Natural Disasters

Laurens de Haan

Erasmus University Rotterdam, NL
University of Lisbon, PT

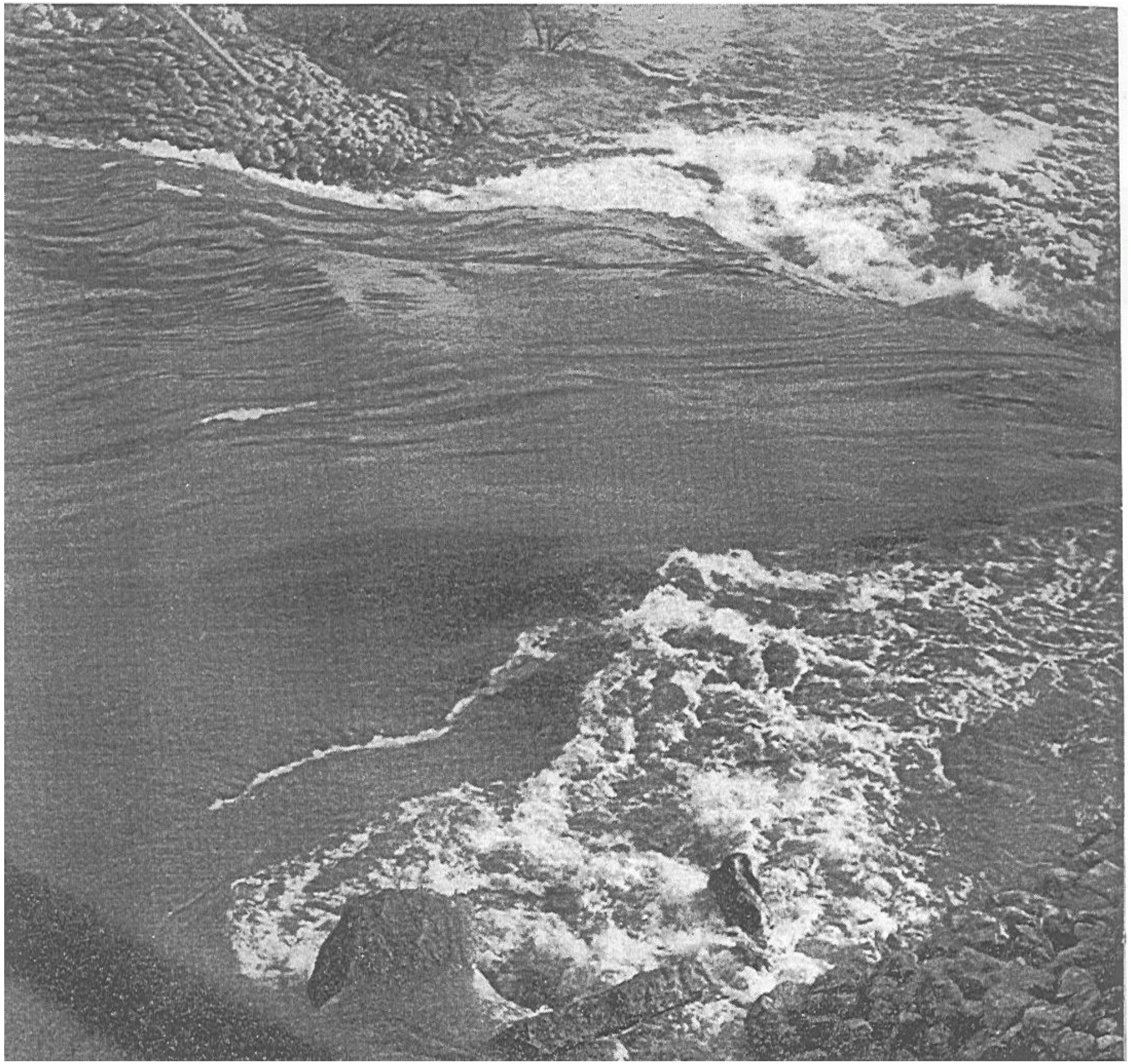
40% of the Netherlands is below sea level.

