Extreme Value Theory

Elisa Ragno

Assistant Professor
Department of Hydraulic Engineering – CiGT- TU Delft

Hydraulic Structures and Flood Risk Section

e.ragno@tudelft.nl

Lecture Topic

- Asymptotic model for Extreme Value Analysis
- Generalized Extreme Value Distribution
- Generalized Pareto Distribution
- Example application

TU Delft

Learning Objectives

At the end of the lecture, you will be able to:

- **LO1. Discuss** extreme events in statistical terms
- **LO2. Calculate** statistical characteristics of extremes
- **LO3. Identify** distribution functions for modelling extremes
- **LO4. Perform** Extreme Value Analysis in a case study

Asymptotic model for Extreme Value Analysis

e.ragno@tudelft.nl

TUDelft

Modelling Extremes – asymptotic model

- $\boldsymbol{X} = (\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$ sequence of independent and identically distributed (i.i.d.) random variables (e.g., daily discharge, daily traffic load)

- $\boldsymbol{F}(\boldsymbol{x})$ distribution function

We are interested in modeling the statistical behavior of

$$\boldsymbol{M_n} = \mathbf{max}(\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$$

- $\boldsymbol{M_n}$ maximum of the process

Modelling Extremes – asymptotic model

- $\boldsymbol{X} = (\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$ sequence of independent and identically distributed (i.i.d.) random variables (e.g., daily discharge, daily traffic load)

- $\boldsymbol{F}(\boldsymbol{x})$ distribution function

We are interested in modeling the statistical behavior of

$$\boldsymbol{M_n} = \mathbf{max}(\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$$

- $\boldsymbol{M_n}$ maximum of the process

The theoretical distribution of $\boldsymbol{M_n}$ ($\boldsymbol{F}$ *known*) is $\boldsymbol{F}(\boldsymbol{x})^{\boldsymbol{n}}$

TU Delft

Modelling Extremes – asymptotic model

- $\boldsymbol{X} = (\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$ sequence of independent and identically distributed (i.i.d.) random variables (e.g., daily discharge, daily traffic load)

- $\boldsymbol{F}(\boldsymbol{x})$ distribution function

We are interested in modeling the statistical behavior of

$$\boldsymbol{M_n} = \mathbf{max}(\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$$

- $\boldsymbol{M_n}$ maximum of the process

The theoretical distribution of $\boldsymbol{M_n}$ ($\boldsymbol{F}$ <u>*known*</u>) is

$$\begin{aligned} \boldsymbol{Pr}\,(\boldsymbol{M_n} \leq \boldsymbol{x}) \quad &= Pr(\max(X) \leq x) \\ &= Pr\,(X_1 \leq x, \dots, X_n \leq x) \\ &= Pr(X_1 \leq x) \cdot \ \dots \cdot Pr(X_n \leq x) \ = \boldsymbol{F}(\boldsymbol{x})^{\boldsymbol{n}} \end{aligned}$$

Modelling Extremes – asymptotic model

- $\boldsymbol{X} = (\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$ sequence of independent and identically distributed (i.i.d.) random variables (e.g., daily discharge, daily traffic load)

- $\boldsymbol{F}(\boldsymbol{x})$ distribution function

We are interested in modeling the statistical behavior of

$$\boldsymbol{M_n} = \mathbf{max}(\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$$

- $\boldsymbol{M_n}$ maximum of the process

The theoretical distribution of $\boldsymbol{M_n}$ ($\boldsymbol{F}$ <u>*known*</u>) is

$$\begin{aligned} \boldsymbol{Pr}\,(\boldsymbol{M_n} \leq \boldsymbol{x}) \quad &= Pr(\max(X) \leq x) \\ &= Pr\,(X_1 \leq x, \dots, X_n \leq x) \\ &= Pr(X_1 \leq x) \cdot \ \dots \cdot Pr(X_n \leq x) \ = \boldsymbol{F}(\boldsymbol{x})^{\boldsymbol{n}} \end{aligned}$$

Example: Estimate $\boldsymbol{F(x)^n}$

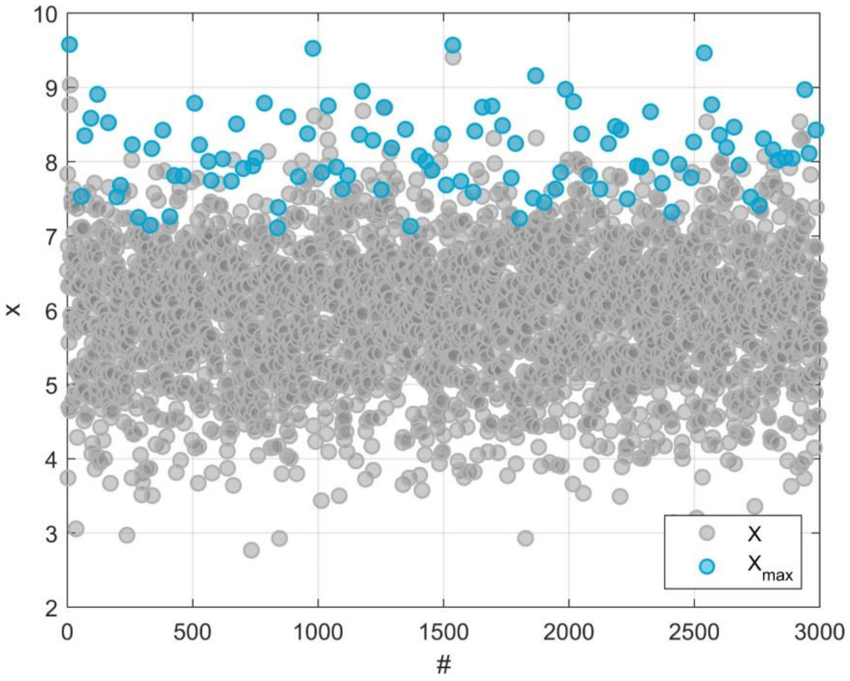
- Random generate 100 samples of length N = 30 from N~(6,1)
- Store the maximum for each sample (x_{max}) – *block maximum*

