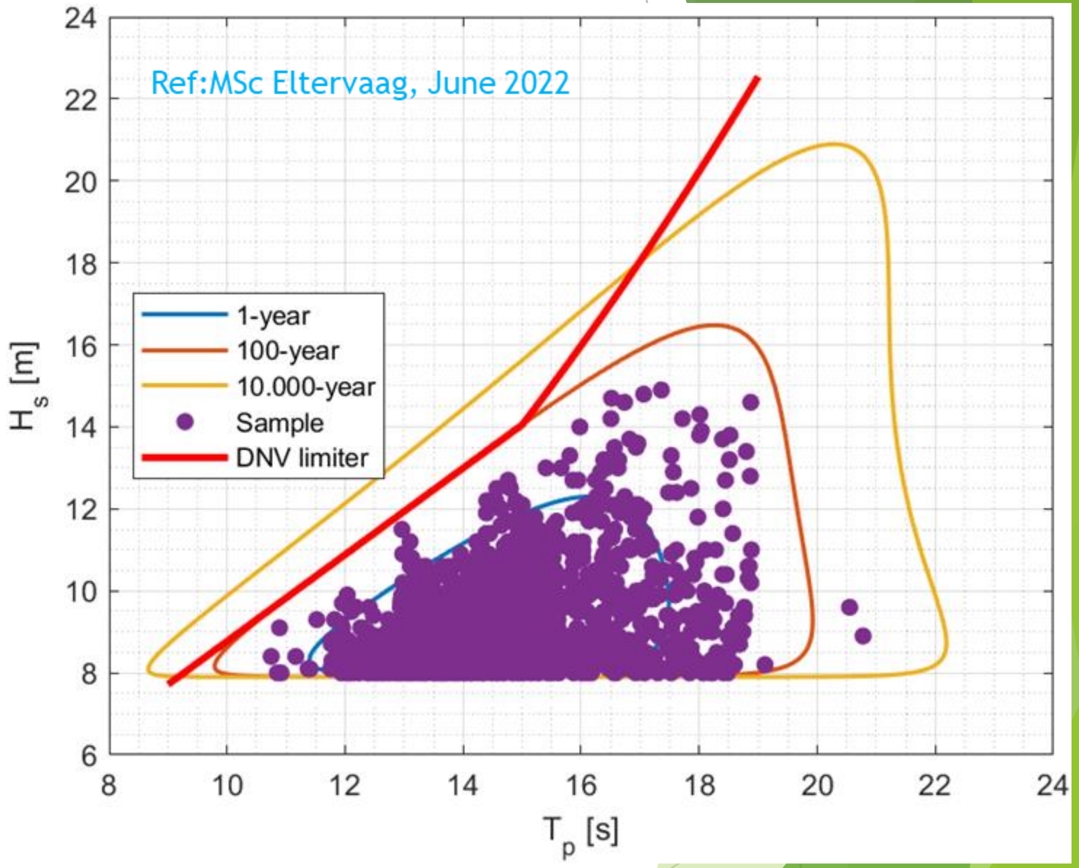
- ALS (Accidental limit state: 10-4 per year

$$1.0x_p + 1.0x_v + 1.0x_e \leq \frac{y_f}{1.0}$$

10000-year action effect

All or all important) «sea-states-approach»





$$F_{H_s}(h; h \geq 8m) = 1 - exp\left\{-\left(\frac{h-\alpha}{b}\right)^{\gamma}\right\}$$

α = 7.9m, β=1.3 and γ = 1.1

→ Dommain of attraction
of block maximum of Hs_n
Is corresponding to Type I extrem
value dist. as $n \rightarrow \infty$

Asymptotic behaviour of P[X> $y+u|X>u$]

(Storm-Peaks-over-Treshold Approach)

- It is suggested by many experts that as u becomes large enough, the distribution of the variable $Y = X - u|X > u$ should for a general case approach the Generalized Pareto Distribution.

$$F_Y(y;a,b,u) = 1 - \left(1+\frac{ay}{b}\right)^{-\frac{1}{a}}, \mathrm{a} \neq 0$$ (Type II and III domain of attraction)

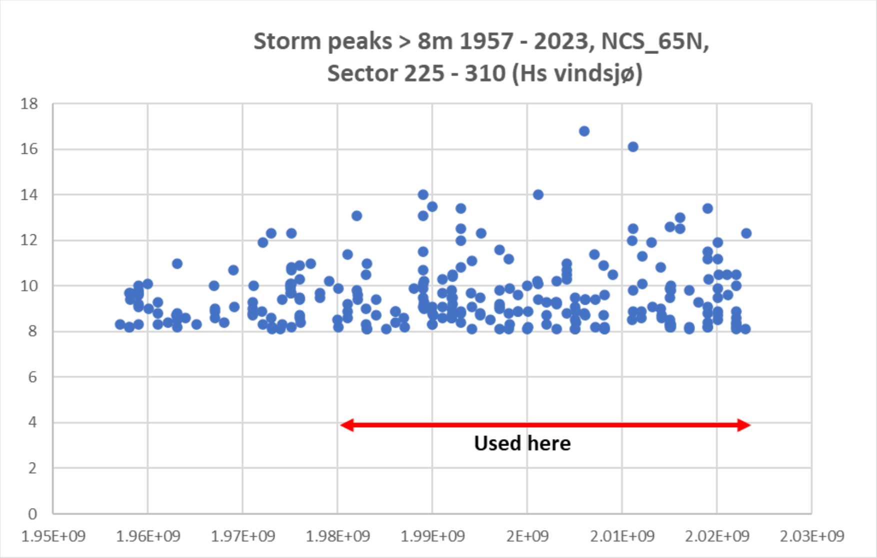
$$F_Y(y;a,b,u) = 1 - \exp\left(-\frac{y}{b}\right), \mathrm{a} = 0$$ (Type I domain of attraction.)

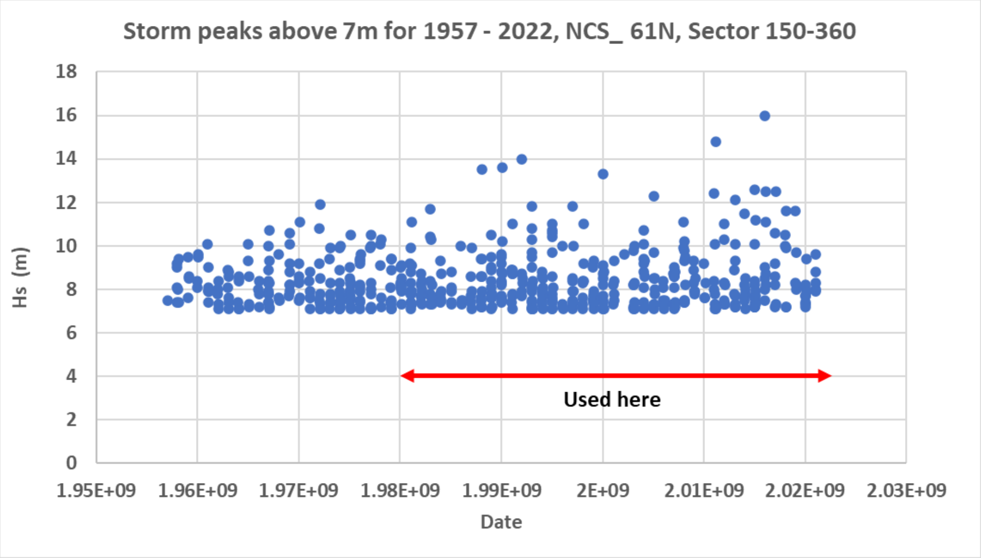
- An alternative distribution function, 2-p Weibull distribution, is frequently used based on empirical grounds:

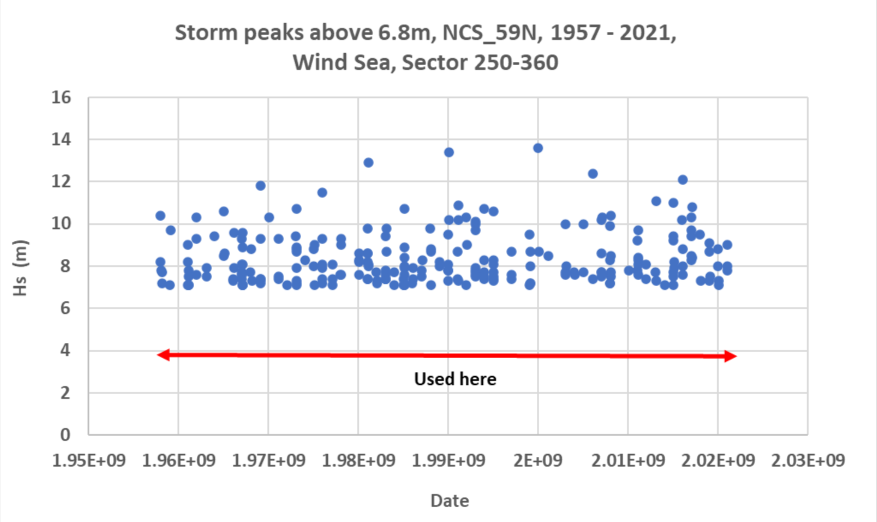
$$F_Y(y;\beta,\gamma,u) = 1 - \exp\left\{-\left(\frac{y}{\beta}\right)^{\gamma}\right\}$$ (For $\gamma = 1$ Weibull ➔ Exponential)