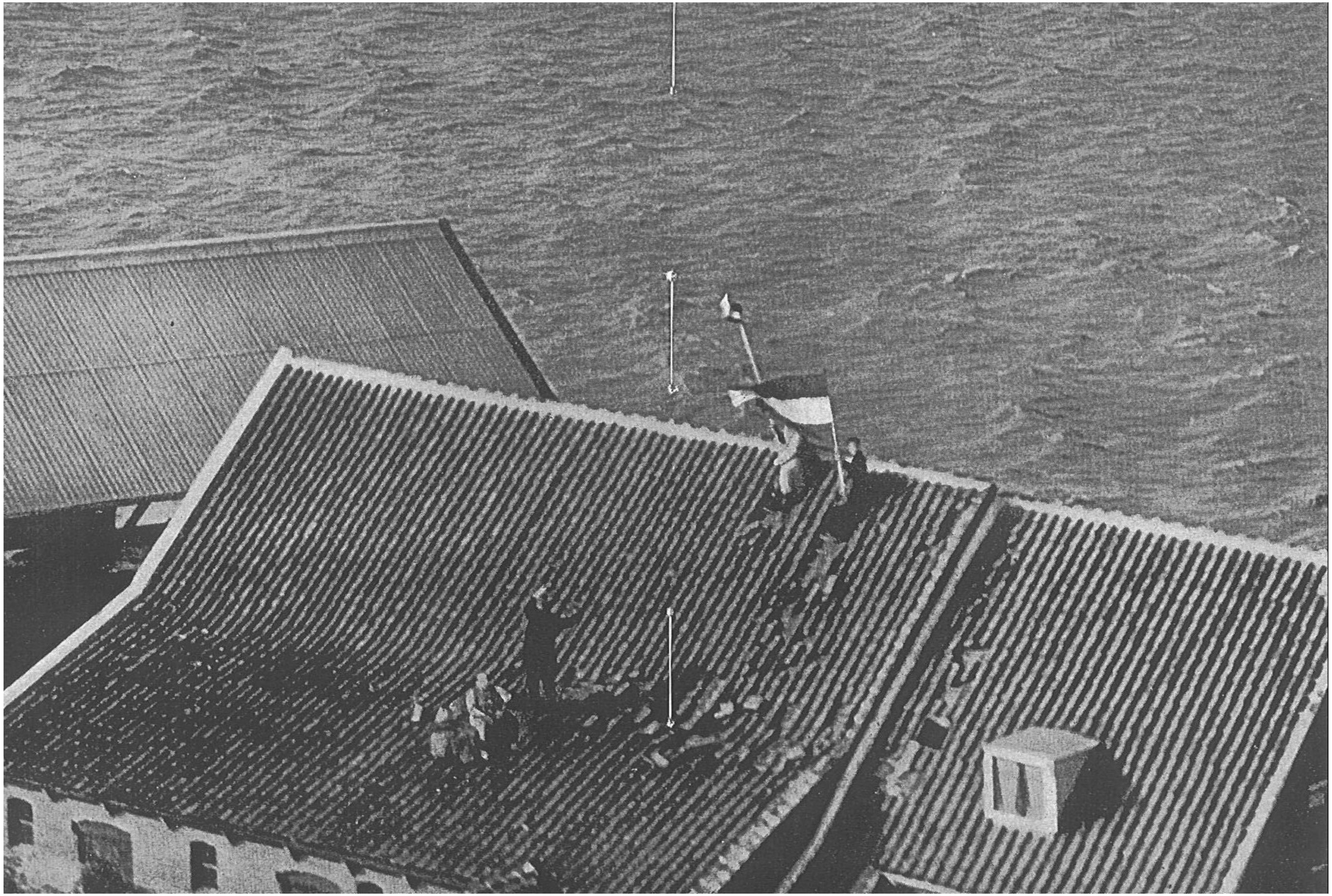
Part of it is protected by sea walls.

