Extreme Value Analysis in engineering

Patricia Mares Nasarre
(p.maresnasarre@tudelft.nl)

Hydraulic Structures and Flood Risk

TU Delft

Learning objectives

1. Identify what is an **extreme value** and apply it within the engineering context

2. Interpret and apply the concept of **return period**

3. Apply extreme value **sampling techniques** to datasets:

   a. Block maxima

   b. Peak over threshold

TU Delft

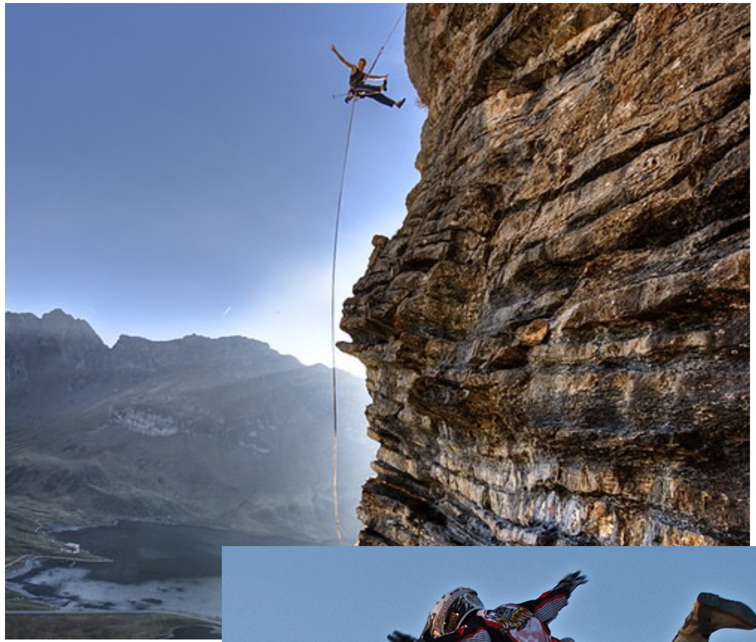
Concept of extreme value

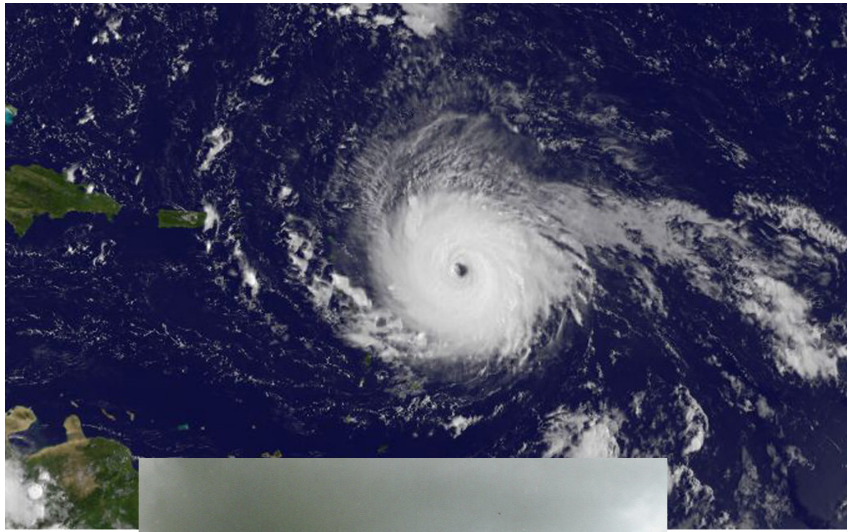
Patricia Mares Nasarre
(p.maresnasarre@tudelft.nl)

Hydraulic Structures and Flood Risk

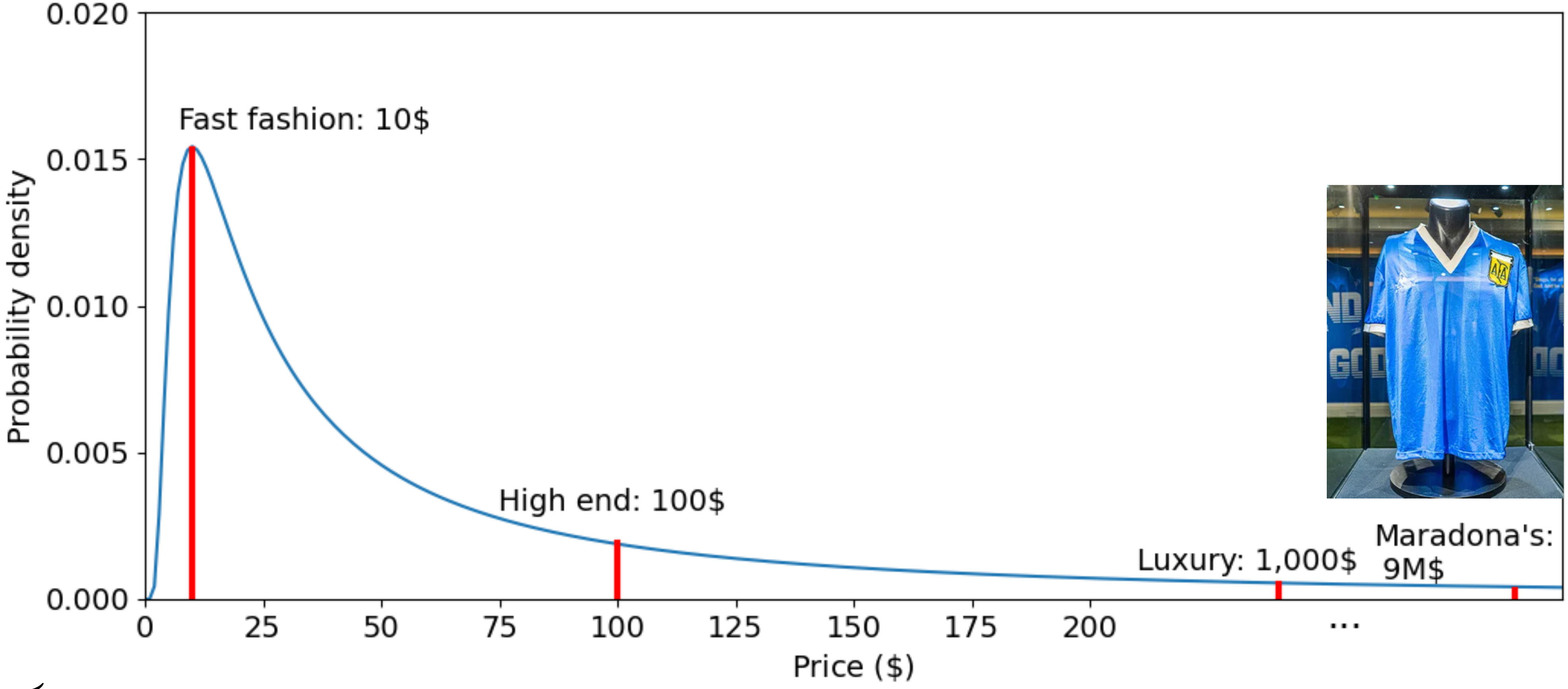
TU Delft

What is an extreme?





Example: t-shirt price



What is an extreme?

An **extreme observation** is an observation that **deviates from the average observations**

TU Delft

Why are we interested in extremes?

Infrastructures and systems are designed to **withstand extreme conditions (ULS)**.

- Breakwater → wave storm
- Flood defences → precipitation
- Bridge → maximum load
- Energy systems → max. and min. consumption
- Ecological discharges → drought

Minimum values are also extreme values!