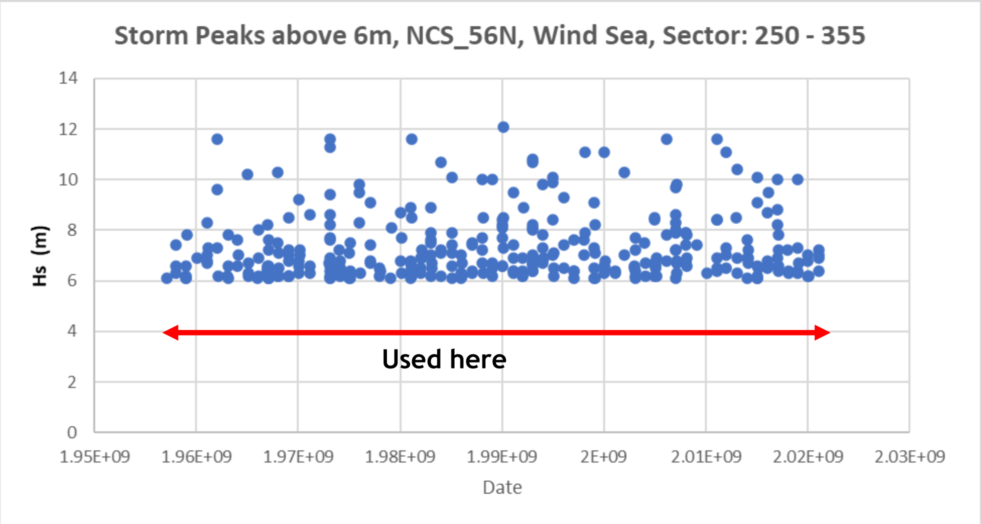
Storm peaks versus time for selected positions

NCS: Norwegian Continental Shelf, Data: NORA10 data base, Reistad et al.(2011)



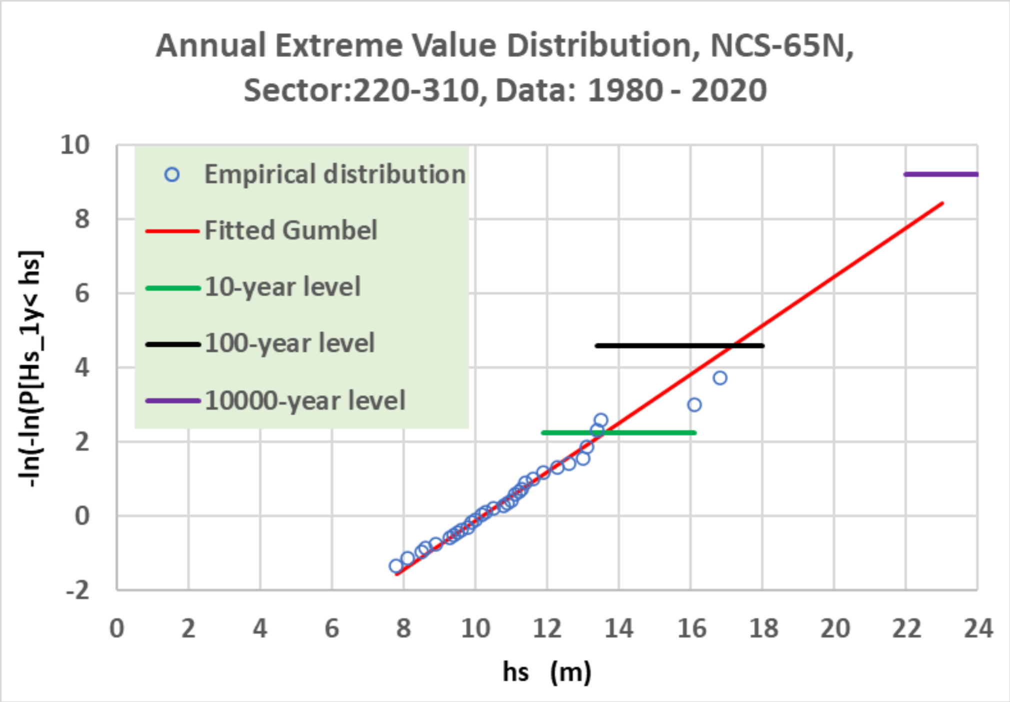


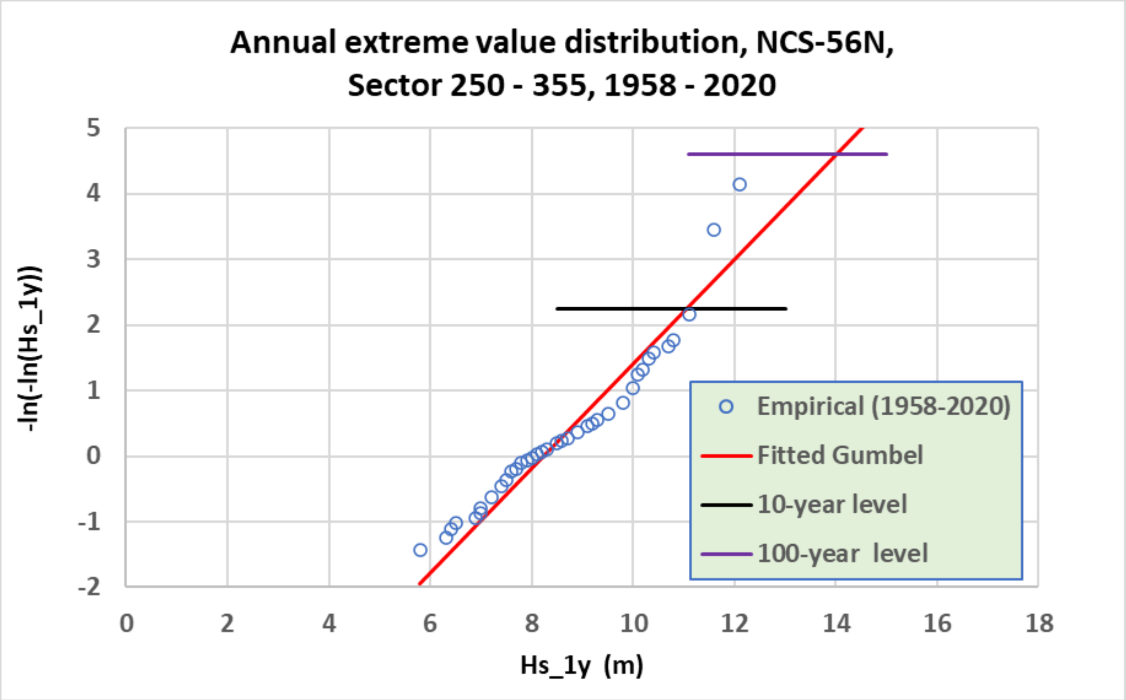




Target variable and its domain of attraction of block extremes

Annual extreme value distribution of significant wave height:





NCS: Norwegian Continental Shelf

Suggested criteria for selecting threshold, u

- Standard practice is to adopt as low a threshold as possible, providing a reasonable fit is obtained, Coles (2004), MacKay et al. (2011)

- Two possible approaches is suggested by Coles(2004):

  (1) Provided that the Pareto model seems to be an acceptable model for excesses of u_0,it should be a valid model also for a threshold $u > u_0$, and the mean residual life (mean av data above u) should be linear function of u.
  (2) Above u_0, the shape parameter, a (see page 4), of the GPD should be approximately constant.