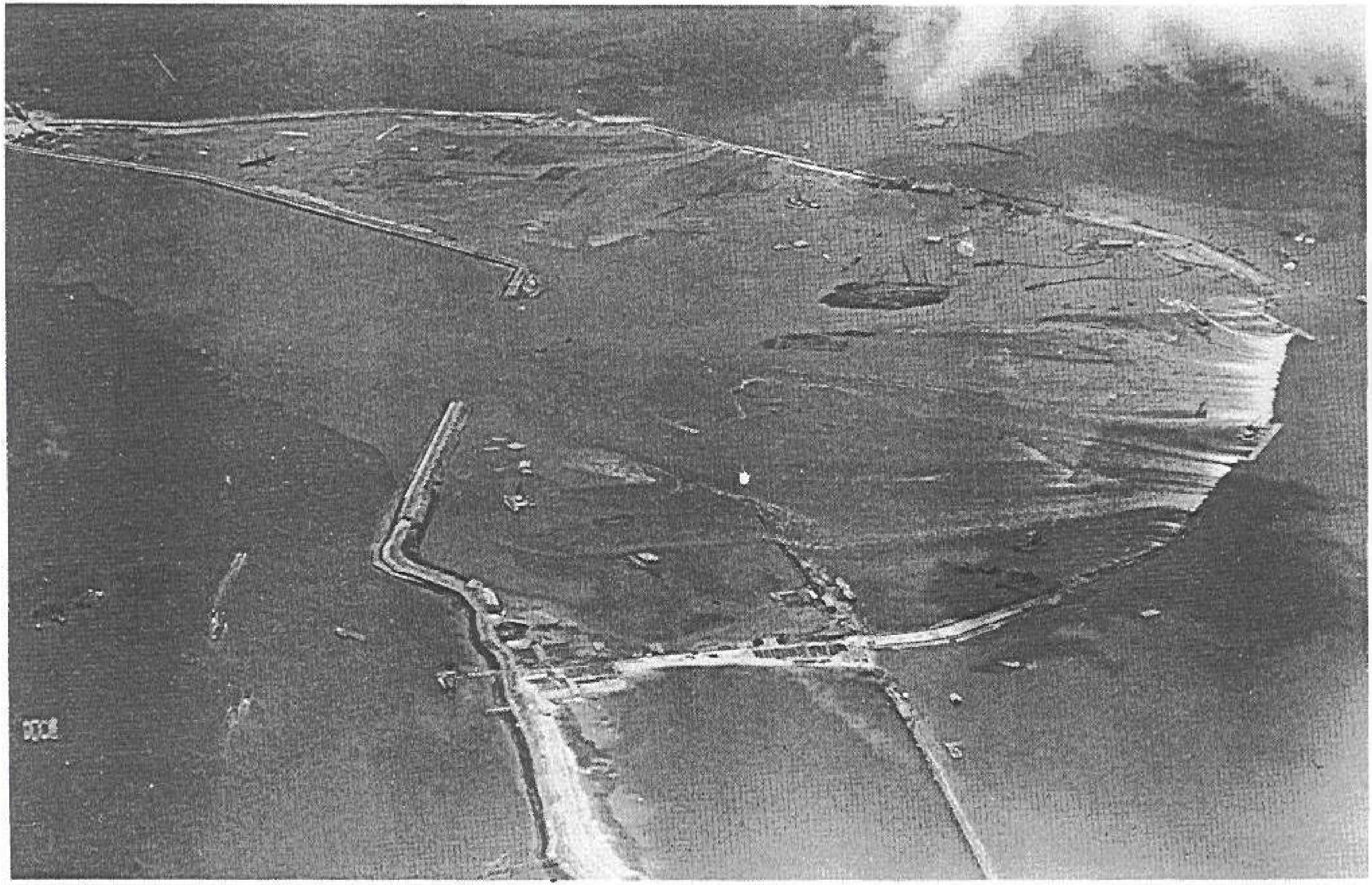
In 1953 during a severe storm many walls broke and big areas were flooded.

About 2000 people died.

Clearly the walls were too low (overtopping).







Walls broke because of unprecedented high water level (perhaps once in 500 years i.e. with probability 1/500 in some given year).