To properly design and assess infrastructures and system **we need to characterize the uncertainty of the loads.**

TU Delft

Extreme Value Analysis

Based on historical observed extremes (limited)…

- Allows us to **model** the stochastic behaviour of extreme events
- Allows us to **infer** extremes we have not observed yet (extrapolation)

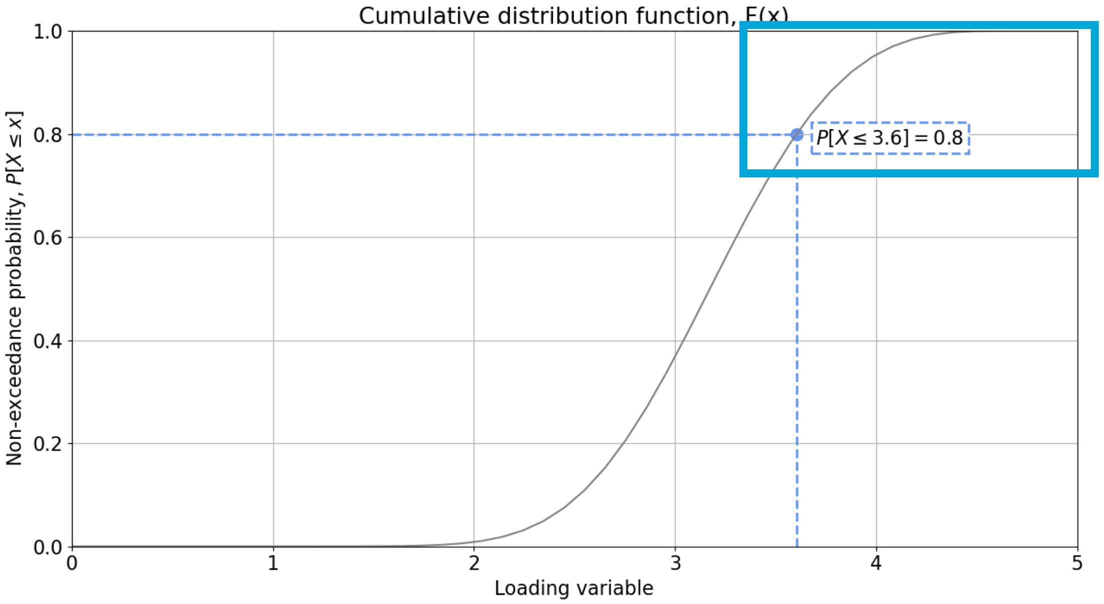
What do we need?

Time series of observations of the loading variable

Extreme Value Analysis

Design value of the loading

Uncertainty in my loading



Summary

✓ Identify what is an **extreme value** and apply it within the engineering context

Return period

Patricia Mares Nasarre
(p.maresnasarre@tudelft.nl)

Hydraulic Structures and Flood Risk

Percentile and Exceedance Probability

Consider $\boldsymbol{x_q}$ such that $\mathbf{Pr}(\boldsymbol{X} \leq \boldsymbol{x_q}) = \boldsymbol{F}(\boldsymbol{x_q}) = \boldsymbol{q}$

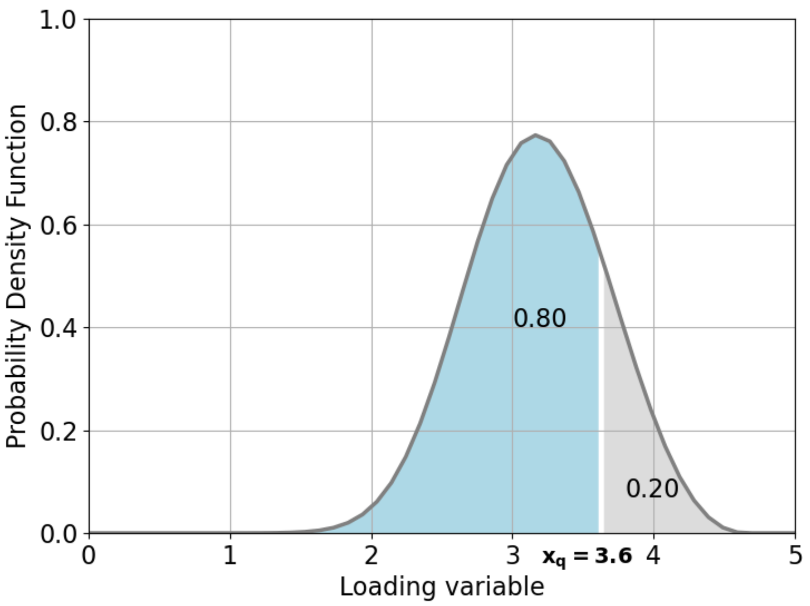
- $\boldsymbol{x_q}$ is the **qth – percentile**
- $\mathbf{Pr}(\boldsymbol{X} > \boldsymbol{x_q}) = 1 - F(x_q) = \mathbf{1} - \boldsymbol{q} = \boldsymbol{p}$ is the **exceedance probability**

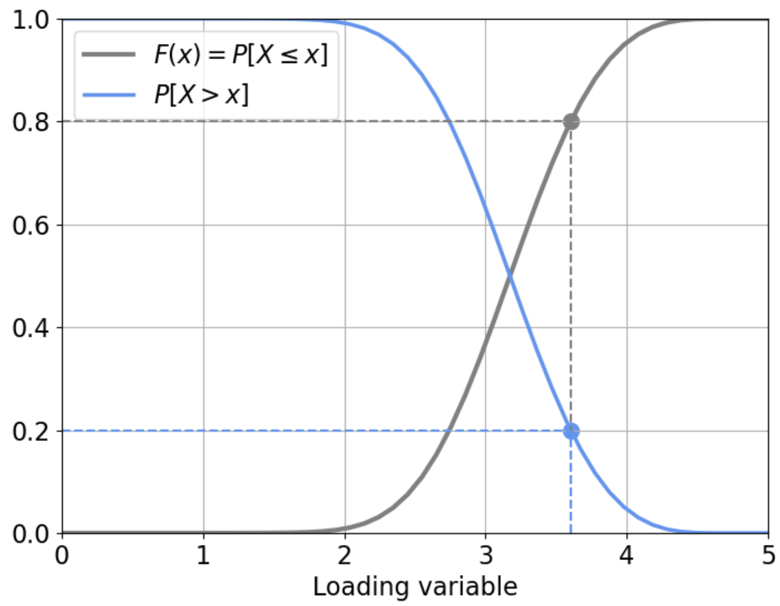
TU Delft

Percentile and Exceedance Probability

Consider $\boldsymbol{x_q}$ such that $\mathbf{Pr}\left(\boldsymbol{X} \leq \boldsymbol{x_q}\right) = \boldsymbol{F}\left(\boldsymbol{x_q}\right) = \boldsymbol{q}$

- $\boldsymbol{x_q}$ is the **qth – percentile**
- $\mathbf{Pr}\left(\boldsymbol{X} > \boldsymbol{x_q}\right) = 1 - F\left(x_q\right) = \mathbf{1} - \boldsymbol{q} = \boldsymbol{p}$ is the **exceedance probability**





80th-percentile: $x_q = 3.60$

$Pr(X \leq 3.6) = 0.8$

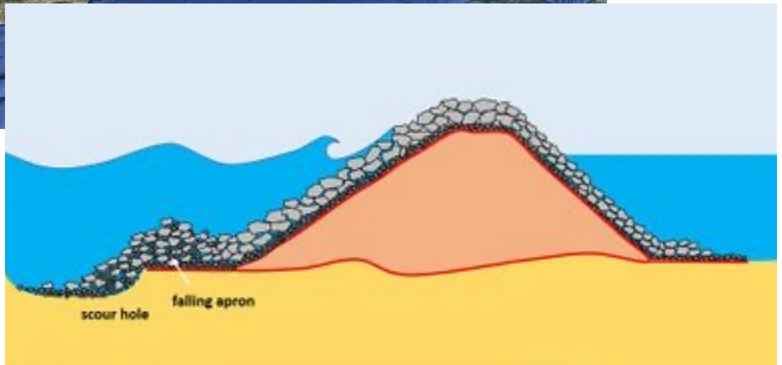
Exceedance probability

$Pr\left(X > x_q\right) = 0.20$

Ready?

Let's apply Extreme Value Analysis together!!

Example case: intervention in the Mediterranean coast



- It may be a coastal structure, a water intake, the restoration of a sandy beach, between others.
- Here: **design a mound breakwater**
- Mound breakwater must resist wave storms
- But which one?