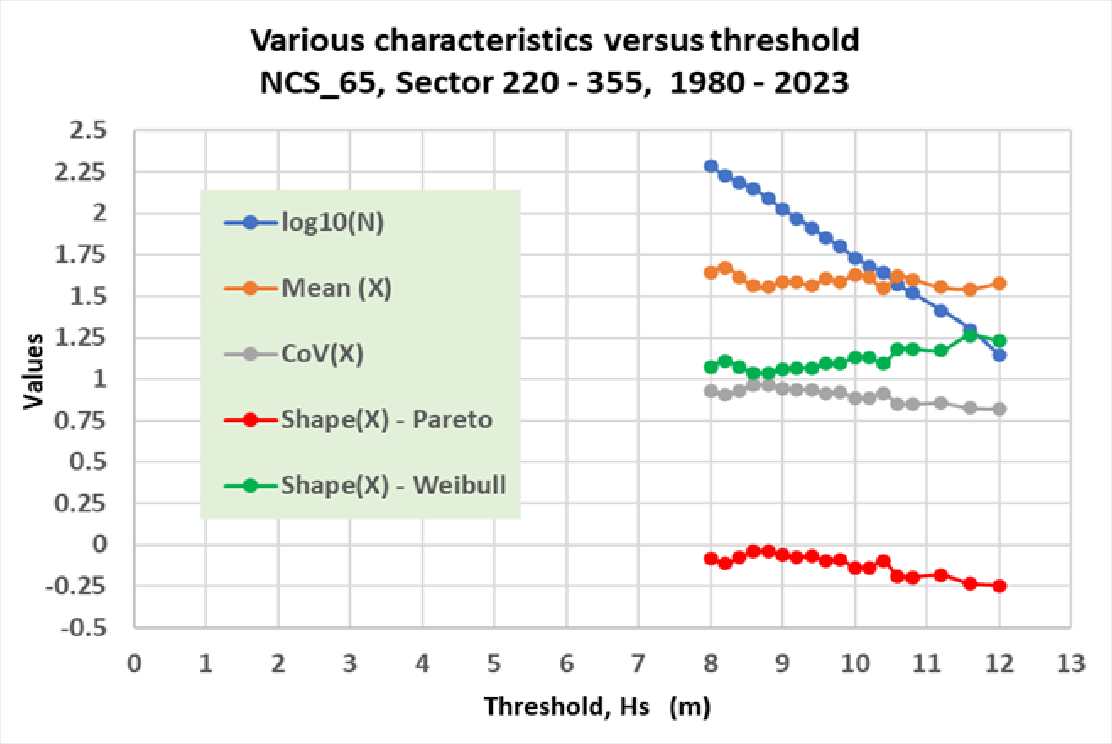
  In practise, the proposed approaches does not work very well regarding a robust prediction of extremes.

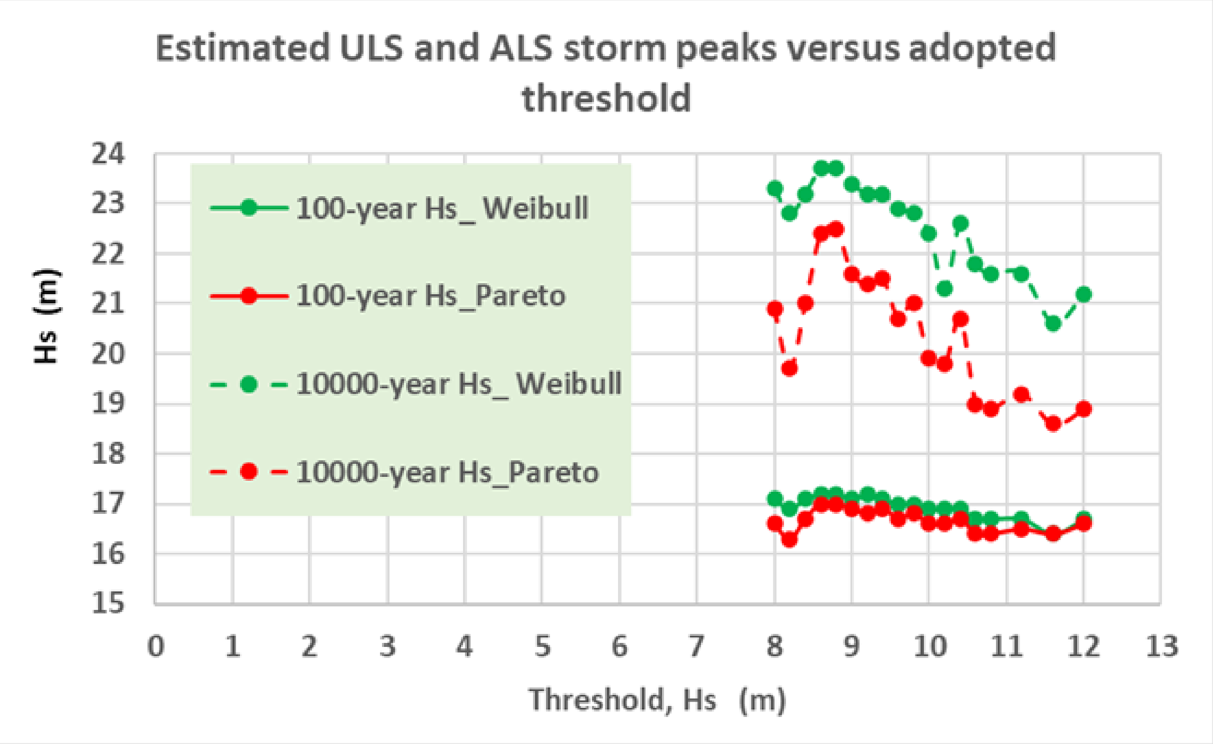
- Herein both GPD and 2-p Weibull is fitted to data above various thresholds using method of moments. For the involved sample sizes the adequacy of the moment estimates are just as good as estimates obtained by the maximum likehood method, MacKay et al.(2011)

Results: NCS_65N, Sector: 220-355, Data: 1980 - 2022

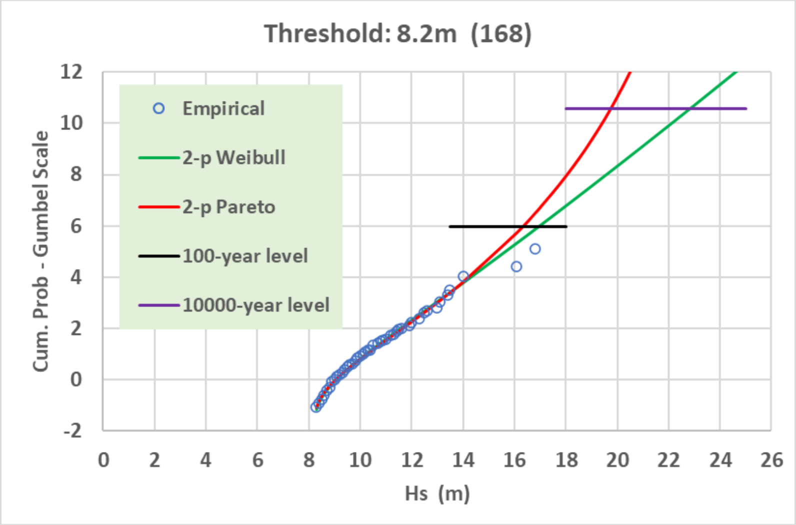
Distributions are fitted to data above threshold for many thresholds in the range 8-12 m.
For lowest threshold we have about 200 observations, for upper threshold close to 10.

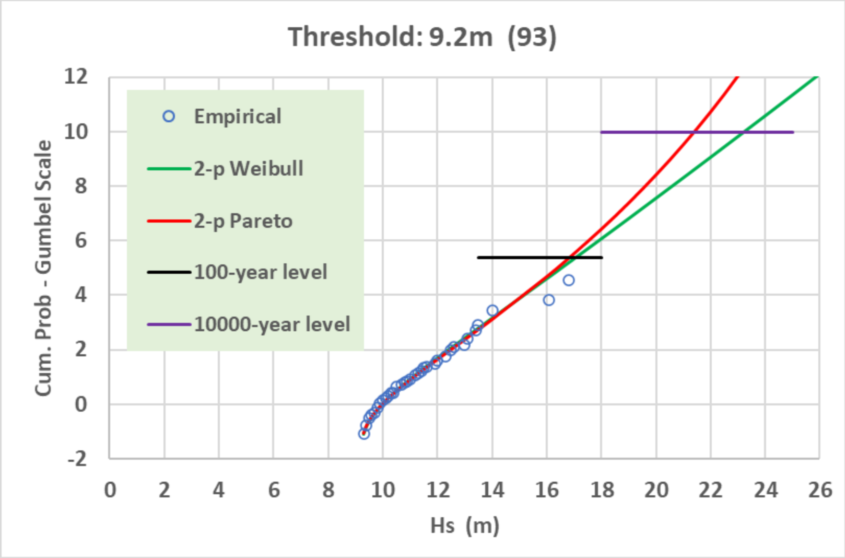
Results:

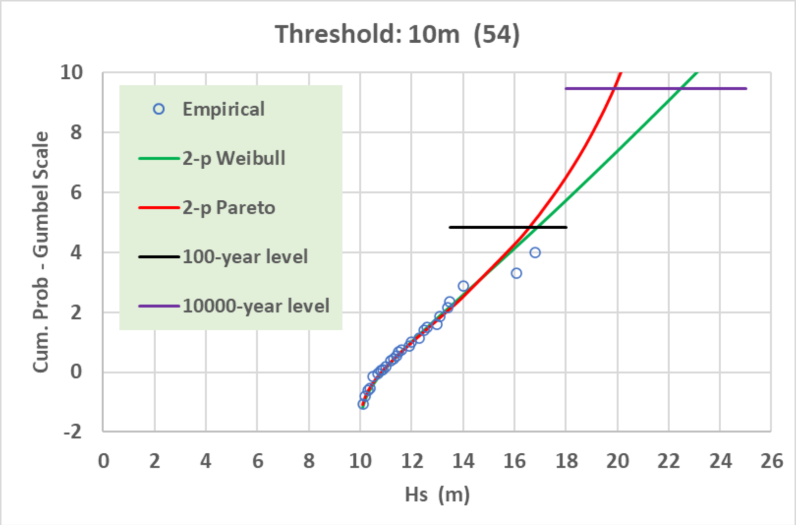


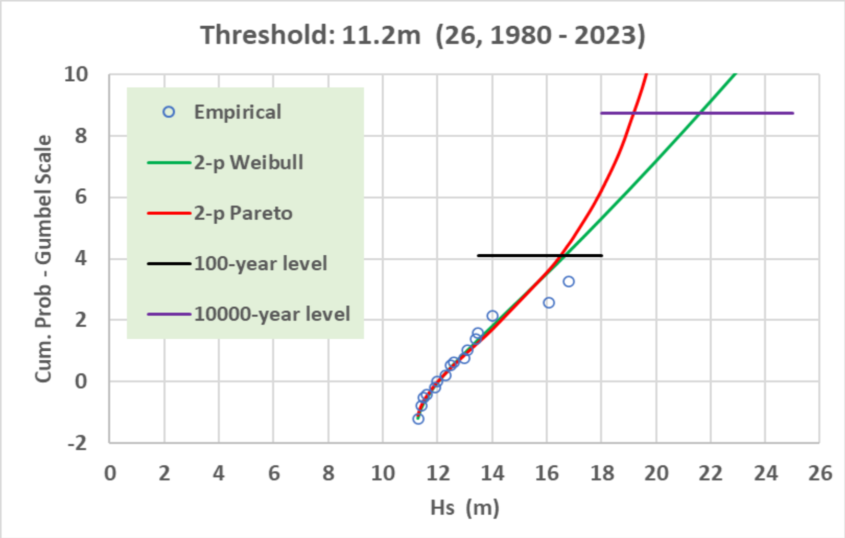


Fitted peak distributions for various thresholds, NCS_65N

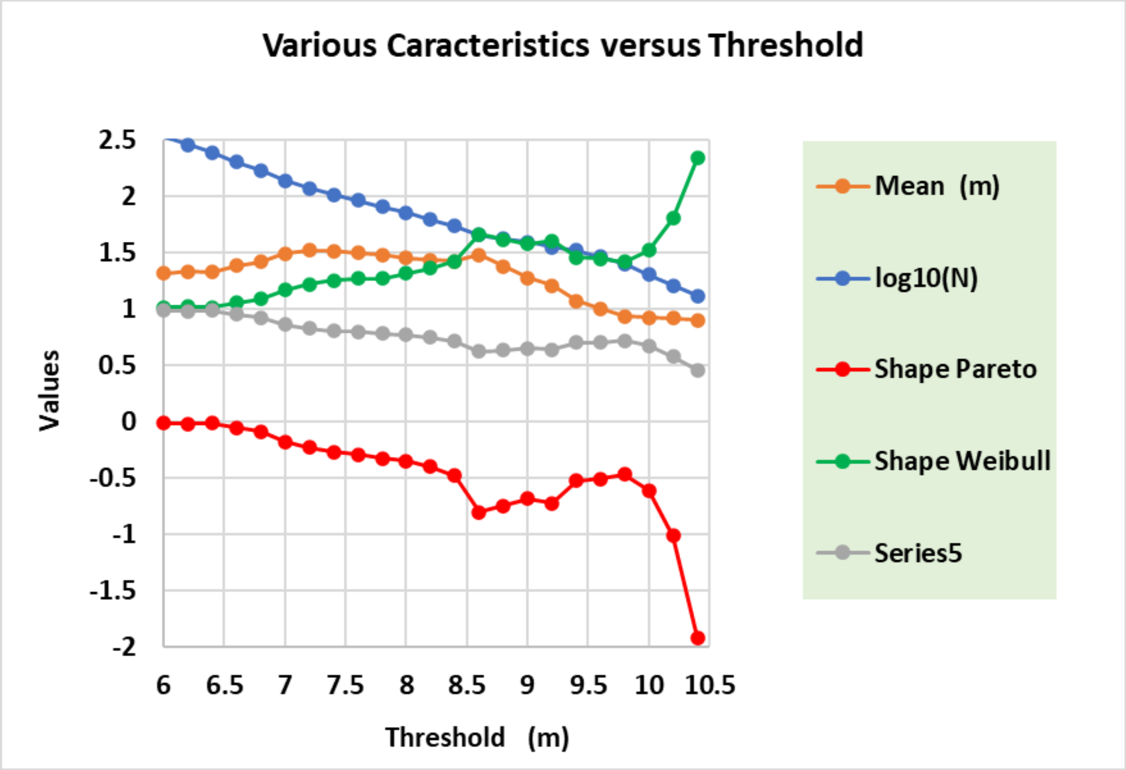


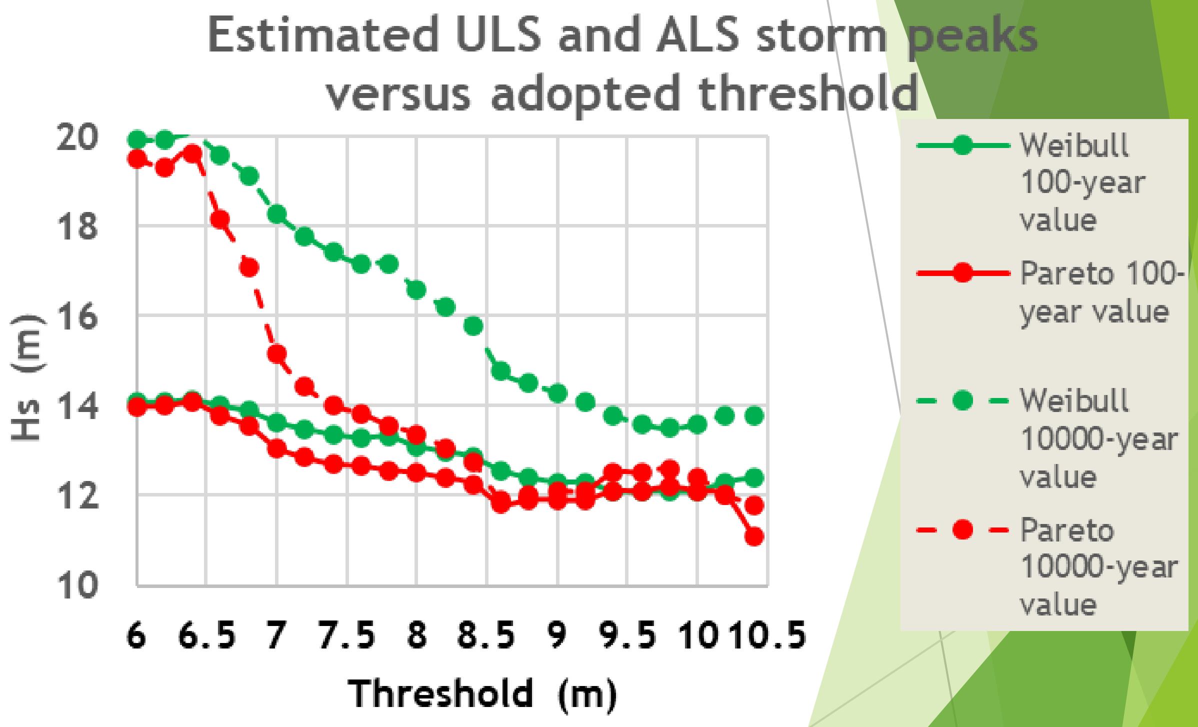




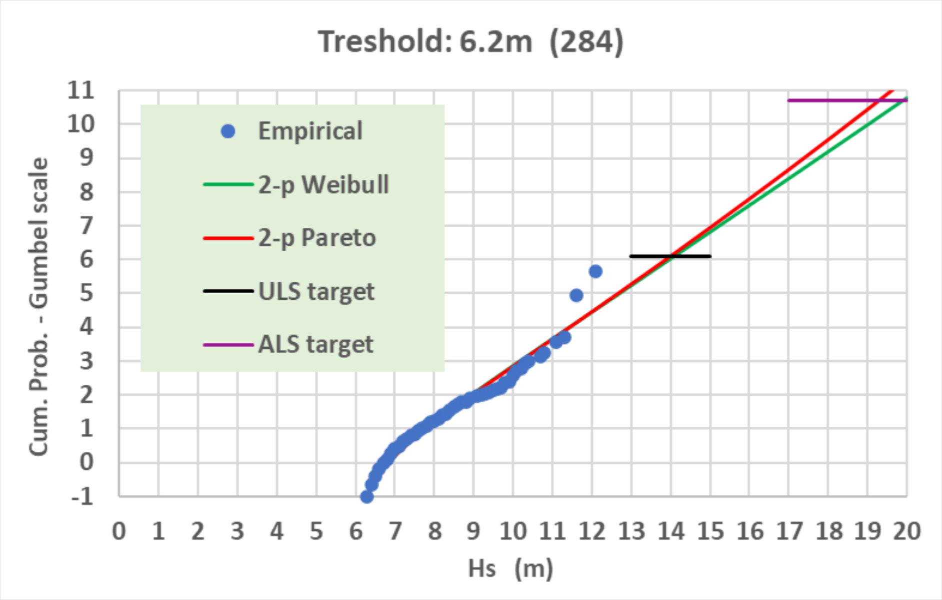


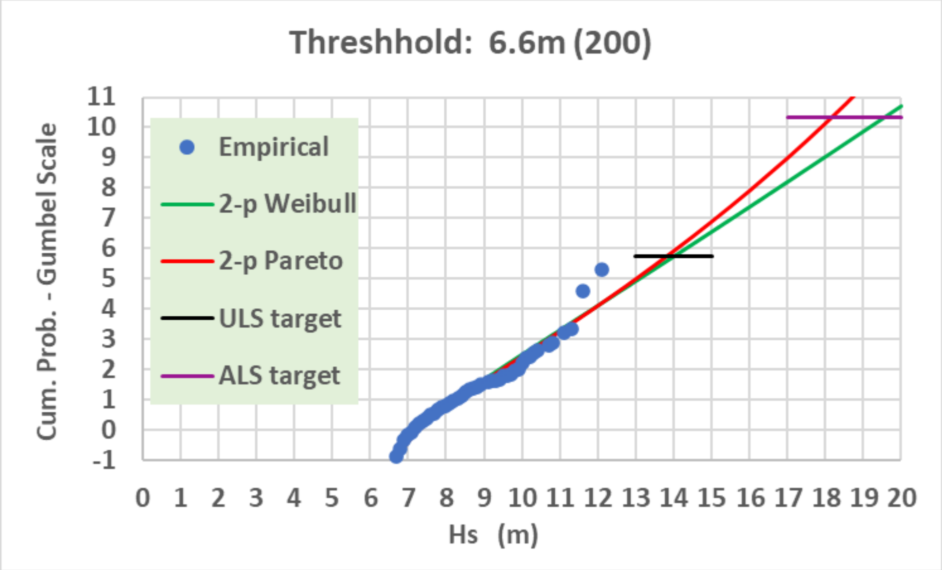
Results:NCS_56N, Sector: 250-355, Data: 1957 - 2021

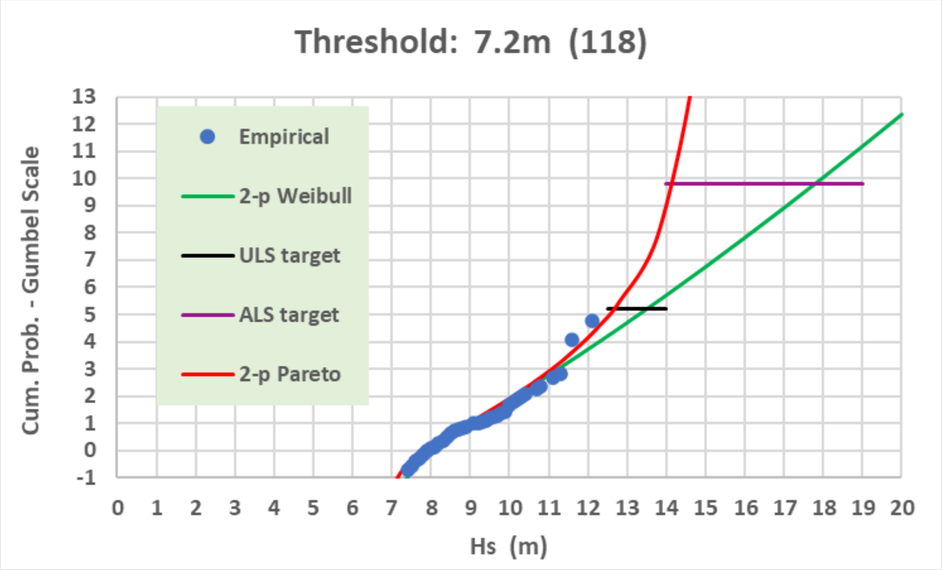