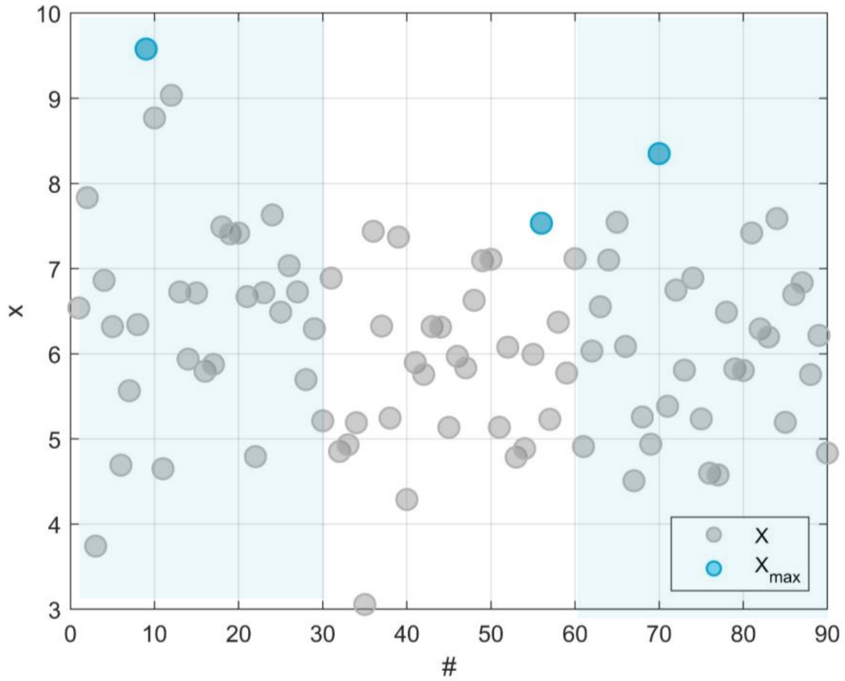
Example: Estimate $F(x)^n$

- Random generate 100 samples of length N = 30 from N~(6,1)
- Store the maximum for each sample (x_{max}) – *block maximum*

```
mu    = 6
Sigma = 1
N     = 30

for i = 1:100
        x(:,i)     = normrnd(mu,sigma,N,1)
        x_max(i)  = max(x(:,i))
end
```





Example: Estimate $F(x)^n$

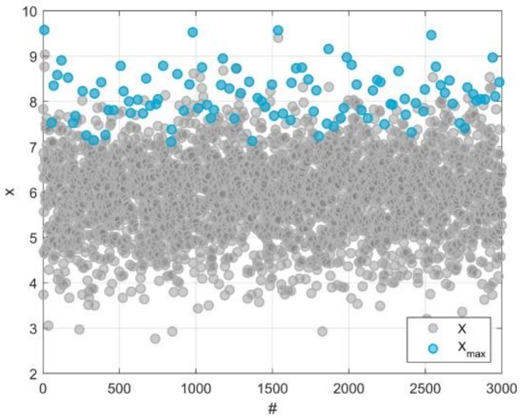
- Random generate 100 samples of length N = 30 from N~(6,1)
- Store the maximum for each sample (x_{max}) – *block maximum*