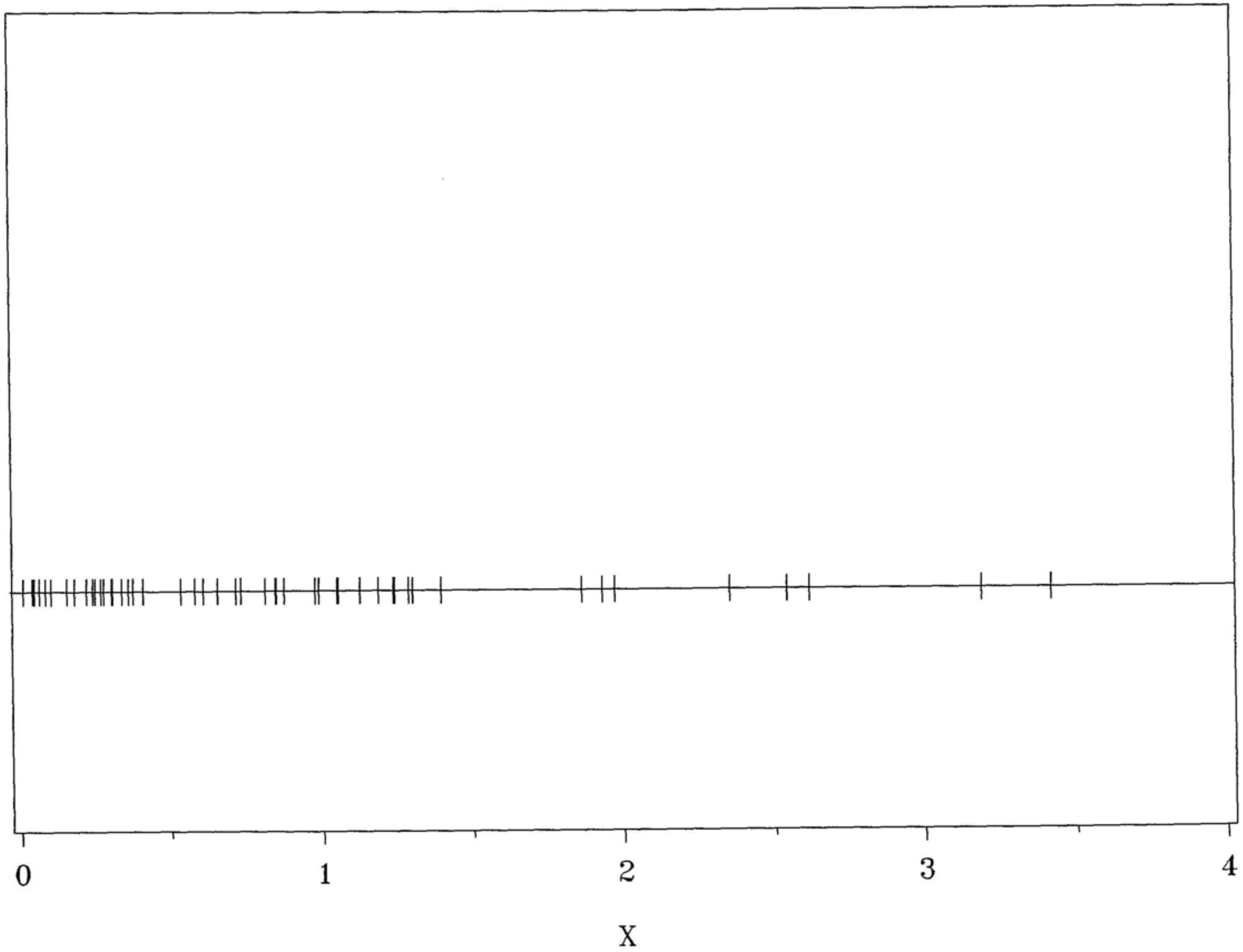
Government wants: probability of flood 1/10,000 in a year.

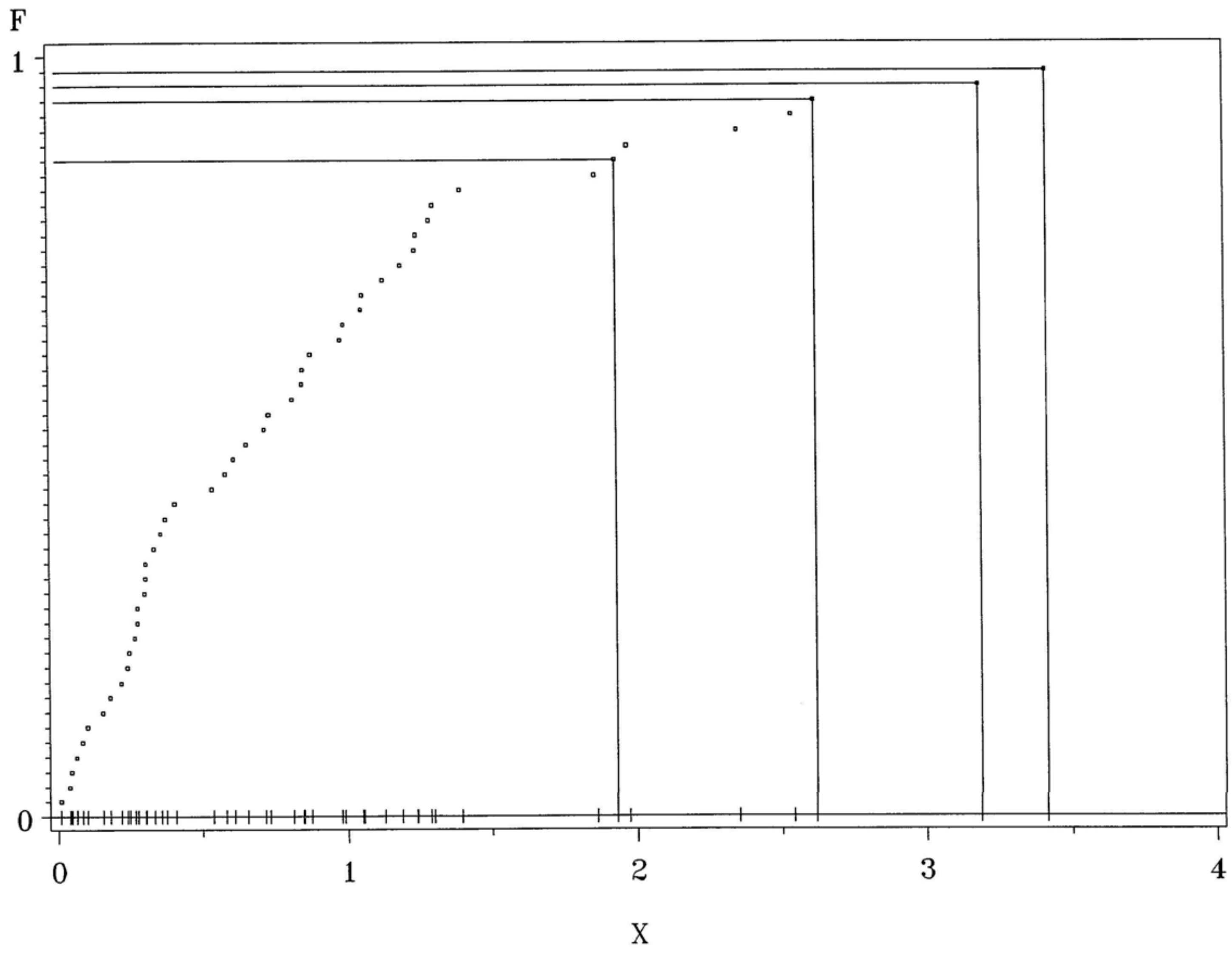
How does this probability translate into a height of the sea wall?