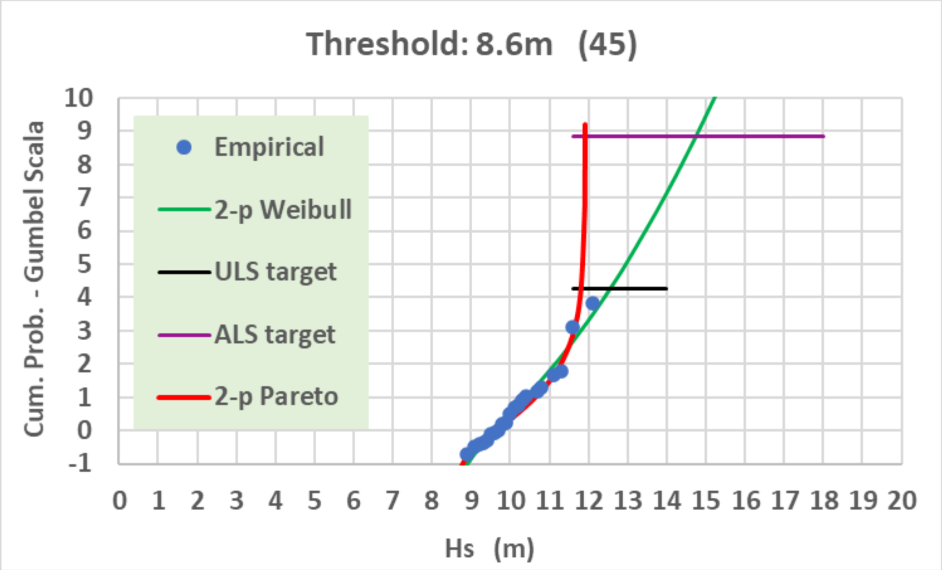


Fitted peak distributions for various thresholds









Calculation of asymptotic distribution if we know distribution function for an initial threshold

- For NCS_56N, the exceedance probability of an initial threshold of 7m is given by:

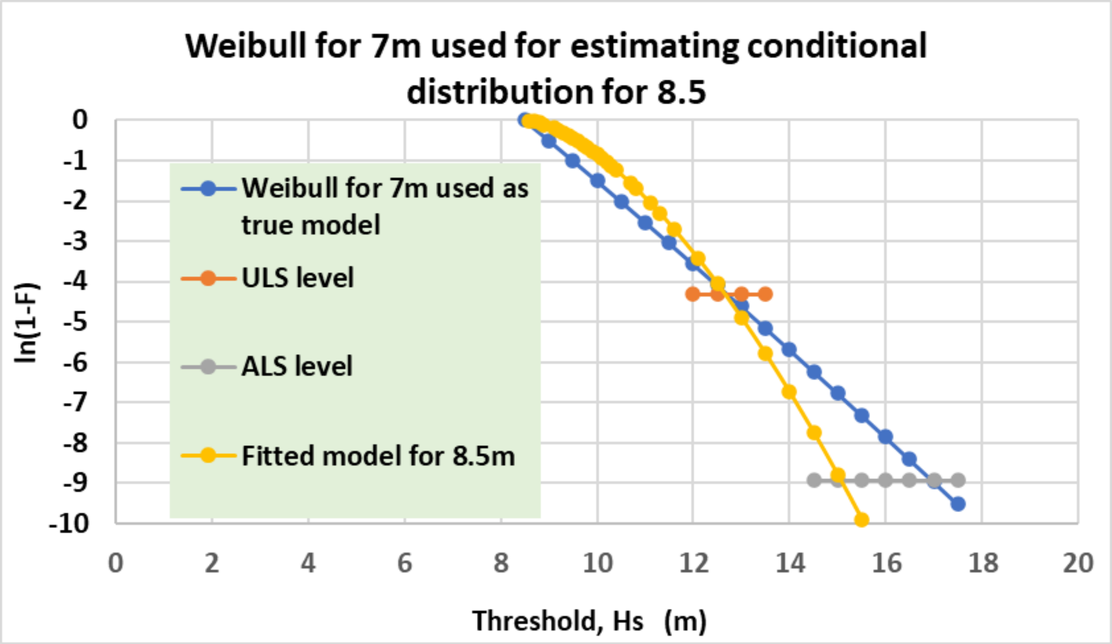
$$1 - F_X(x) = exp\left\{-\left(\frac{x}{\beta}\right)^{\gamma}\right\}; x > 0, \beta = 1.575, \gamma = 1.1691$$