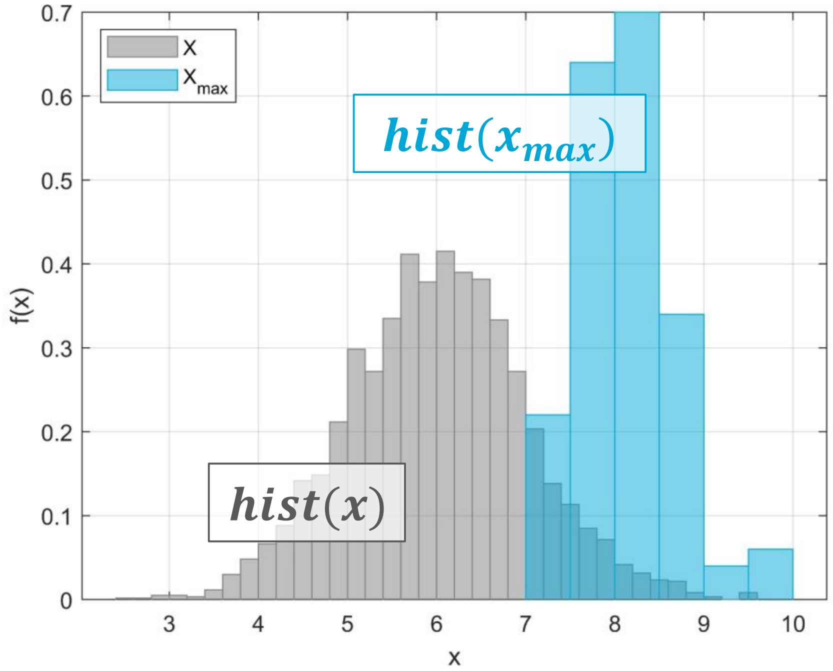
- Plot the distribution of the maxima

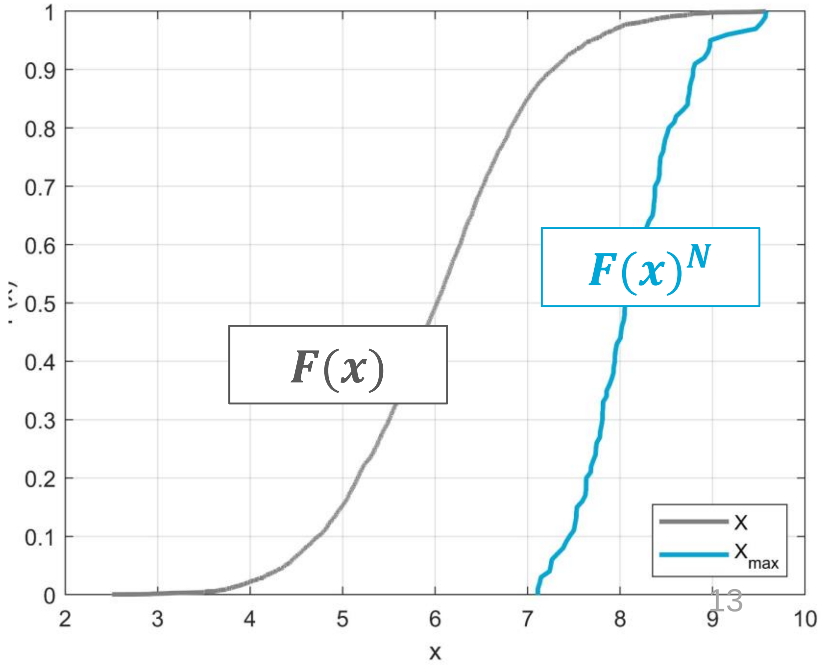
```
mu    = 6
Sigma = 1
N     = 30

for i = 1:100
        x(:,i)    = normrnd(mu,sigma,N,1)
        x_max(i)  = max(x(:,i))
end

plot histogram(x_max)
plot x_max empirical cdf
```







TU Delft

Generalized Extreme Value Distribution

e.ragno@tudelft.nl

Generalized Extreme Value Distribution

- We are interested in modeling the statistical behavior of the maximum of the sequence $X_1, \dots, X_n$ of independent and identically distributed (i.i.d.) random variables, $M_n = \max(X_1, \dots, X_n)$, where n is the number of observations in a block (e.g., annual maximum).

- We can prove that, for **large $\mathbf{n}$**

$$\mathbf{Pr(M_n \leq x) \rightarrow G(x)}$$

- Where G belongs to the **Generalized Extreme Value family of distributions** *regardless of the distribution of X*

Generalized Extreme value distribution

GEV for block maxima

$$\boldsymbol{G(x) = exp\left\{-\left[1 + \xi \cdot \left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\},} \qquad \left(1 + \xi \cdot \frac{x-\mu}{\sigma}\right) > 0$$

parameters: location ($-\infty < \boldsymbol{\mu} < \infty$), scale ($\boldsymbol{\sigma} > 0$), and shape ($-\infty < \boldsymbol{\xi} < \infty$)

TU Delft

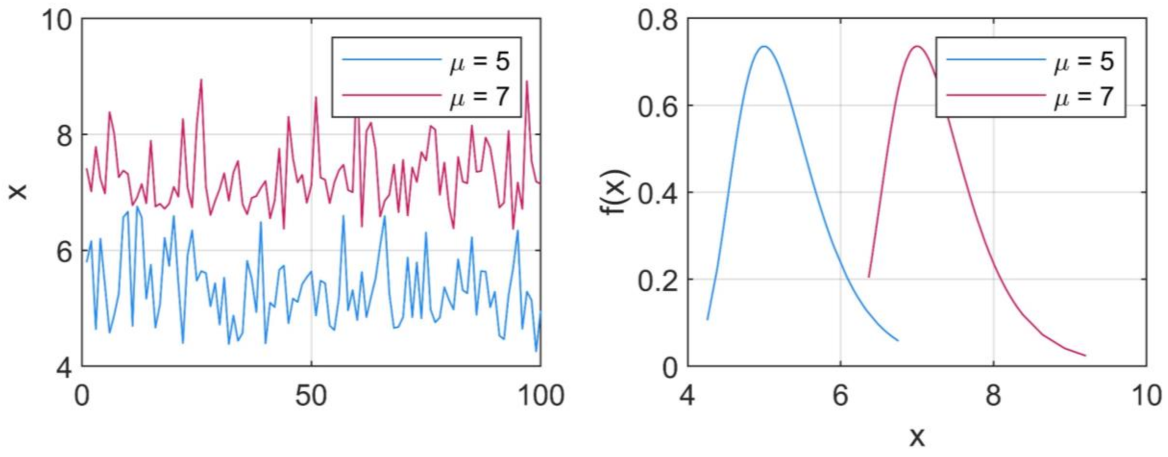
Generalized Extreme value distribution

GEV for block maxima

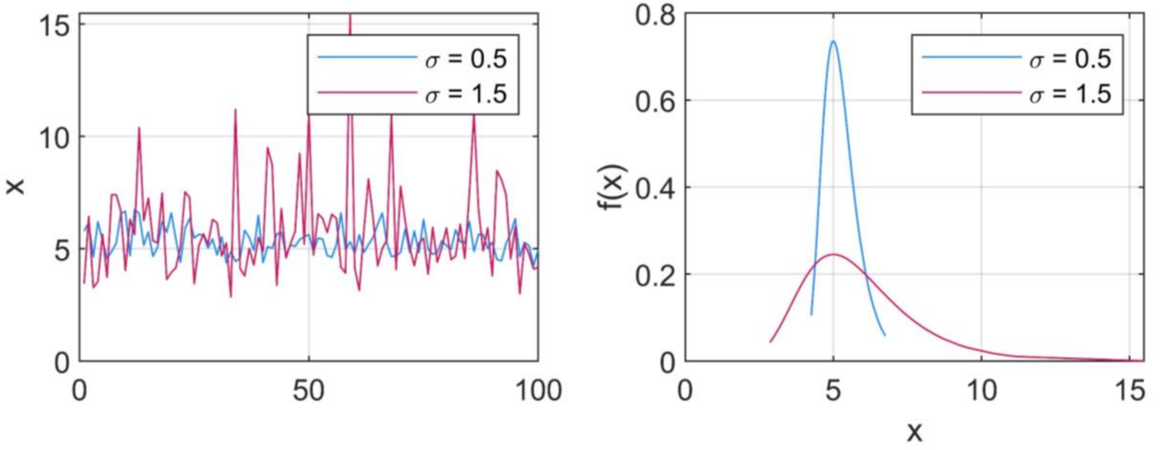
$$\boldsymbol{G(x) = exp\left\{-\left[1 + \xi \cdot \left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}},\qquad \left(1 + \xi \cdot \frac{x-\mu}{\sigma}\right) > 0$$

parameters: location ($-\infty < \boldsymbol{\mu} < \infty$), scale ($\boldsymbol{\sigma} > 0$), and shape ($-\infty < \boldsymbol{\xi} < \infty$)