TU Delft

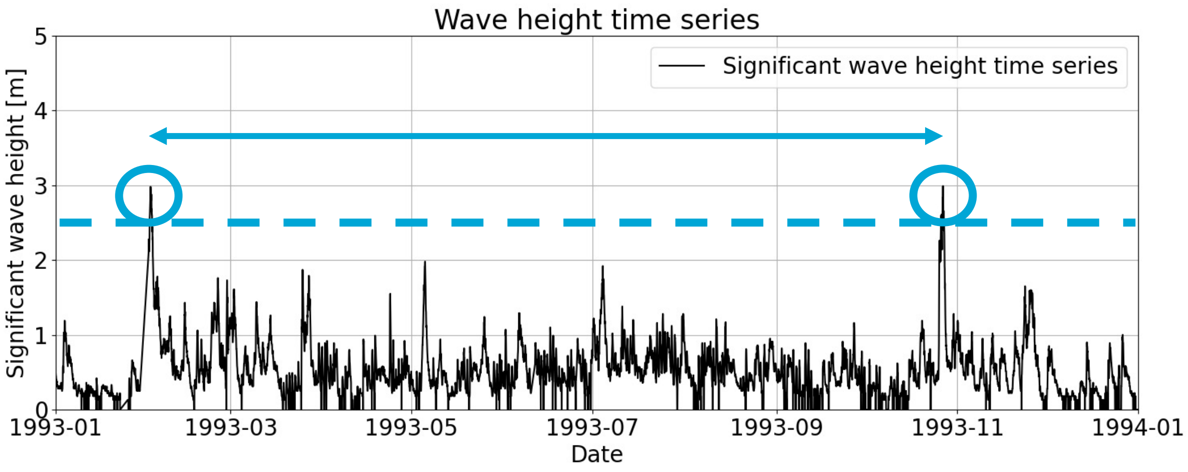
Design requirements

Regulations and recommendations → Exceedance probability or **return period**

Country	Standard	**T_R (years)**	DL (years)	p_{DL} (-)
England	BS 6349-1-1:2013	**50-100***	50-100	0.05*
Japan	TS Ports-2009	**50-100**	50	0.40-0.64
Spain	ROM 0.0-01/1.0-09	**113-4,975**	25-50	0.01-0.2

*Not well defined

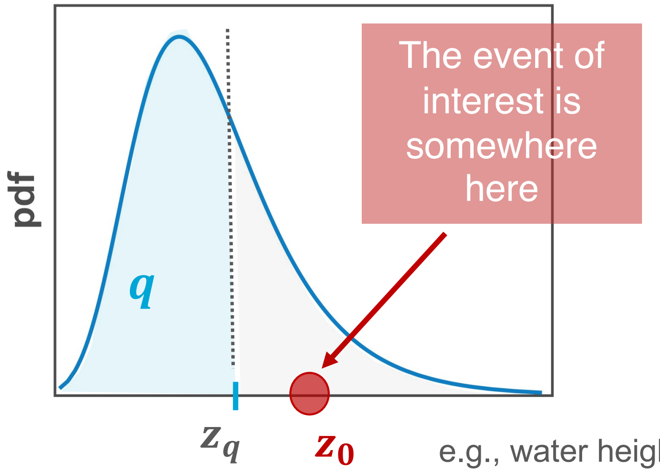
But what is return period?



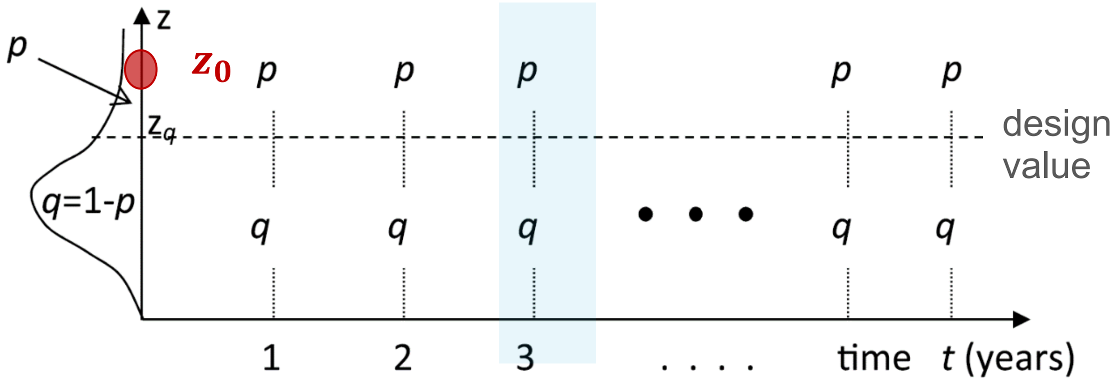
TU Delft

Return Period - Derivation

We are interested in estimating, on average, the time (e.g., year(*)) at which an event (here, the wave height) higher than a given threshold, (e.g. design value), occurs.
We know that $\Pr(Z > z_q) = 1 - q = p$



Every year the probability of the event being higher/lower than the threshold is always the same



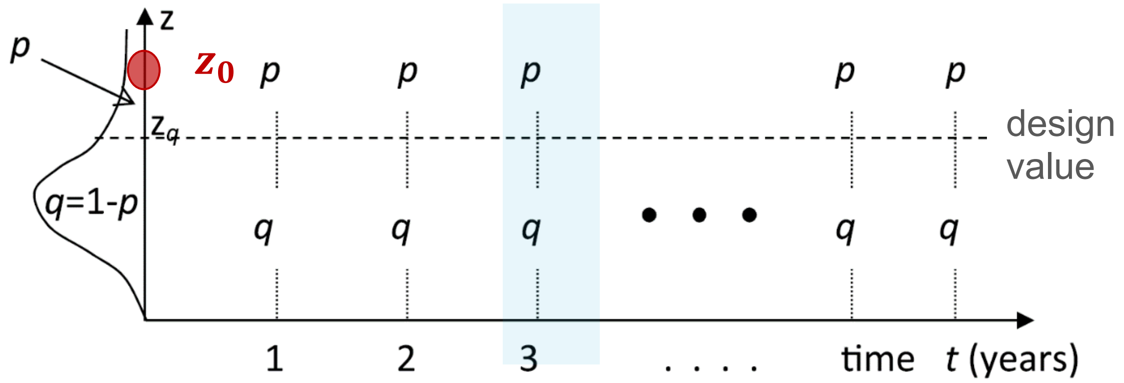
(*) the unit time reflects the interval time in which the observations are taken

Right figure from Salas, et la (2013). *Journal of Hydrologic Engineering*, *19*(3), 554-568.

Return Period - Derivation

Every year the probability of the event being higher/lower than the threshold is always the same

Let's calculate the probability that an event z_0 higher than the design value z_q occurs at time t



$$f(t) = Pr(z_0 \; at \; time \; t) = (1-p)(1-p) \dots (1-p)p$$

$$\boldsymbol{f(t) = Pr(z_0 \; at \; time \; t) = q^{t-1}p}$$

Geometric Distribution

it models the number of trials up to the first success (included)

$$\boldsymbol{T(t) = \frac{1}{p}}$$

T(t) expectation

it will take on average 1/p trials to get a success

T is also defined as **Return Period** (in unit time).
"We have to make, on average, 1/p trials in order that the event happens once" (Gumbel)
or **wait 1/p years before the next occurrence**

TU Delft

Design requirements

Regulations and recommendations → Exceedance probability or **return period**

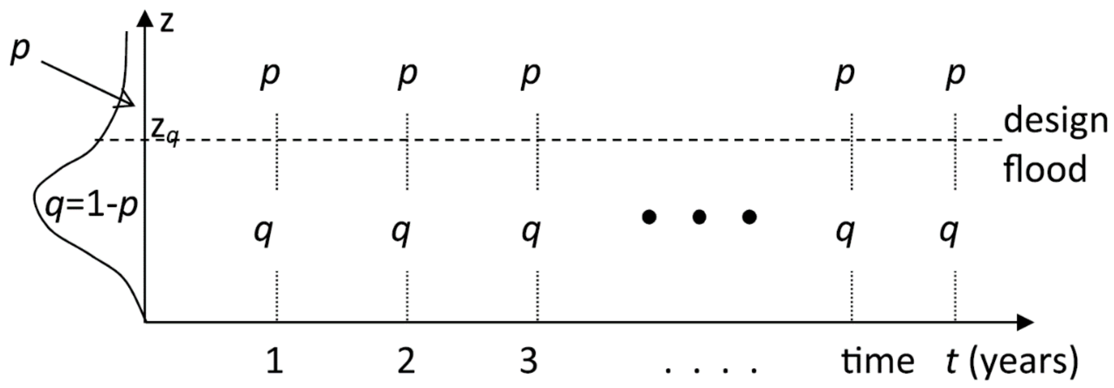
$$T_R = \frac{1}{p} = \frac{1}{1-(1-p_{DL})^{1/DL}}$$

Country	Standard	T_R (years)	DL (years)	p_{DL} (-)
England	BS 6349-1-1:2013	50-100*	50-100	0.05*
Japan	TS Ports-2009	50-100	50	0.40-0.64
Spain	ROM 0.0-01/1.0-09	113-4,975	25-50	0.01-0.2