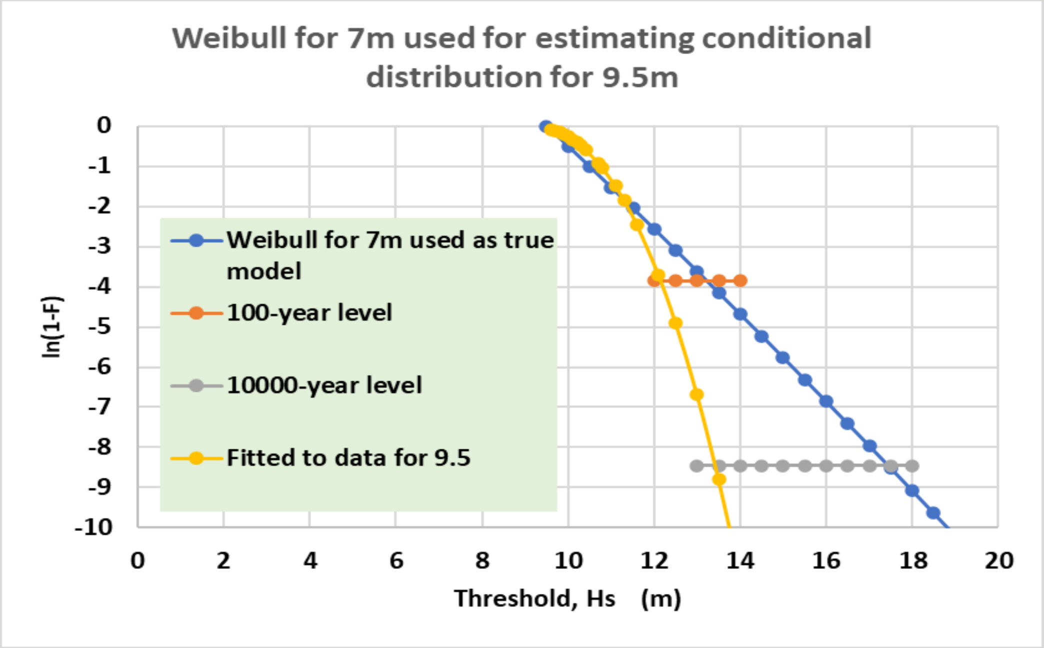
- A higher threshold of u = 8.4m is selected. We will use model above to calculate the conditional probability of exceeding u with a value y:

$$1 - F_Y(y) = \ P[X > u + y|\ X > u] = \frac{1 - F_X(u + y)}{1 - F_X(u)} = \frac{exp\left\{-\left(\frac{u + y}{\beta}\right)^{\gamma}\right\}}{exp\left\{-\left(\frac{u}{\beta}\right)^{\gamma}\right\}}$$

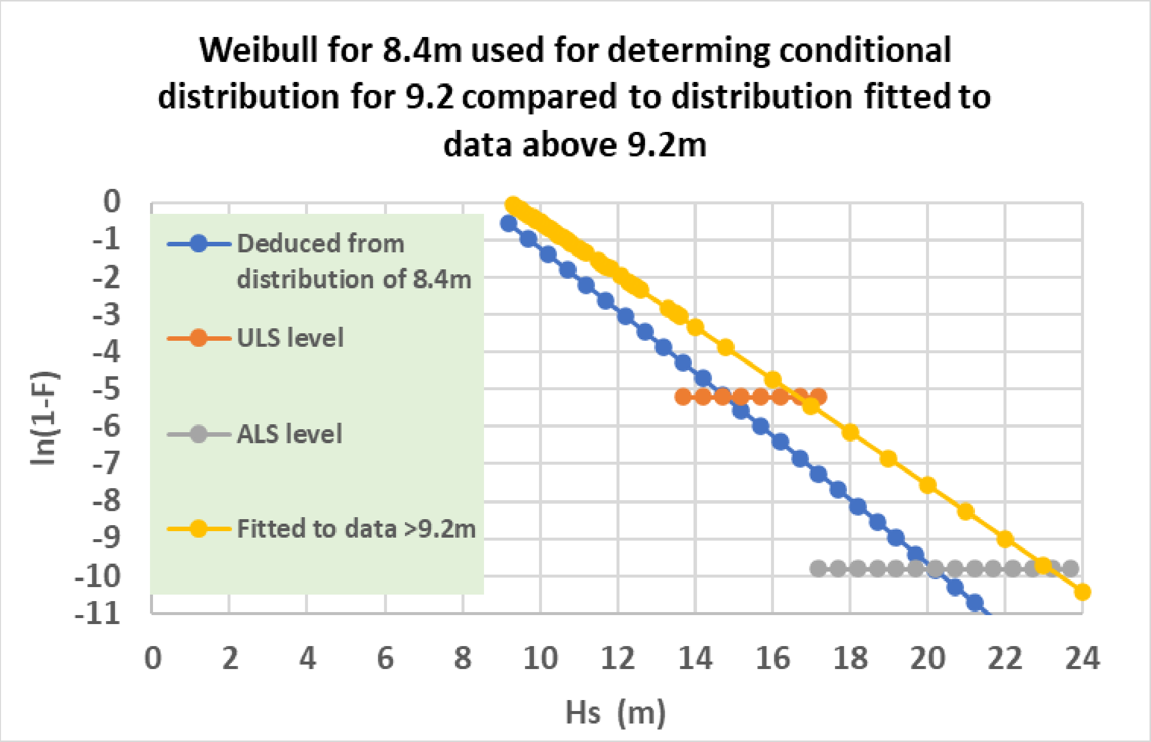
$$\Rightarrow \quad \boldsymbol{F_Y(y) = 1 - \exp\left\{-\left[\left(\frac{u+y}{\beta}\right)^{\gamma} - \left(\frac{u}{\beta}\right)^{\gamma}\right]\right\}}$$

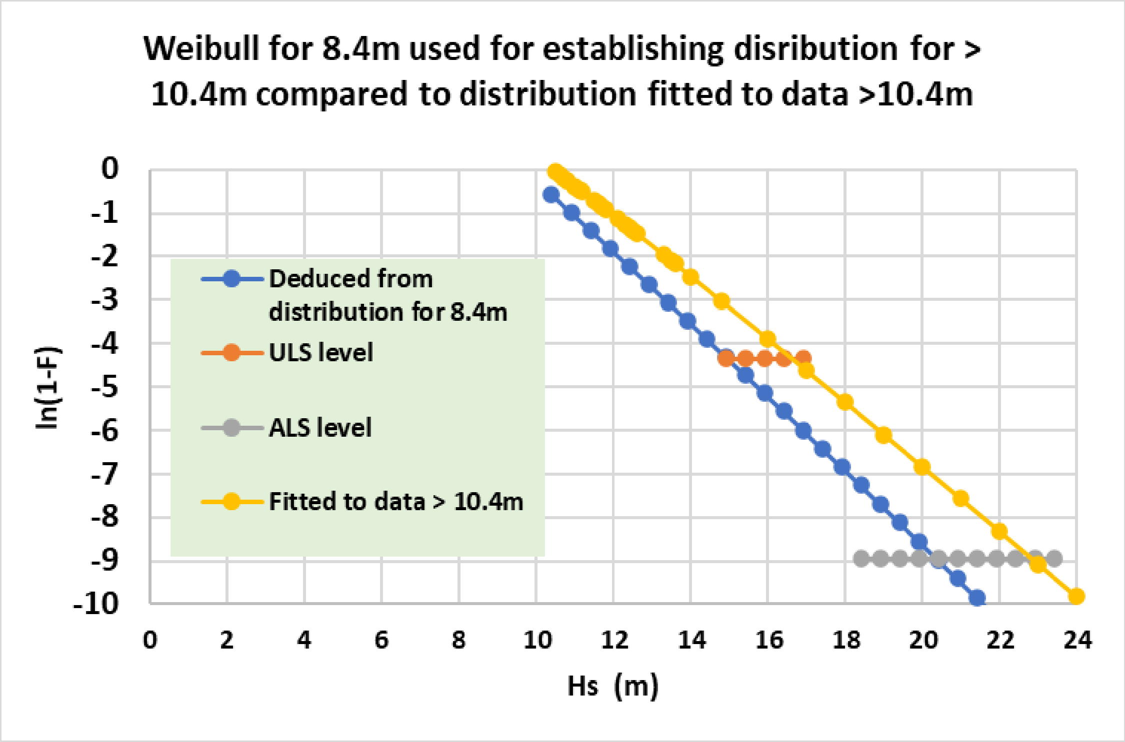
Results for NCS_56N





Results for NCS_61N





Summary

- If block extremes of actual variable have a domain of attraction corresponding to Type I Extreme Value Distribution, the exponential distribution is the asymptotic model for peaks over high threshold.
- Both 2-p Weibull and 2-p Pareto give more less the same distribution up to ULS level for the cases considered here. Selecting the treshold is (and will continue to be) the major difficultiy.
- For ALS the extremes differ - in particular as threshold level increasing. Can we believe in this? Determining a low lower bound based on a rather limited number of obs. seems questionable.
- Design recipes should give clear guidance regarding selection of threshold levels. What is a proper minimum acceptable number of data above selected threshold?
- In Norway 2-P Weibull is preferred. Regarding ALS level I do expect that 2-p Weibull will be conservative, too conservative? The 2-p Weibull model yields lower ALS estimates than the exponential distributions for Type I problems - may be Weibull is not too conservative.
- A British JIP have recommended a Generalized Pareto model. But they are accounting for statistical uncertainties in threshold level and distribution parameters. When merging inherent randomness and statistical unceratainties for the Generalized Pareto, results come rather close to the results obtained using Weibull. NB! I have not followed this JIP so I may have missed important points in their work.
- I prefer to continue with a semi-deterministic design recipe.