Location - $\boldsymbol{\mu}$



Scale - $\boldsymbol{\sigma}$



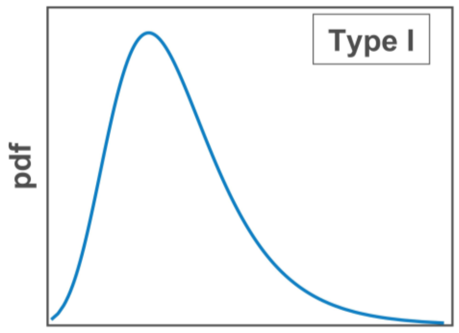
Generalized Extreme value distribution

GEV for block maxima

$$G(x) = exp\left\{-\left[1 + \xi \cdot \left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \qquad \left(1 + \xi \cdot \frac{x-\mu}{\sigma}\right) > 0$$

parameters: location ($-\infty < \mu < \infty$), scale ($\sigma > 0$), and shape ($-\infty < \xi < \infty$)

Gumbel

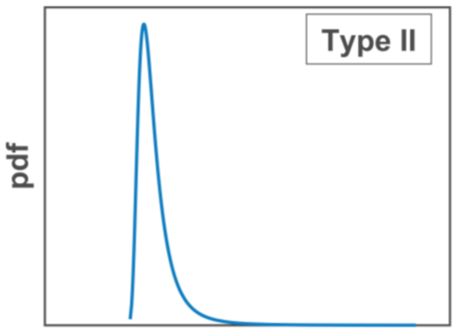


x

$\xi \to 0$

Exponential decay

Fréchet

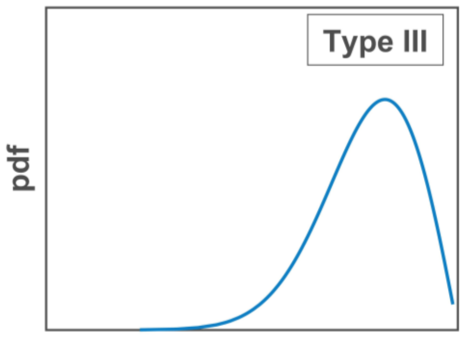


x

$\xi > 0$

Polynomial decay

Reverse Weibull



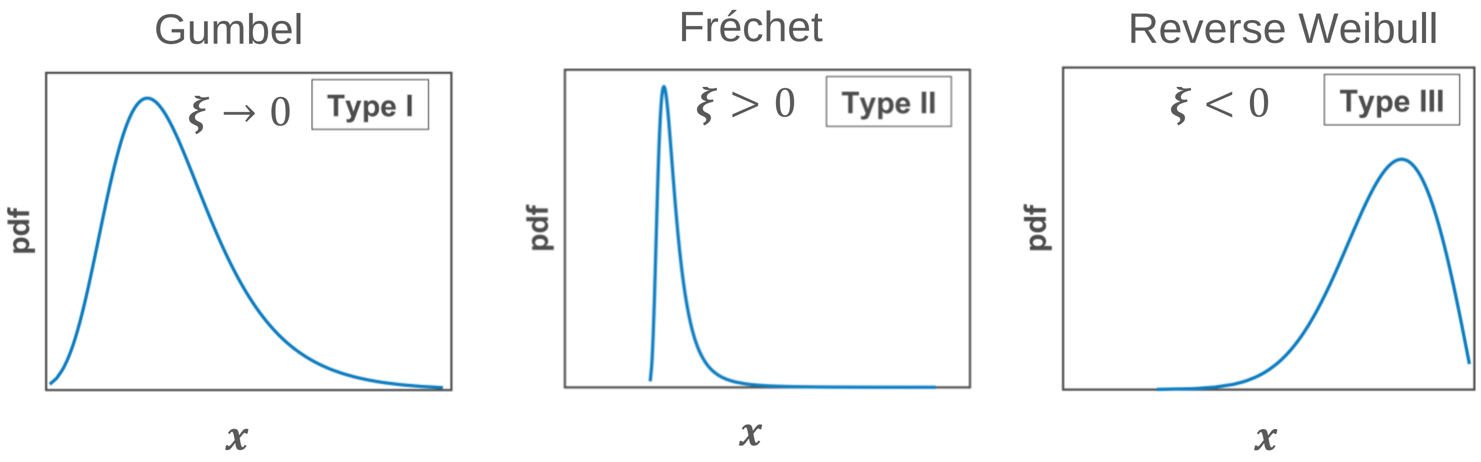
x

$\xi < 0$

Upper bound ($\mu - {}^{\sigma}/_{\xi}$)