*Not well defined

Return Period and Design Life

Let's calculate the probability to observe an event z_0 higher than the design value z_q <u>at least once</u> in DL years of design life. Under *iid* conditions:



$$p_{DL} = 1 - (1-p)(1-p)\ldots(1-p) = 1 - \prod_{i=1}^{DL}(1-p_i) = 1-(1-p)^{DL}$$

$$p_{DL} = 1-(1-p)^{DL} \rightarrow p = 1-(1-p)^{\frac{1}{DL}}$$

$$T_R = \frac{1}{p} = \frac{1}{1-(1-p_{DL})^{1/DL}}$$

TU Delft

Design requirements

Regulations and recommendations → Exceedance probability or **return period**

$$T_R = \frac{1}{p} = \frac{1}{1-(1-p_{DL})^{1/DL}}$$

Country	Standard	T_R (years)	DL (years)	p_{DL} (-)
England	BS 6349-1-1:2013	50-100*	50-100	0.05*
Japan	TS Ports-2009	50-100	50	0.40-0.64
Spain	ROM 0.0-01/1.0-09	113-4,975	25-50	0.01-0.2

*Not well defined

Design requirements – Regulator example

Figure 2.2.33. ***ERI, SERI and minimum useful life for different types of sheltered area***

TYPE OF SHELTERED OR PROTECTED AREA			ERI [7]		MINIMUM USEFUL LIFE (L_m) [7] (years)
HARBOR AREAS	COMMERCIAL PORT	All vessel types	r_3	High	50
HARBOR AREAS	COMMERCIAL PORT	Specific vessel types	$r_2(r_3)$ [1]	Medium (high) [1]	25 (50) [1]
HARBOR AREAS	FISHING PORT		r_2	Medium	25
HARBOR AREAS	MARINA		r_2	Medium	25
HARBOR AREAS	INDUSTRIAL PORT		$r_2(r_3)$ [1]	Medium (High) [1]	25 (50) [1]
HARBOR AREAS	NAVAL PORT		$r_2(r_3)$ [2]	Medium (High) [2]	25 (50) [2]
HARBOR AREAS	PROTECTION OF FILL MATERIAL OR SHORELINE		r_2 (r_3) [3]	Medium (High) [3]	25 (50) [3]
COASTAL AREAS	DEFENSE AGAINST EXTREME FLOOD EVENTS [4]		r_3	High	50
COASTAL AREAS	PROTECTION OF WATER INTAKE OR DISCHARGE STRUCTURE		$r_2(r_3)$ [5]	Medium (High) [5]	25 (50) [5]
COASTAL AREAS	SHORELINE PROTECTION AND DEFENSE		$r_1(r_3)$ [6]	Low (High) [5]	15 (50) [7]
COASTAL AREAS	BEACH DEFENSE AND NOURISHMENT		r_1	Low	15

DL=25years

ROM 1.0-09
Recommendations for the Project Design and Construction of Breakwaters (Part 1: Calculation and Project Factors. Climate Agents)

Figure 2.2.34. ***SERI and joint probability of failure for ULS and SLS***

TYPE OF SHELTERED OR PROTECTED AREA				SERI		$P_{f,ULS}$	$P_{f,SLS}$
HARBOR AREAS	COMMERCIAL PORT	Storage areas or areas for passengers and/or cargo handling adjacent to the breakwater [1]	Hazardous cargo [2]	s_3	High	0.01	0.07
HARBOR AREAS	COMMERCIAL PORT	Storage areas or areas for passengers and/or cargo handling adjacent to the breakwater [1]	Passengers and non-hazardous cargo [1]	s_2	Low	0.10	0.10
HARBOR AREAS	COMMERCIAL PORT	No storage areas or areas for passengers and/or cargo handling adjacent to the breakwater		s_1	Insignificant	0.20	0.20
HARBOR AREAS	FISHING PORT	Storage or operational areas adjacent to the breakwater		s_2	Low	0.10	0.10
HARBOR AREAS	FISHING PORT	No storage or operational areas adjacent to the breakwater		s_1	Insignificant	0.20	0.20
HARBOR AREAS	MARINA	Storage or operational areas adjacent to the breakwater		s_2	Low	0.10	0.10
HARBOR AREAS	MARINA	No storage or operational areas adjacent to the breakwater		s_1	Insignificant	0.20	0.20
HARBOR AREAS	INDUSTRIAL PORT	Storage or cargo handling areas adjacent to the breakwater [1]	Hazardous cargo [2]	s_3	High	0.01	0.07
HARBOR AREAS	INDUSTRIAL PORT	Storage or cargo handling areas adjacent to the breakwater [1]	Non-hazardous cargo	s_2	Low	0.10	0.10
HARBOR AREAS	INDUSTRIAL PORT	No storage or cargo handling areas adjacent to the breakwater		s_1	Insignificant	0.20	0.20
HARBOR AREAS	NAVAL PORT	Storage or operational areas adjacent to the breakwater [1]		s_3	High	0.01	0.07
HARBOR AREAS	NAVAL PORT	No storage or operational areas adjacent to the breakwater		s_1	Insignificant	0.20	0.20
HARBOR AREAS	PROTECTION *	Storage area adjacent to the breakwater [1]	Hazardous cargo [2]	s_3	High	0.01	0.07
HARBOR AREAS	PROTECTION *	Storage area adjacent to the breakwater [1]	Non-hazardous cargo	s_2	Low	0.10	0.10

p_{DL}=0.20

Design requirements

Regulations and recommendations → Exceedance probability or **return period**

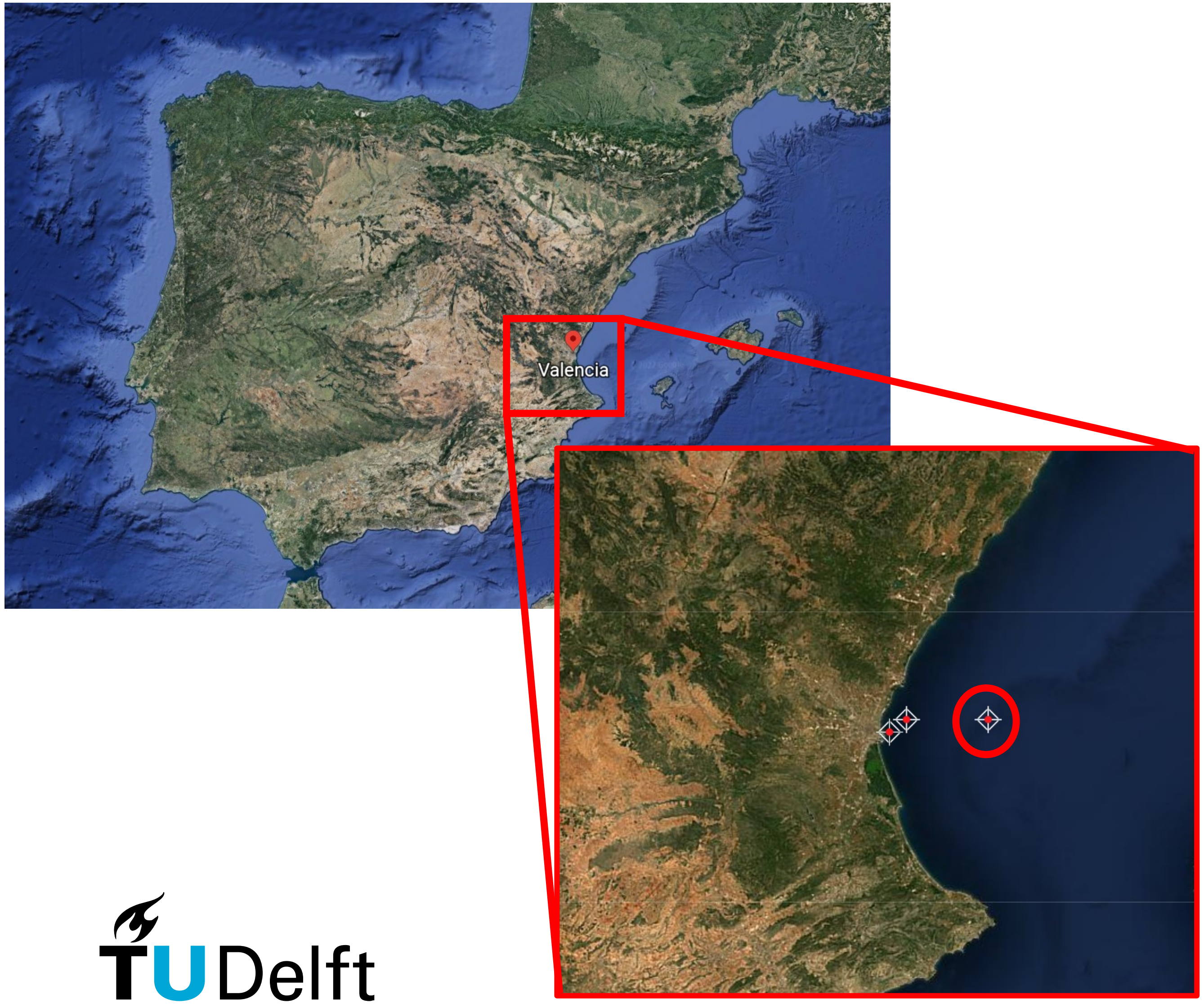
$$T_R = \frac{1}{p} = \frac{1}{1-(1-p_{DL})^{1/DL}}$$

$$T_R = \frac{1}{p} = \frac{1}{1-(1-0.20)^{1/25}} = 112.5\ years$$

$\boxed{T_R \approx 100 \text{ years}}$ ⇒ **Load** $\boxed{H_{s,T=100\ years}}$

TU Delft

Example case: intervention in the Mediterranean coast



- **Load: significant wave height (T_R=100 years)**
- Historical data from a buoy in the Mediterranean sea, in front of Valencia coast
- 20 years of hourly measurements → **infer design value using EVA**