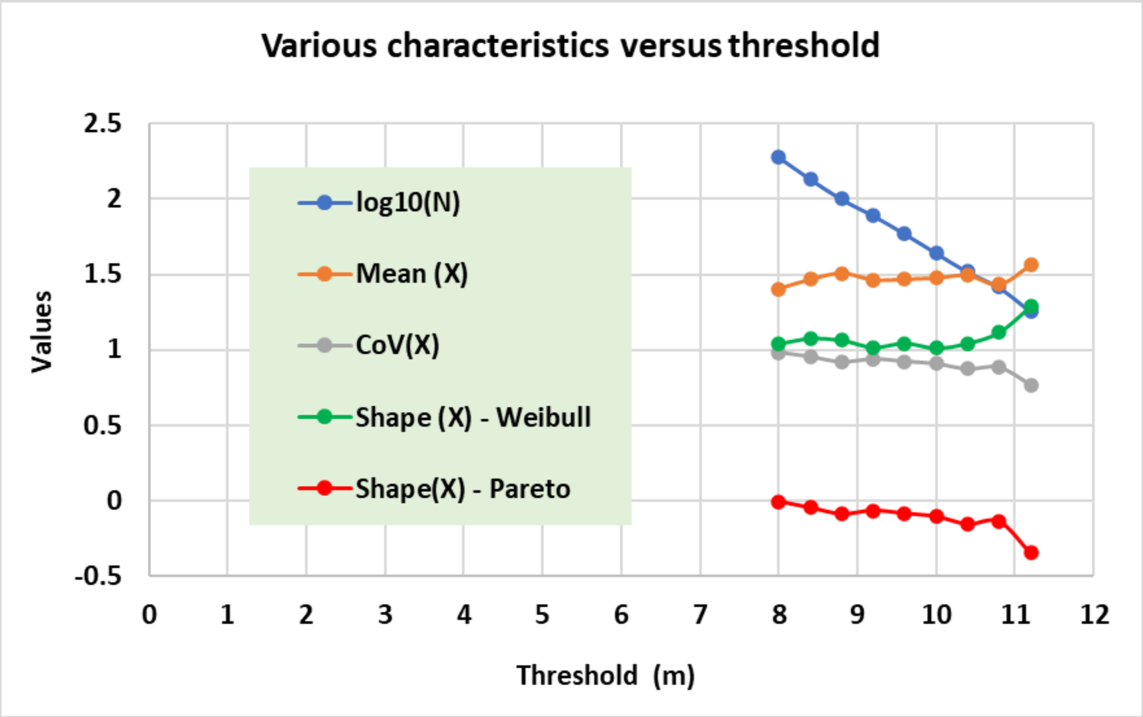
Litterature used

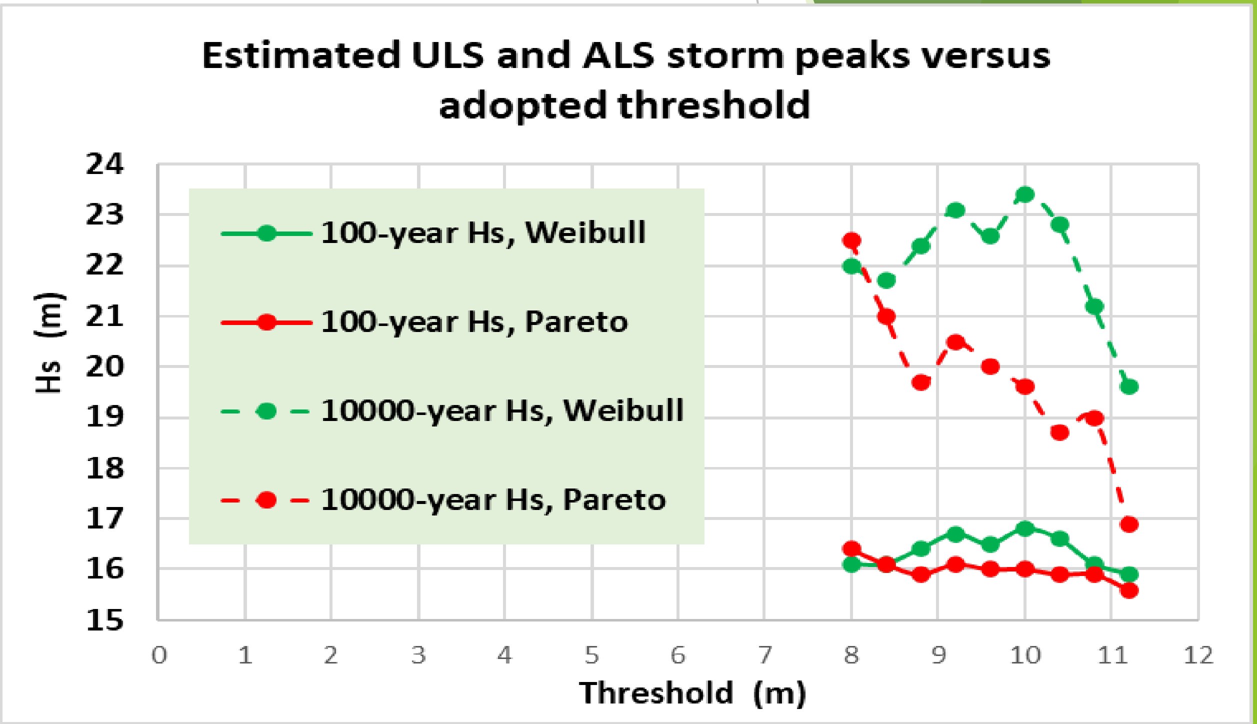
- Leadbetter, M.R. , Lindgren, G. and Rootzen, H.: «Extremes and Related Properties of Random Sequences and Series», Springer Verlag, New York, 1983
- Coles, S.: «An Introduction to Statistical Modelling of Extreme Values», Springer-Verlag, London, 3rd Printing 2004
- Reistad,M., Breivik, Ø.,Haakenstad, H., Aarnes, O.J., Furevik, B.R. and Bidlot, J.R.:»A high resolution hindcast ofwind and waves for the North Sea, the Norwegian Seaand the Barents Sea», Journal of Geophysical Research, Oceans, Vol. 116, 2011.
- Davison, A.C. and Smith, R.L.:»Models for Exceedances over High Thresholds»,Journal of the Royal Statistical Society, Series B (Methodological), Vol. 52, No. 3, 1990.
- Mackay, E.B.L., Challenor, P.G. and Bahaj, A.S.:»A comparison of estimators for the generalised Pareto distribution», Ocean Engineering, 38, 2011.