TU Delft

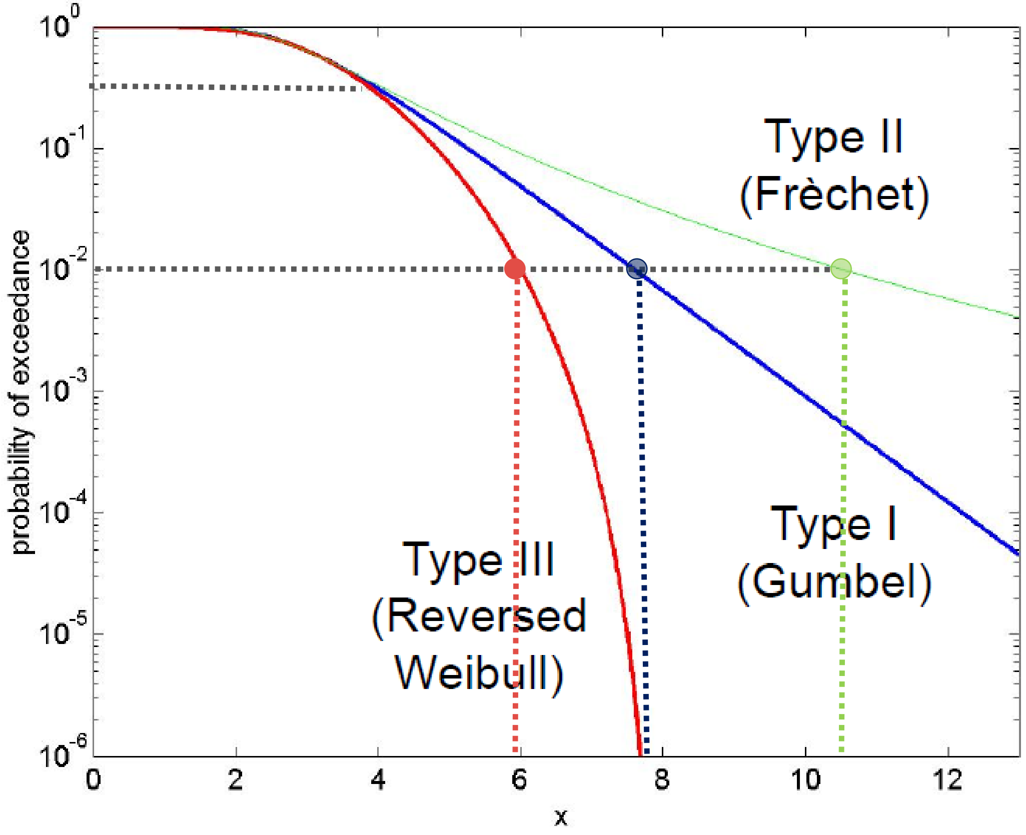
Exceedance Probability - GEV



The three types behave differently at the tails

This is even more evident for very small probability of exceedance, e.g. event with a very low probability of occurrence

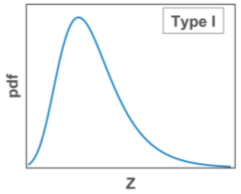
- Gumbel – "Light tail"
- Frechet – "Heavy tail"
- Weibull - Bounded

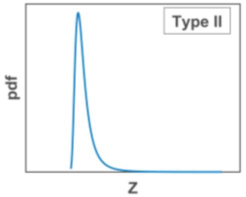


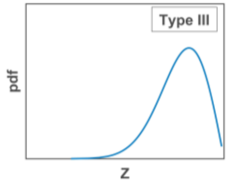
TU Delft

GEV – Domain of Attraction

Parent Distribution	Asymptotic type maximum (D.A.)
Normal, Exponential, Gamma, Lognormal, Weibull	Gumbel
Pareto, Cauchy, Student-t (fat tail)	Fréchet
Uniform, Beta (short tail)	Reverse Weibull



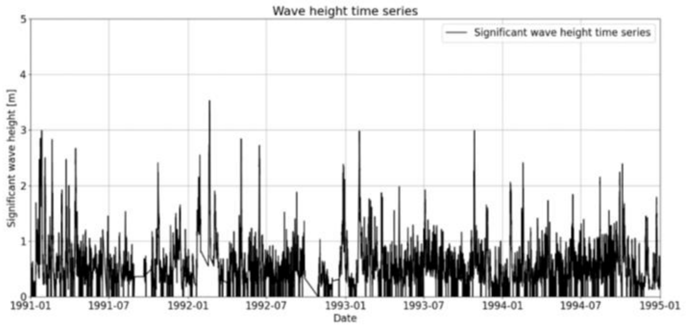




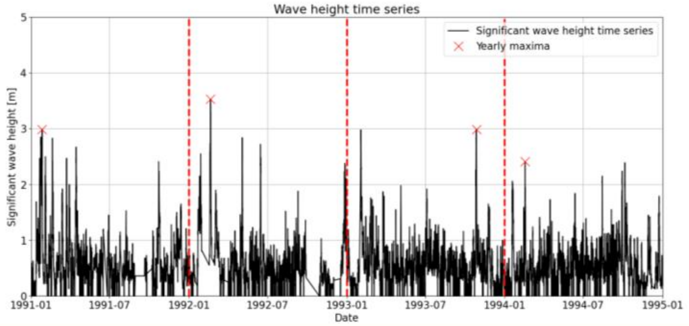
TU Delft

Example Case: how big is the 100-year event?

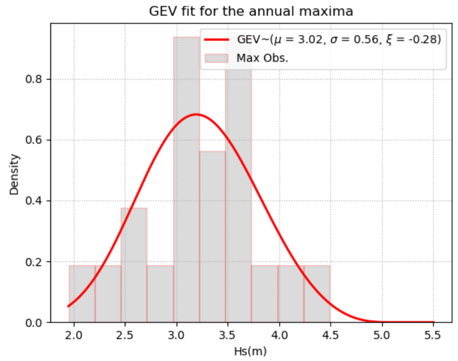
```
>> read observations
```

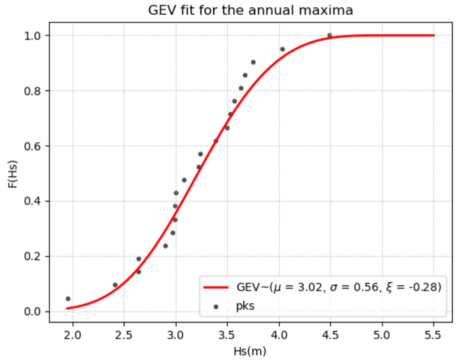



```
>> for each year i
       OBSmax(i) = max(observations in year i)
   end
```




```
>> fit a GEV on OBSmax
```