TU Delft

Learning objectives

1. Identify what is an **extreme value** and apply it within the engineering context
2. Interpret and apply the concept of **return period**
3. Apply extreme value **sampling techniques** to datasets:
   a. Block maxima
   b. Peak over threshold

TU Delft

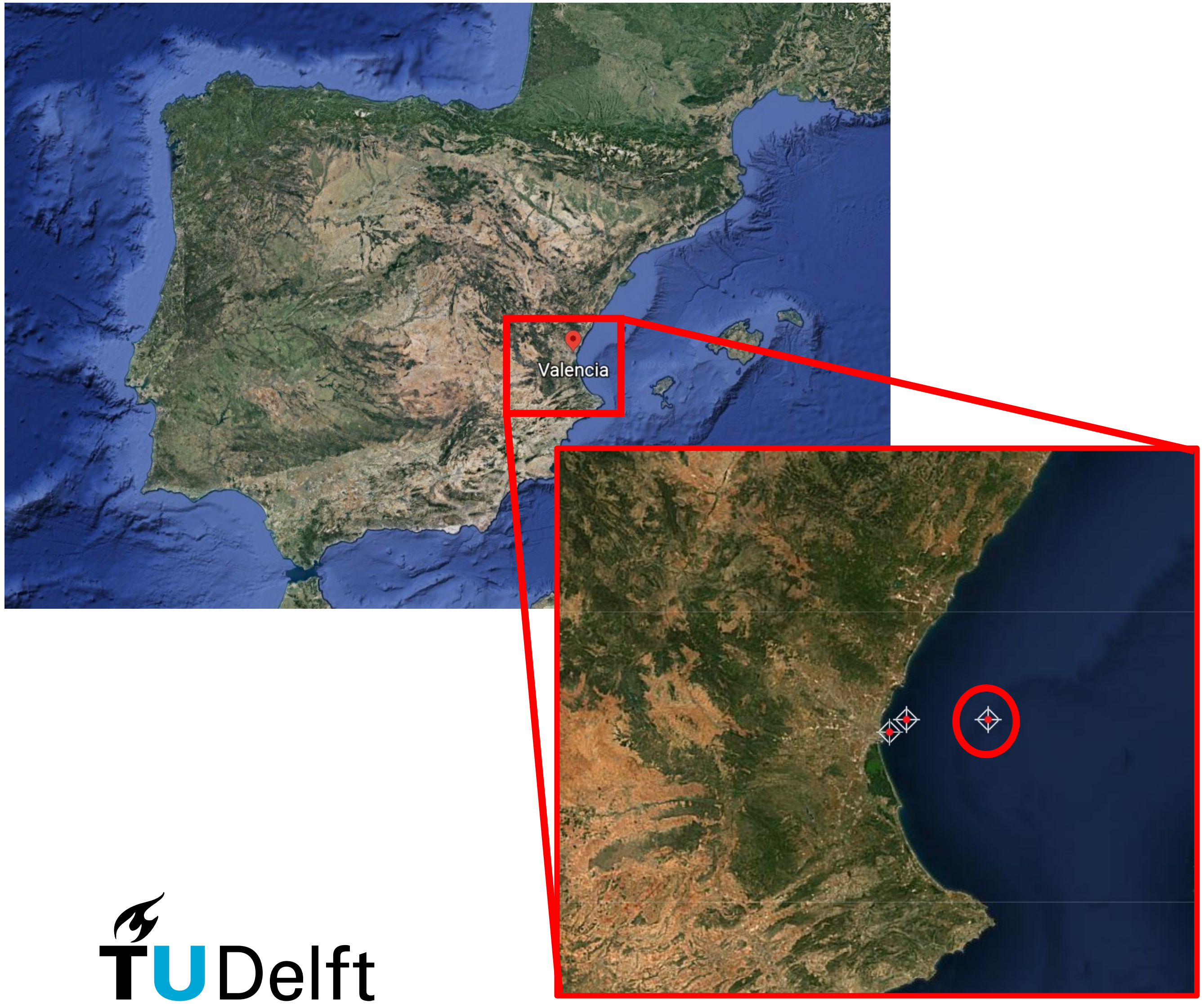
Sampling extremes

Patricia Mares Nasarre
(p.maresnasarre@tudelft.nl)

Hydraulic Structures and Flood Risk

TU Delft

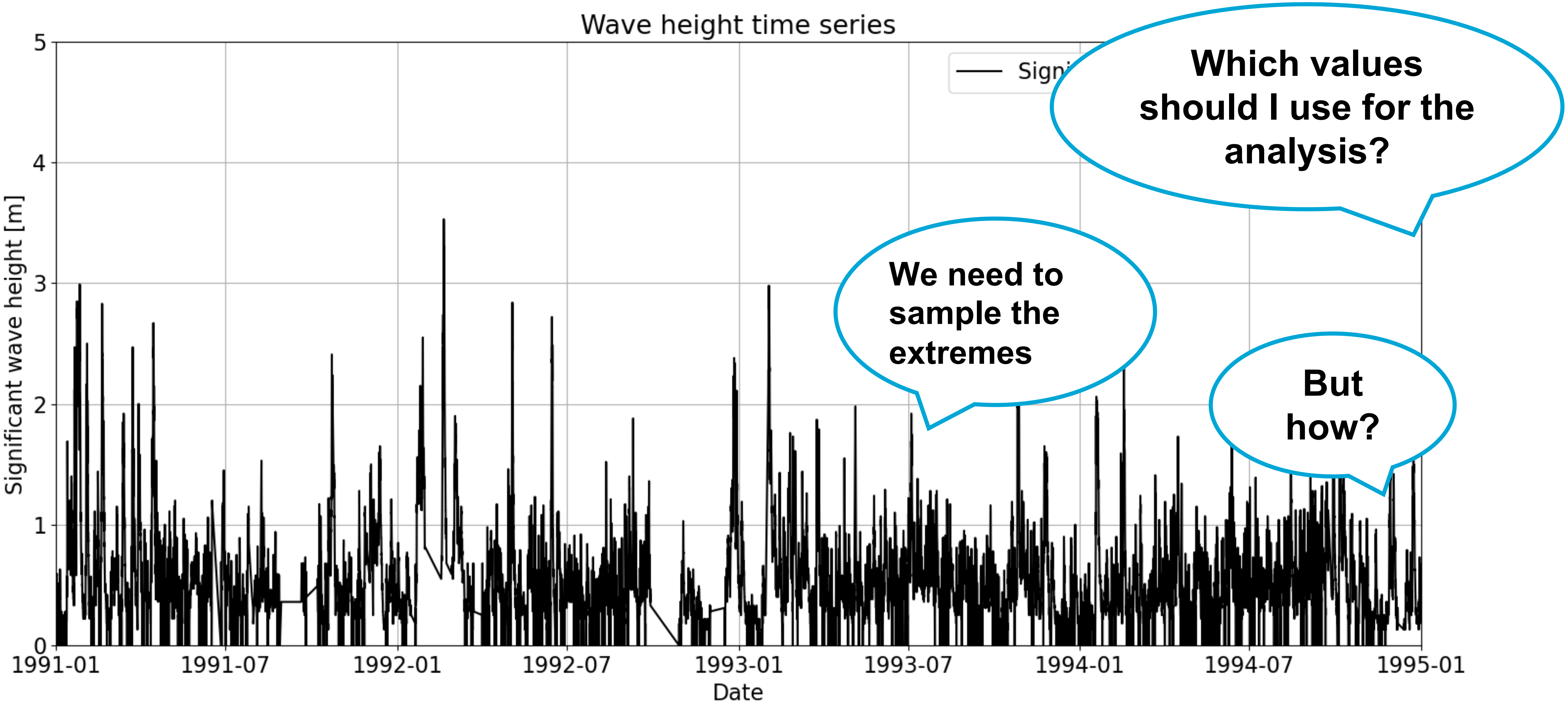
Example case: intervention in the Mediterranean coast