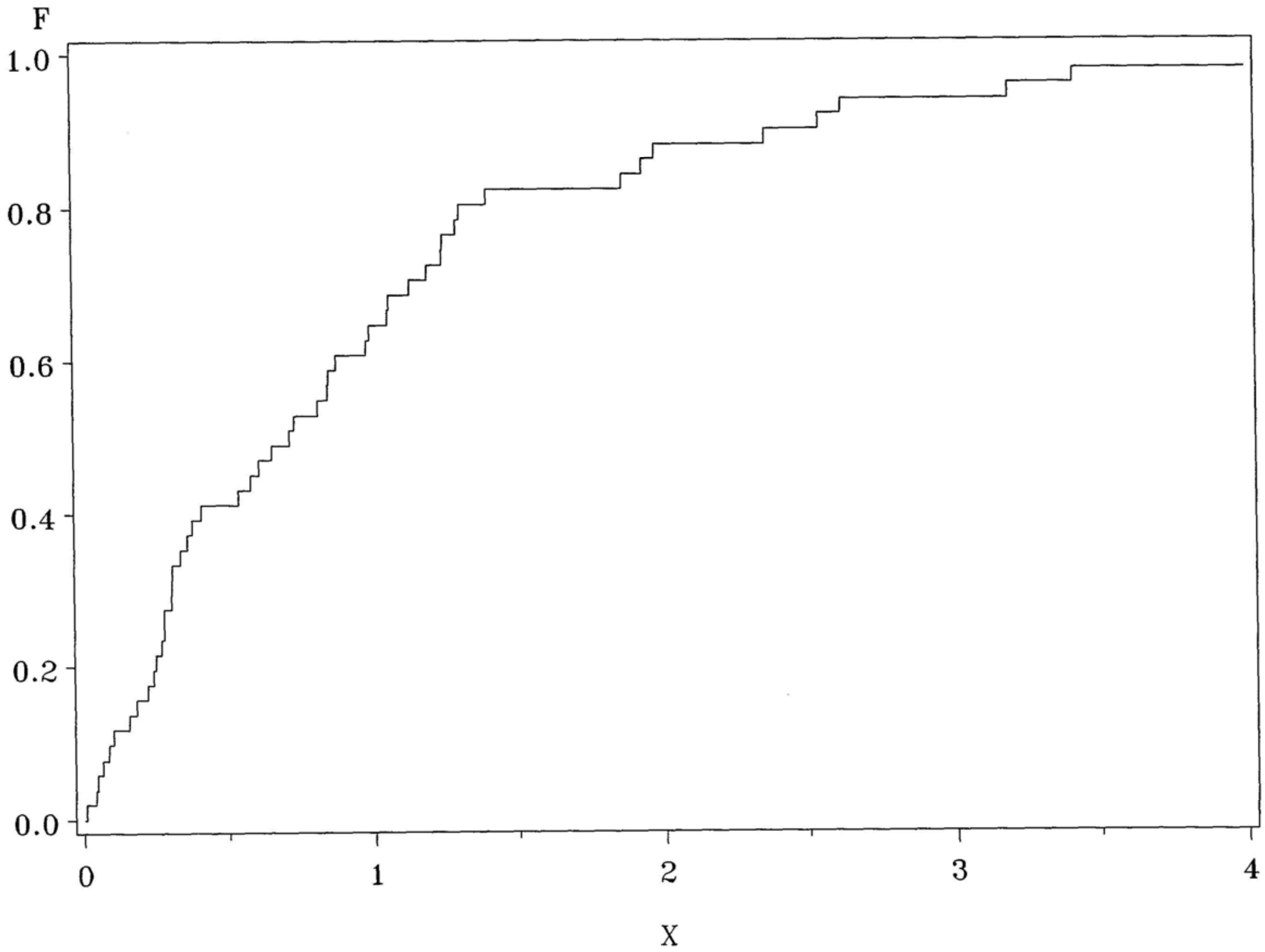
Answer: **EVT**.