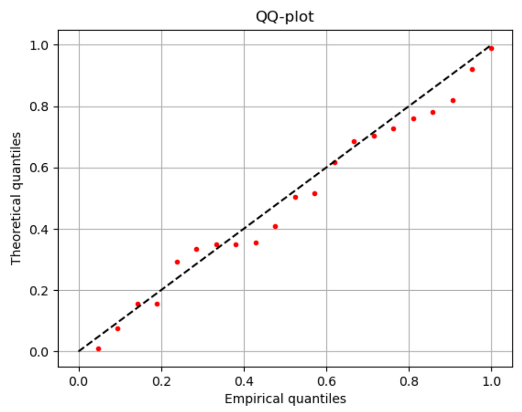




TU Delft

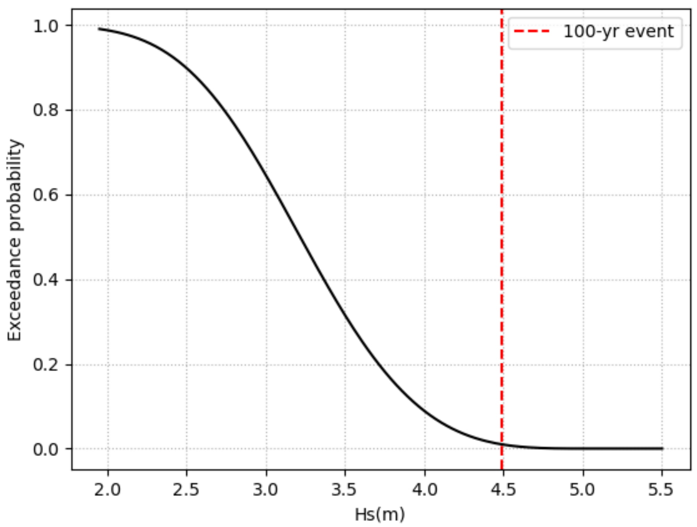
Example Case: how big is the 100-year event?

>> **check** the fit (e.g. QQ-plot)



>> **design event:** inverse of GEV for yearly exceedance probability p_{ex} = 1/100 = 0.01

$$z_p = G^{-1}(1 - p_{ex}) = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-log(1 - p_{ex})\}^{-\xi}\right], & for\ \xi \neq 0 \\ \mu - \sigma log\{-log(1 - p_{ex})\}, & for\ \xi = 0 \end{cases}$$