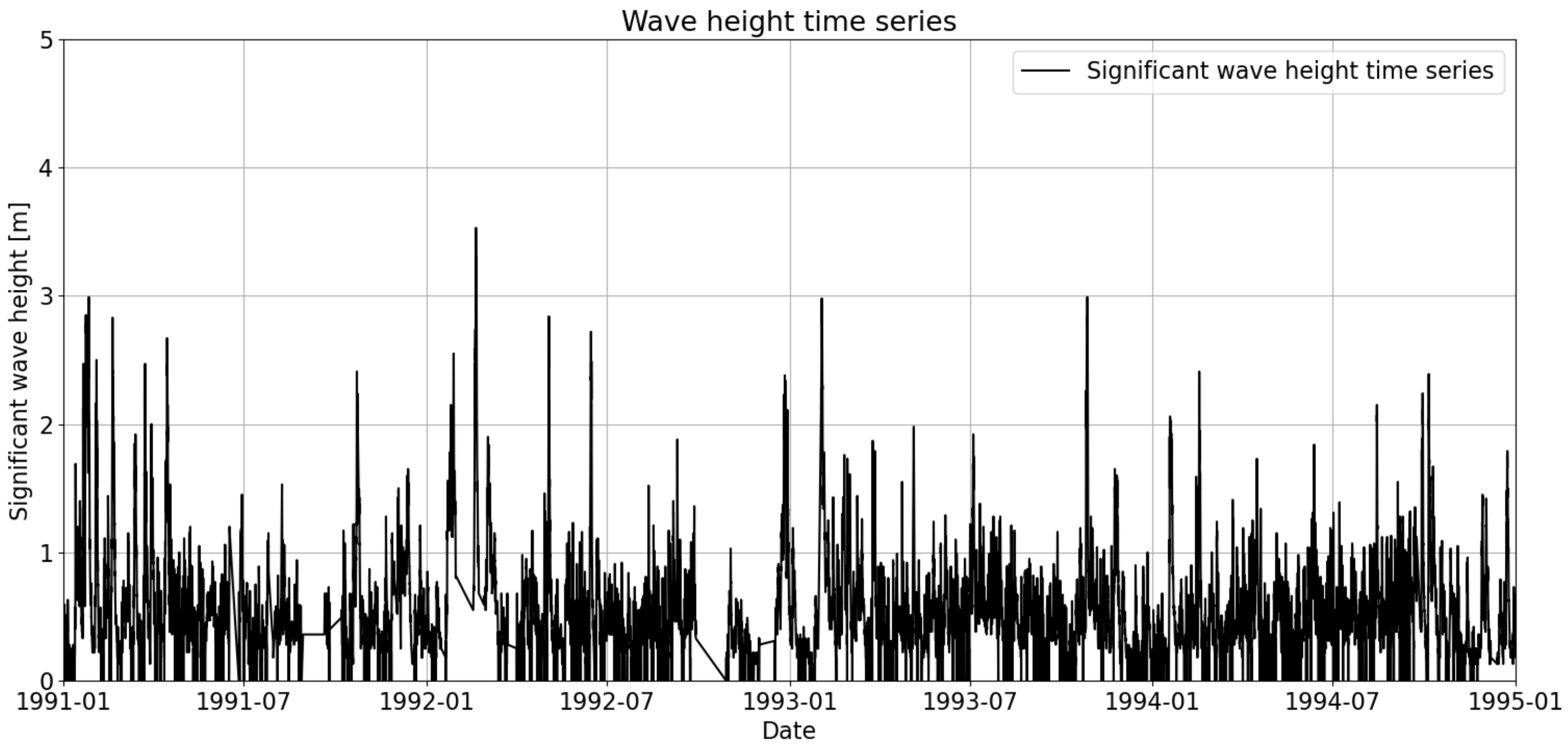
- **Load: significant wave height (T_R=100 years)**
- Historical data from a buoy in the Mediterranean sea, in front of Valencia coast
- 20 years of hourly measurements → **infer design value using EVA**

TU Delft

Time series


Wave height time series
Signi
Significant wave height [m]
0
1
2
3
4
5
1991-01
1991-07
1992-01
1992-07
1993-01
1993-07
1994-01
1994-07
1995-01
Date
Which values should I use for the analysis?
We need to sample the extremes
But how?

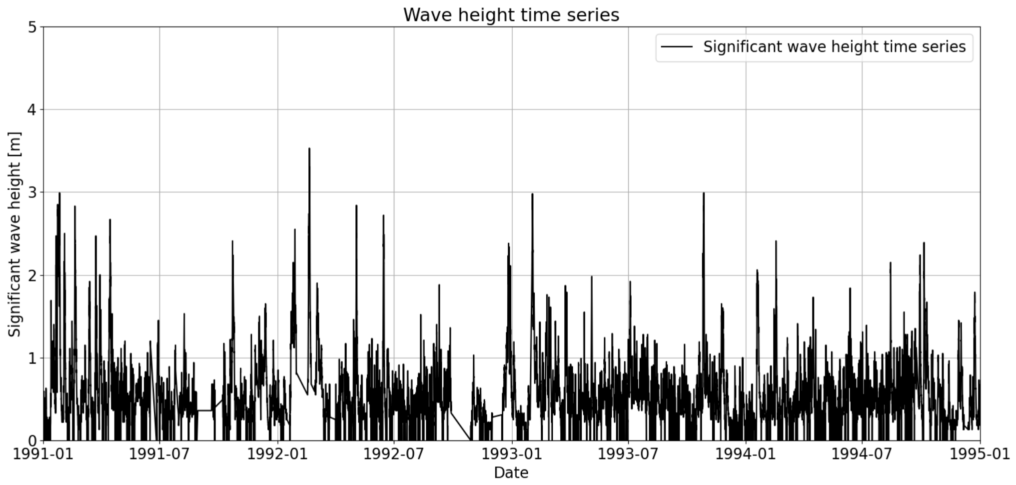
How can we sample extremes?



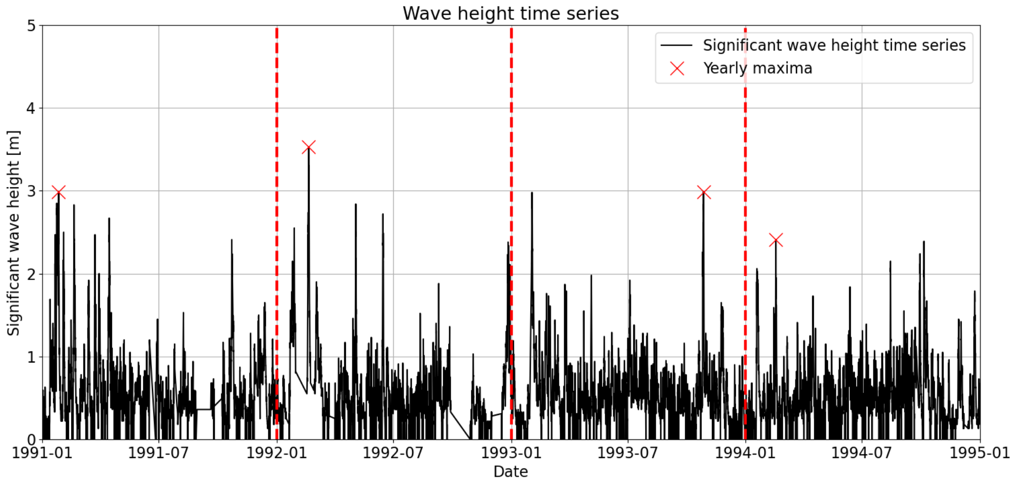
Two techniques:

1. **Block Maxima**
2. **Peak Over Threshold (POT)**

Sampling extremes: Block Maxima

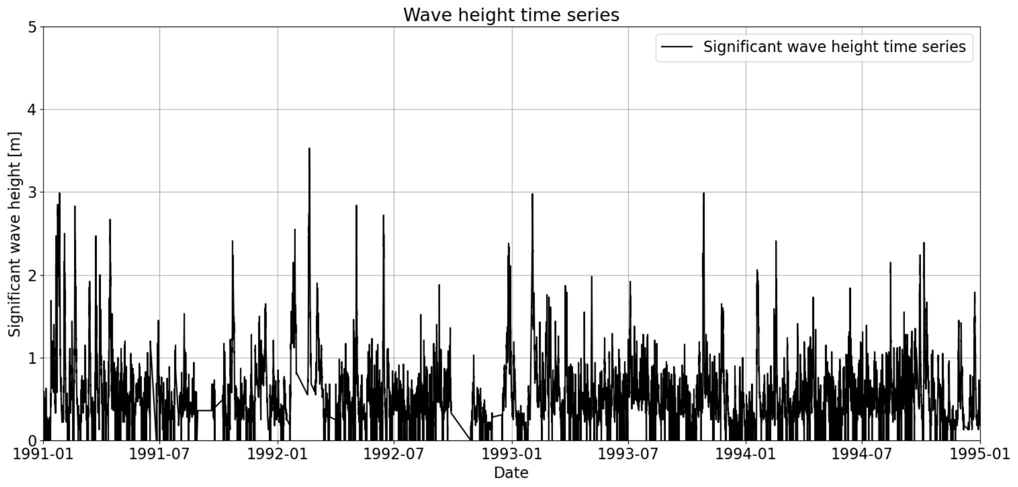


- Maximum value within the block (typically one year)
- Number of selected events=number of blocks
- Easy to implement

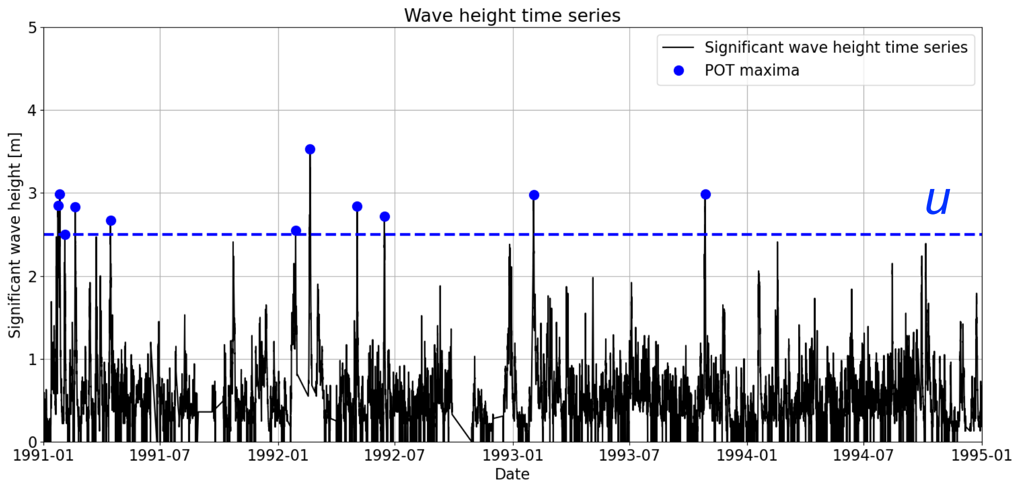



```
>> read observations

>> for each year i
        OBSmax(i)= max(observation in year i)

end
```

Sampling extremes: Peak Over Threshold (POT)



- Excesses over a threshold
- Usually, higher number of identified extremes
- Additional parameters:
  - Threshold
  - Declustering time



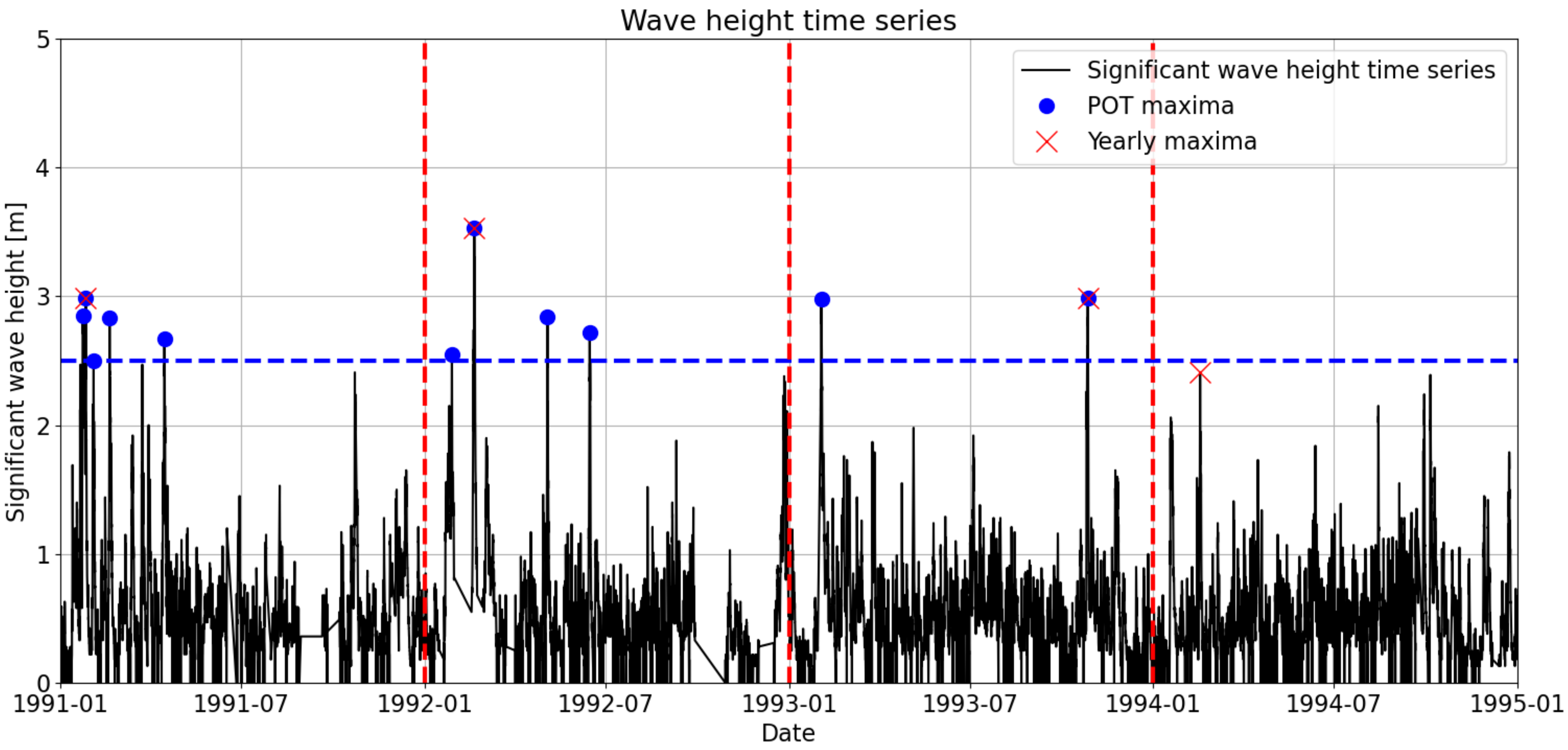

```
>> read observations

>> Define parameters

u=2.5
d=2*24
```

Threshold=2.5m
Declustering time (storm duration) = 2 days (in hours)

```
>>Select Excesses=
find_peaks(OBS, threshold=u,
        distance=d)-u
```

TUDelft

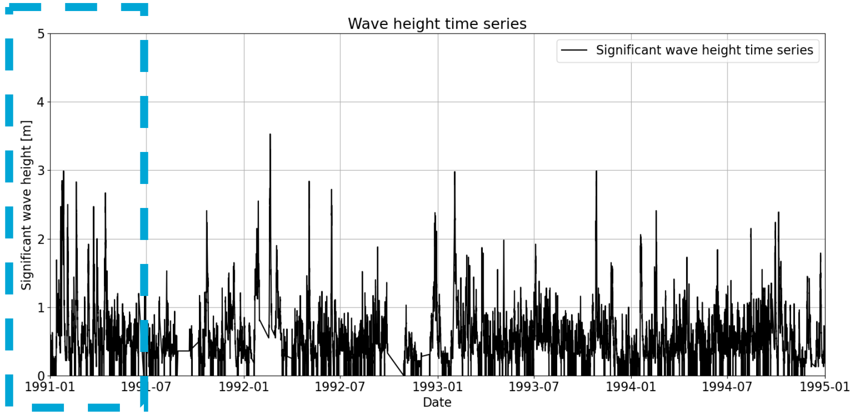
Sampling extremes: Peak Over Threshold (POT)