Appendix

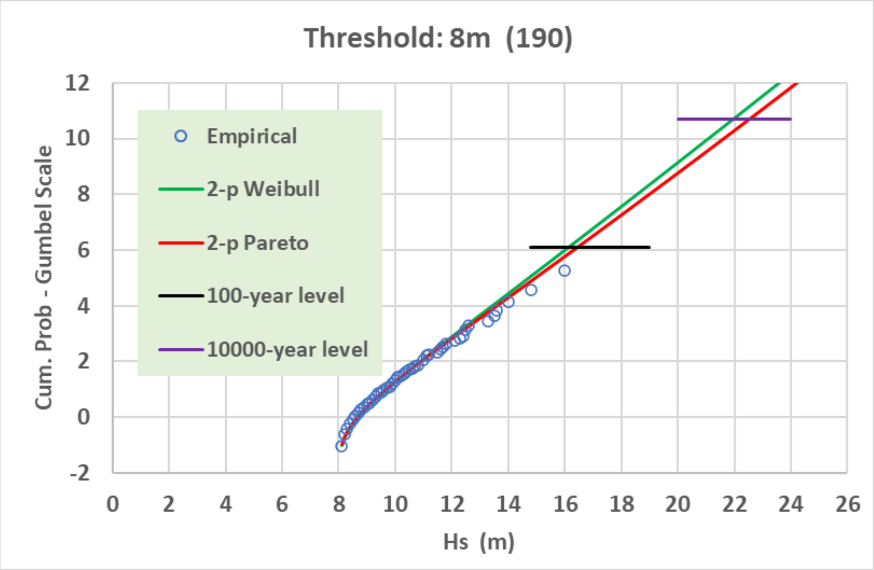
Results for 2 more positions on NCS

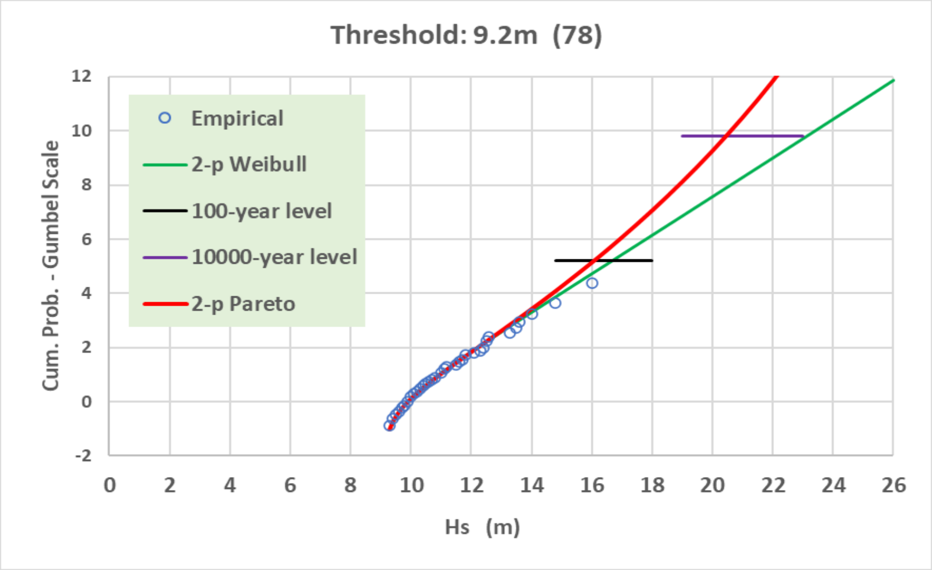
Results: NCS_ 61N, Sector: 150-360, Data: 1980 - 2022

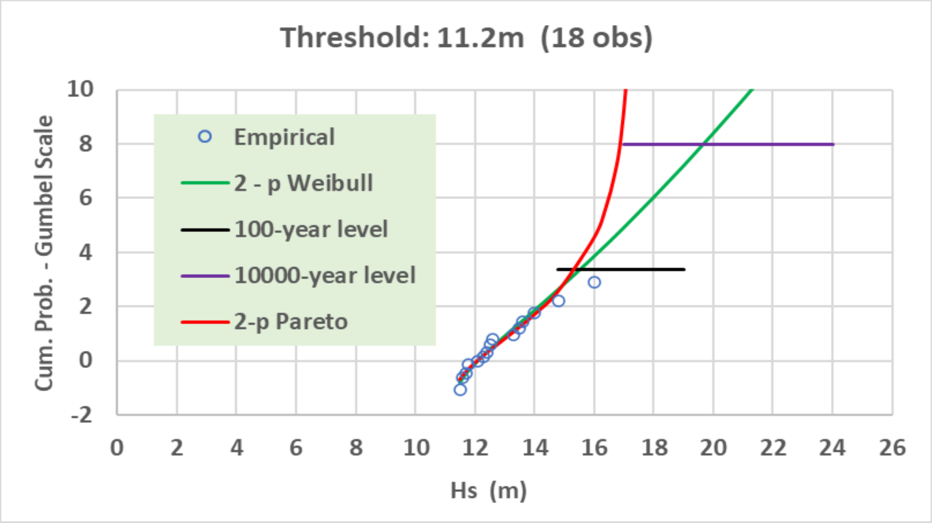




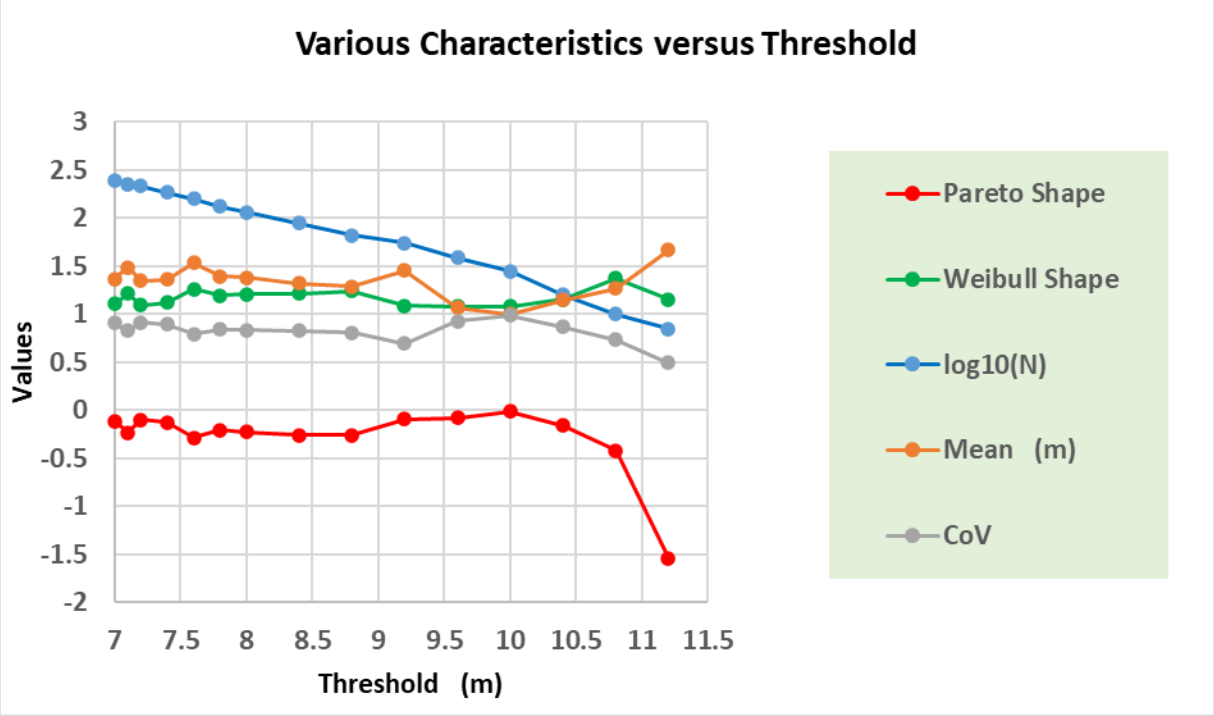
Fitted peak distributions for various thresholds

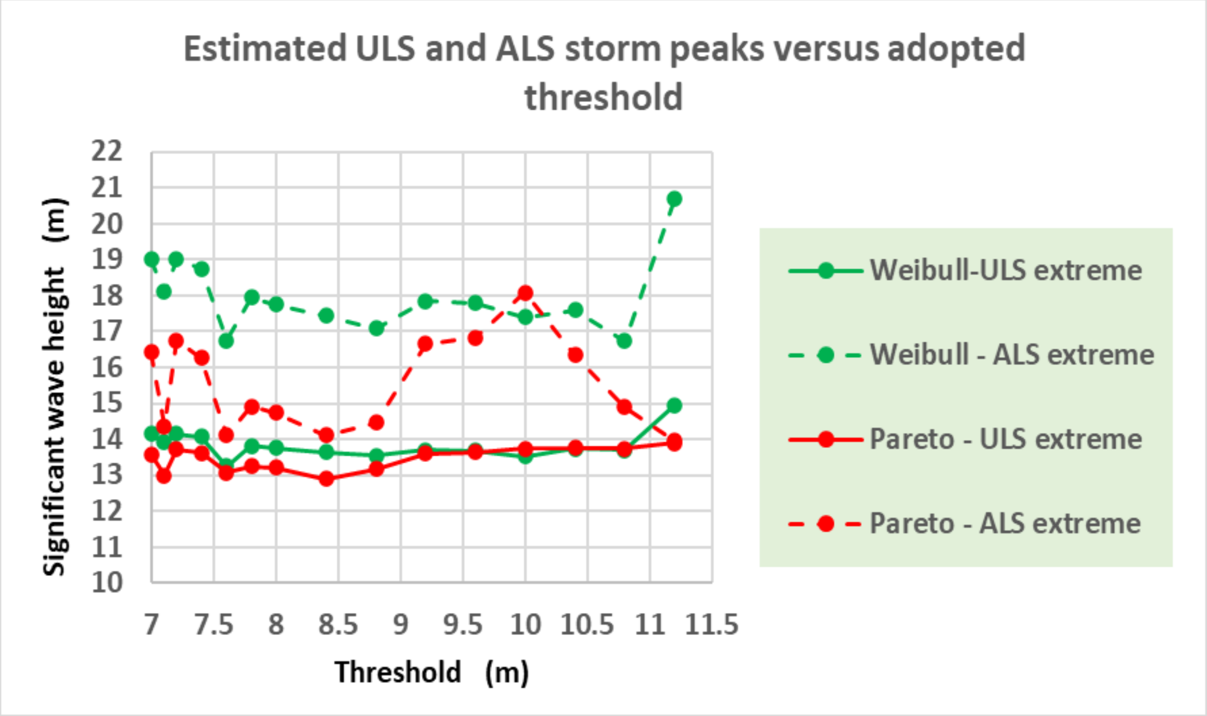






Results NCS_59N, Sector: 250 - 355, Data: 1957- 2021





Fitted peak distributions for various thresholds