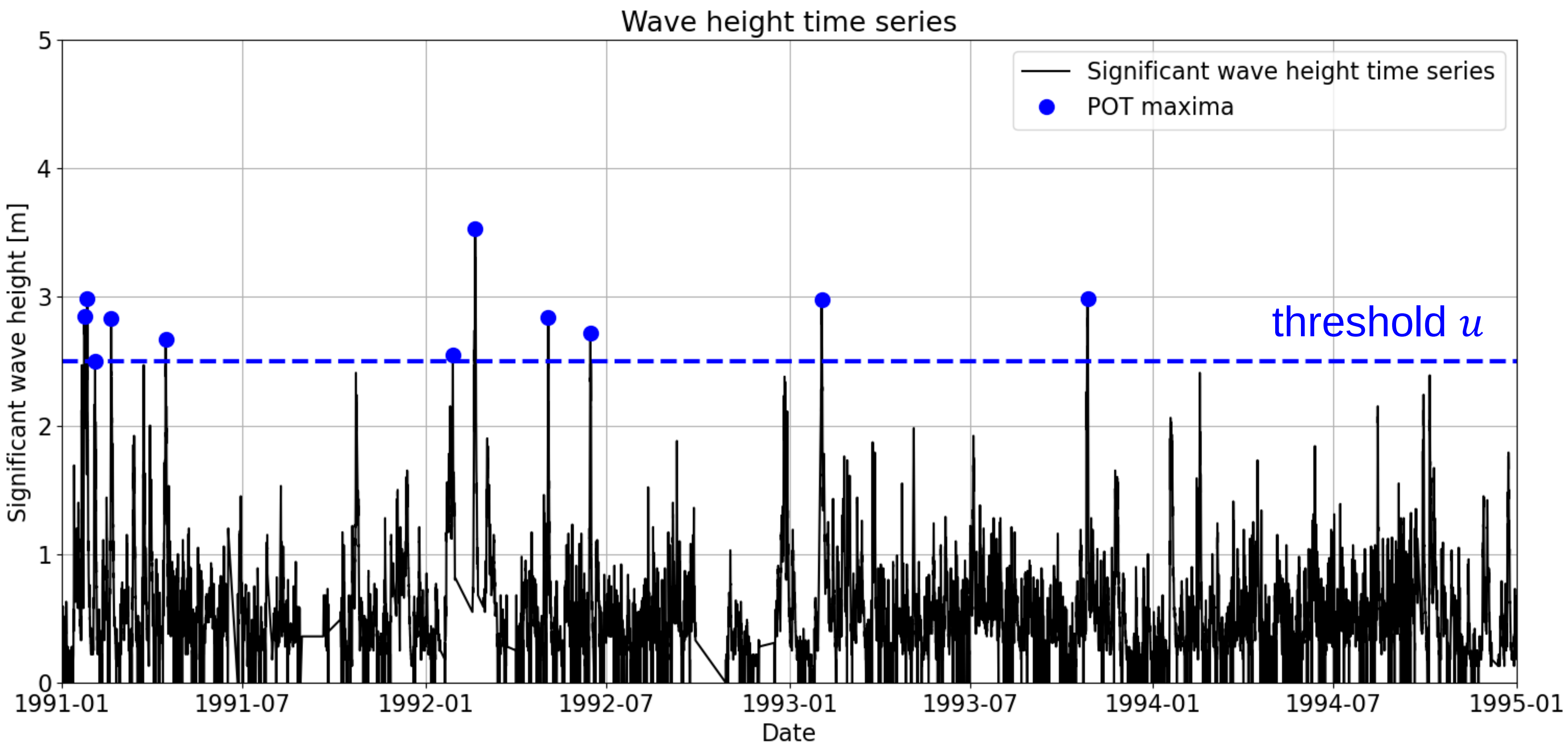
TU Delft

Generalized Pareto Distribution

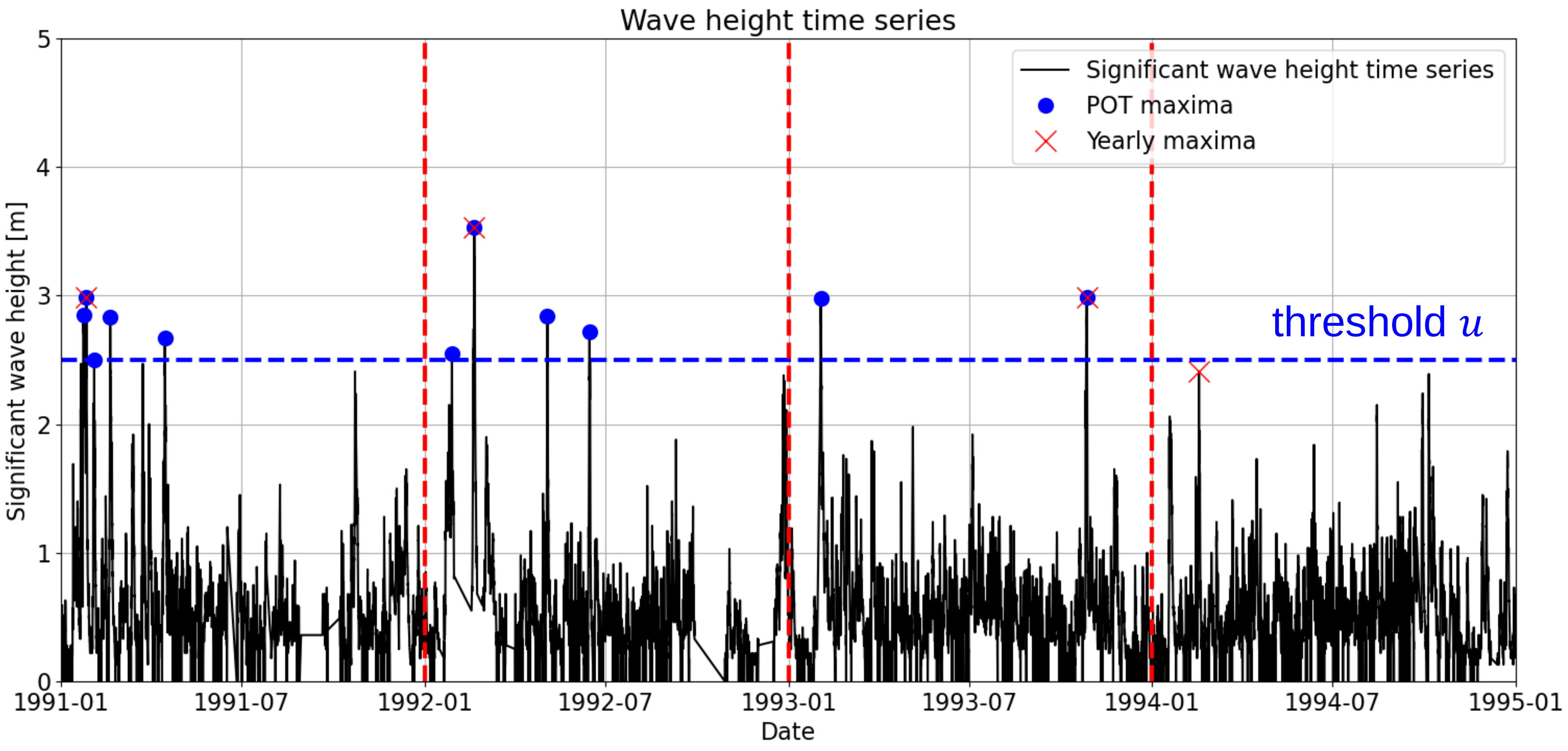
e.ragno@tudelft.nl

TU Delft

Different definition of extreme



Different definition of extreme



Generalized Pareto Distribution

Given, $\boldsymbol{X_1, \dots, X_n}$ a sequence of independent random variables, with a common distribution function $\boldsymbol{F(x)}$, and $\boldsymbol{M_n} = \mathbf{max}(\boldsymbol{X_1, \dots, X_n})$ so that for large $\boldsymbol{n}$

$$\Pr(M_n \leq z) \approx G(z)$$

where:

$$G(z) = exp\left\{-\left[1 + \xi \cdot \left(\frac{z - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}$$

for $\mu, \xi, \ \sigma > 0$.