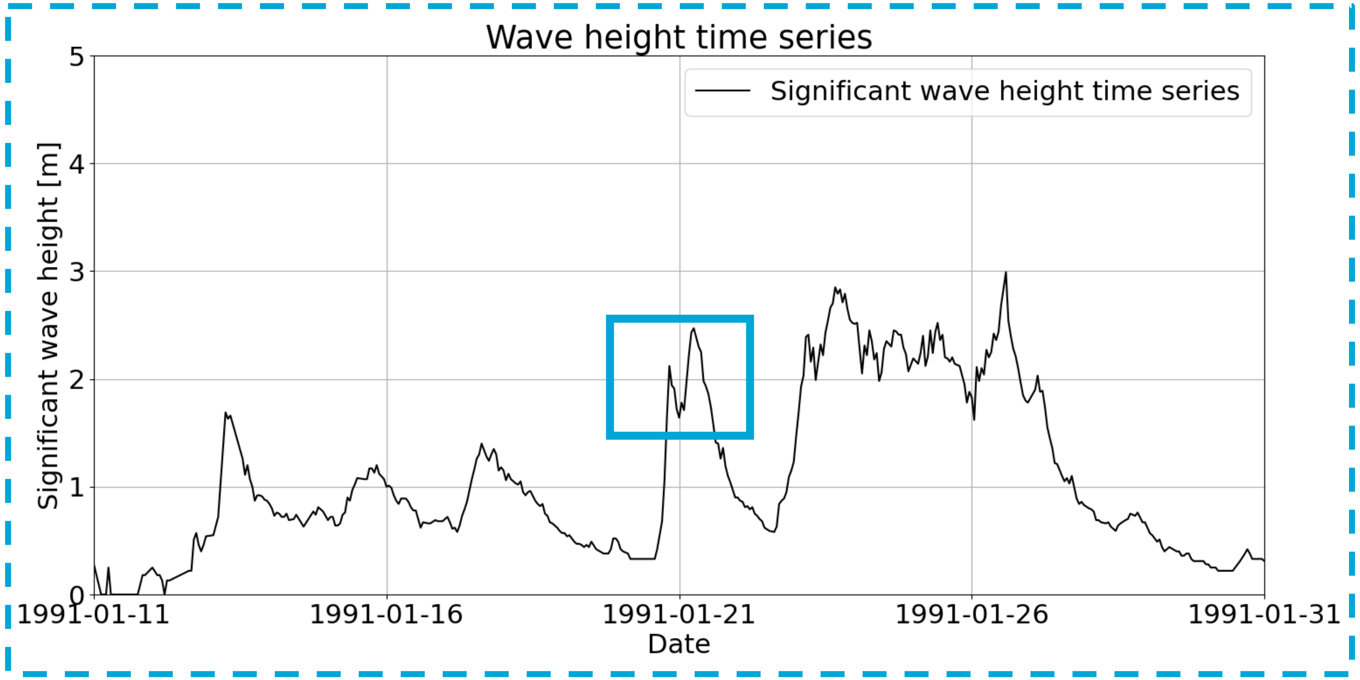


Choosing POT parameters

Parameters for POT (threshold and declustering time) should be chosen so the identified extreme events are independent (*iid* assumption).

But extremes tend to cluster…





TU Delft

Choosing POT parameters

Parameters for POT (threshold and declustering time) should be chosen so the identified extreme events are independent (*iid* assumption).

Under *iid* conditions, we have:

- A series of Bernoulli trials (exceeds or not the threshold)
- Sum the number of excesses each year → Poisson distribution

Number of exceedances per year follows a Poisson distribution.
We can check it using:

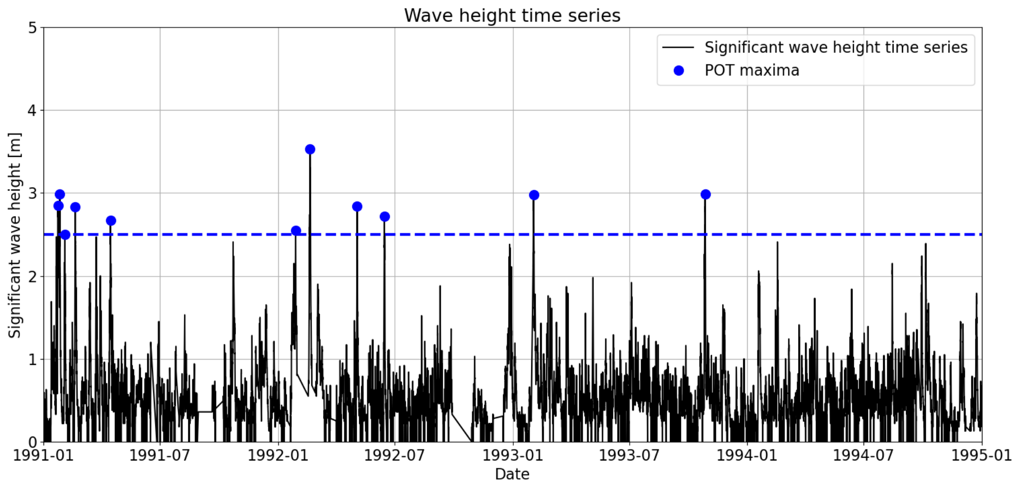
- Mean=variance=parameter (property of Poisson distribution)
- GOF to Poisson distribution (e.g.: Chi Square test for discrete distributions)

TU Delft

Choosing POT parameters

Number of exceedances per year follows a Poisson distribution.
We can check it using:

- Mean=variance=parameter (property of Poisson distribution)
- GOF to Poisson distribution (e.g.: Chi Square test for discrete distributions)

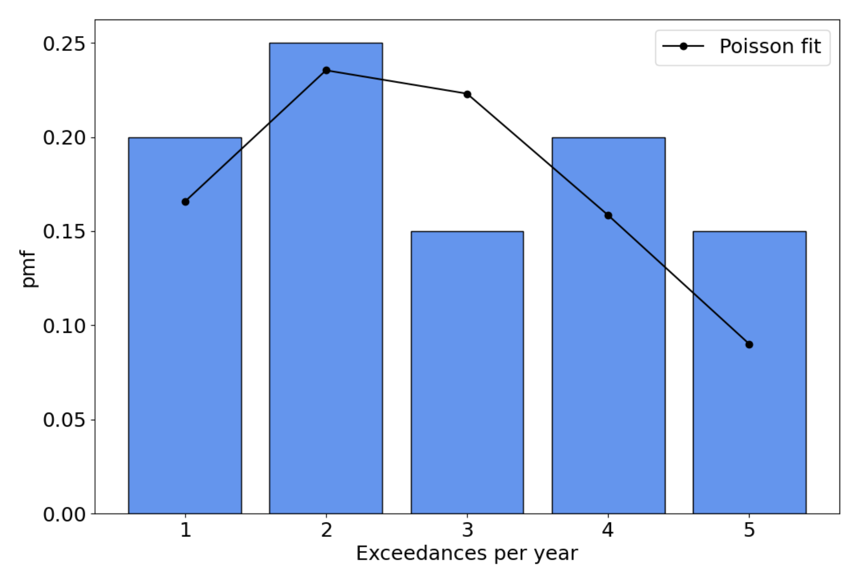



```
>> read observations

>> Define parameters

>> Select Excesses=find_peaks(OBS, threshold=u,
        distance=d)-u
```

Choosing POT parameters

Number of exceedances per year follows a Poisson distribution.
We can check it using:

- Mean=variance=parameter (property of Poisson distribution)
- GOF to Poisson distribution (e.g.: Chi Square test for discrete distributions)




```
>> read observations

>> Define parameters

>> Select Excesses=find_peaks(OBS, threshold=u,
          distance=d)-u

>> for each year i
          nExceedances(i)=count(Excesses in year i)
end

>> plot histogram(nExceedances)

>> fit a Poisson on nExceedances

>> check fit (e.g. Chi Square)
```

TU Delft

Learning objectives

1. Identify what is an **extreme value** and apply it within the engineering context
2. Interpret and apply the concept of **return period**
3. Apply extreme value **sampling techniques** to datasets:

   a. Block maxima

   b. Peak over threshold