Generalized Pareto

Then, for **large enough $\boldsymbol{u}$**

$$\boldsymbol{Pr(X - u \leq y \mid X > u) = \boxed{H(y) = 1 - \left(1 + \frac{\xi \cdot y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}}}$$

defined on $y: y > 0,\ 1 + \left(\frac{\xi \cdot y}{\tilde{\sigma}}\right) > 0$, and $\tilde{\sigma} = \sigma + \xi \cdot (u - \mu)$

Generalized Pareto Distribution

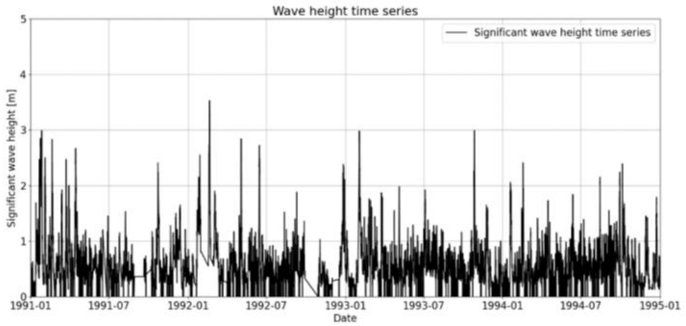
Important aspects to keep in mind when working with POT

- Threshold selection
- Excesses
  - independent
  - Poisson process
- Pareto distribution conditional probability:
  - $Pr(X < x) = P(X > u) \cdot Pr(X < x \mid X > u) = \zeta_u \cdot \left(1 - \left(1 + \frac{\xi \cdot (x - u)}{\sigma}\right)^{-\frac{1}{\xi}}\right)$
- On average more than one excesses per year

TU Delft

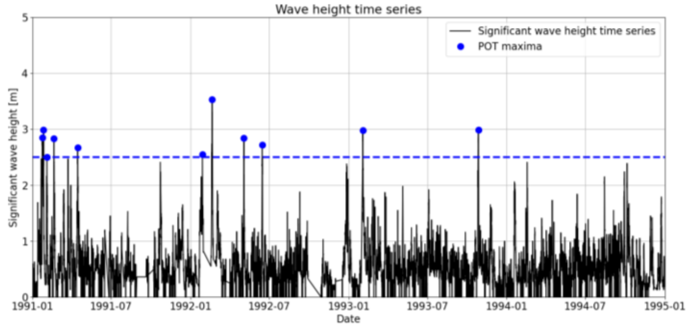
Example Case: how big is the 100-year event?

```
>> read observations
```

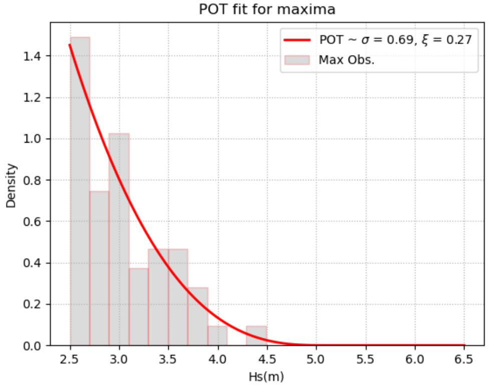


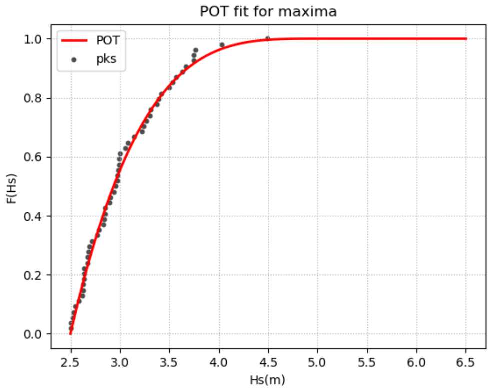

```
>> select Excesses = (OBS > U) - U
```

Important to consider also
- *Threshold selection (u = 2.5 m)*
- *Declustering time (storm duration 2 days)*




```
>> fit a GP on Excesses
```

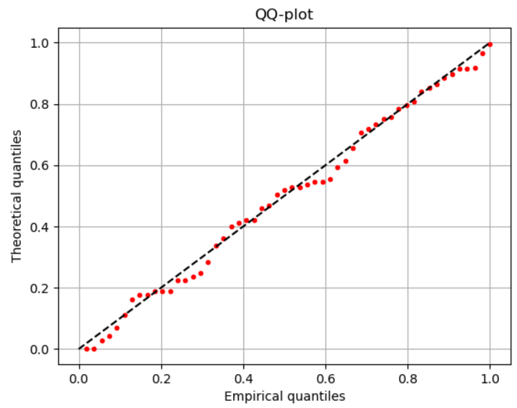




TU Delft

Example Case: how big is the 100-year event?

>> **check** the fit (e.g. QQ-plot)



>> **design event:** inverse of GP for exceedance probability corresponding to 100 yr return period

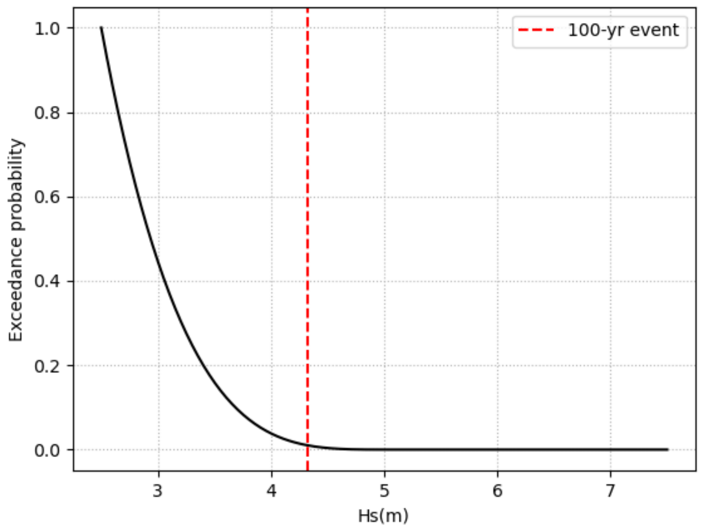
probability of exceedance must be adjusted depending on the number of exceedance per year:

>> **Tadj** $= T^*(n_y{}^*N_{Excesses}/N_{OBS})$

Design value:

>> **Zcrt = U + σ*log(Tadj) if ξ = 0**

>> **Zcrt = U + σ / ξ *((Tadj)^ ξ − 1)** otherwise



TU Delft

Model Selection

When selecting a model:

- Number of observations used
- Results of GoF tests
- Uncertainty