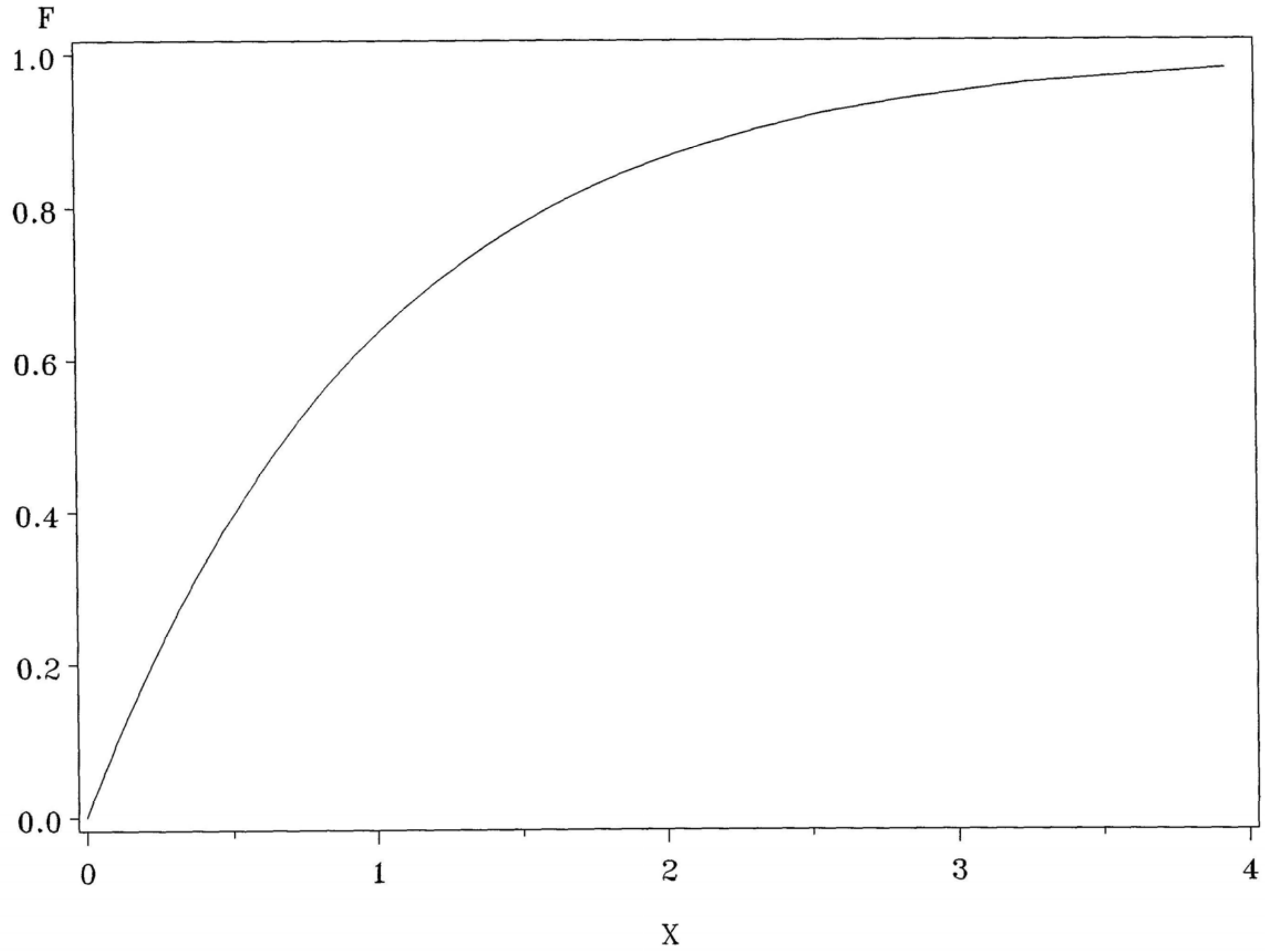
Let us consider the so-called

empirical distribution function

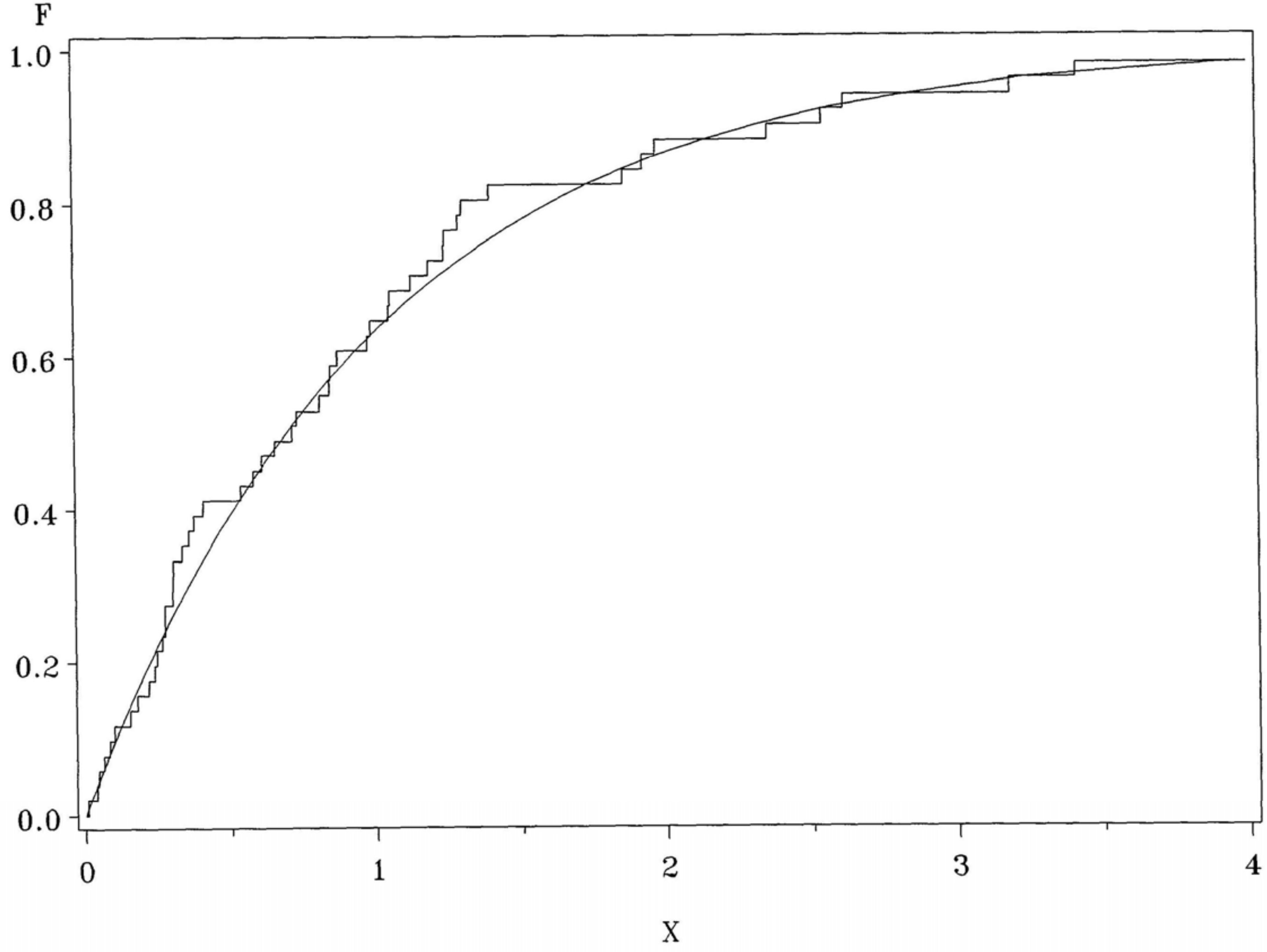

0
1
2
3
4
X


F
1
0
0
1
2
3
4
X

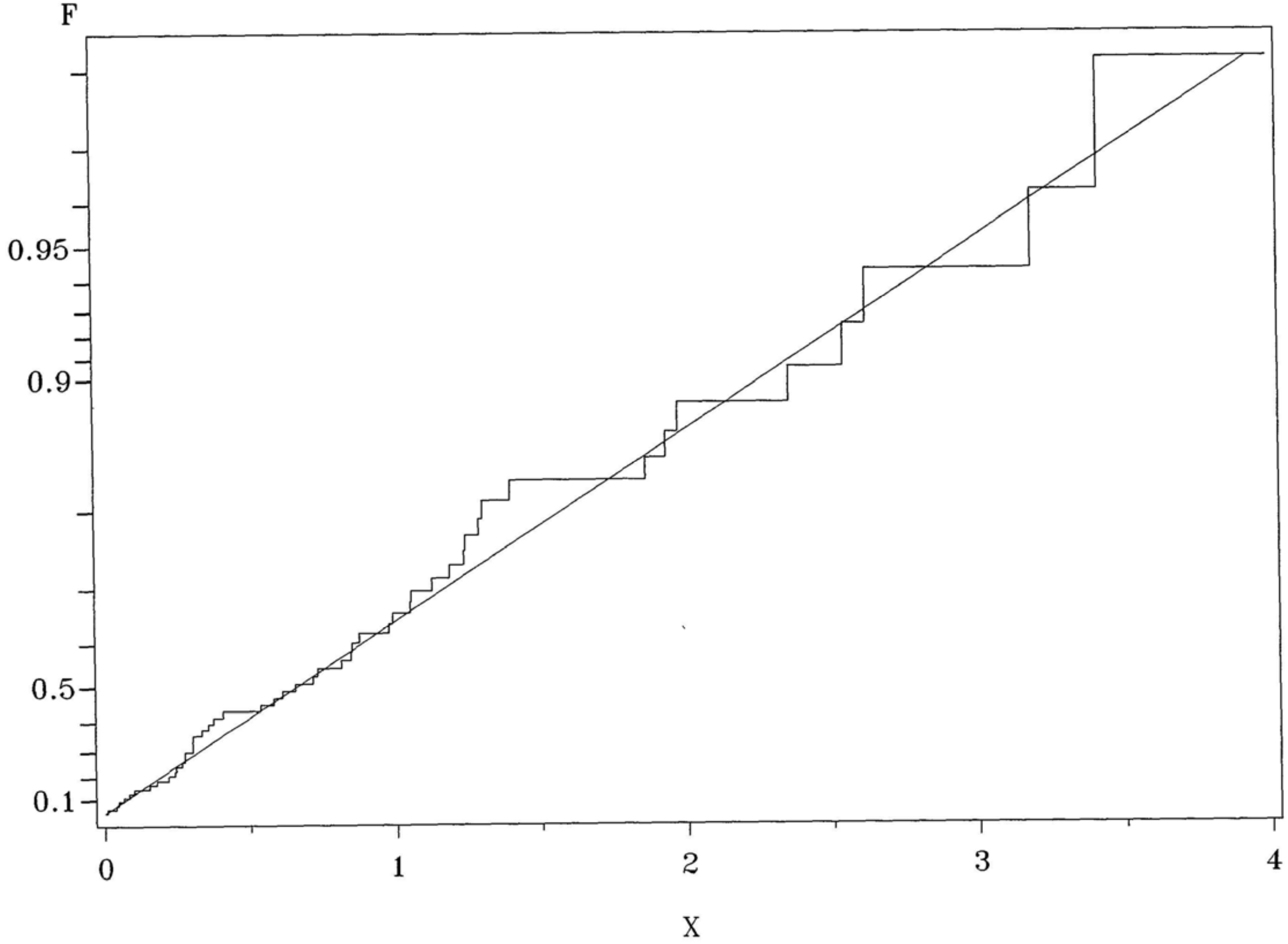

F
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
X


F
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
X

F_n and F



$$\log \frac{1}{1-F_n} \text{ and } \log \frac{1}{1-F}$$



The Dutch government – after the 1953 flood – has decided that the sea dikes should be built so high that the probability of a flood in a given year is $\dfrac{1}{10,000}$.

But we have only 100 years of observations.

$$n = 100 \ , \ p = \frac{1}{10,000} \quad \text{hence} \quad p << \frac{1}{n}.$$

How to extend the graph?

Extreme value theory:

$F^n(a_n x + b_n)$ is the probability distribution function of the maximum of a sample (normalized).

Extreme value condition: suppose that there exists a distribution function (non-degenerate) such that

$$F^n(a_n x + b_n) \to G(x) \qquad (n \to \infty), \qquad \text{all } x.$$

Then essentially

$$G(x) = G_\gamma(x) = \left\{ -(1+\gamma x)^{-1/\gamma} \right\}$$

for some $\gamma \in \mathbb{R}$ (extreme value distributions) .

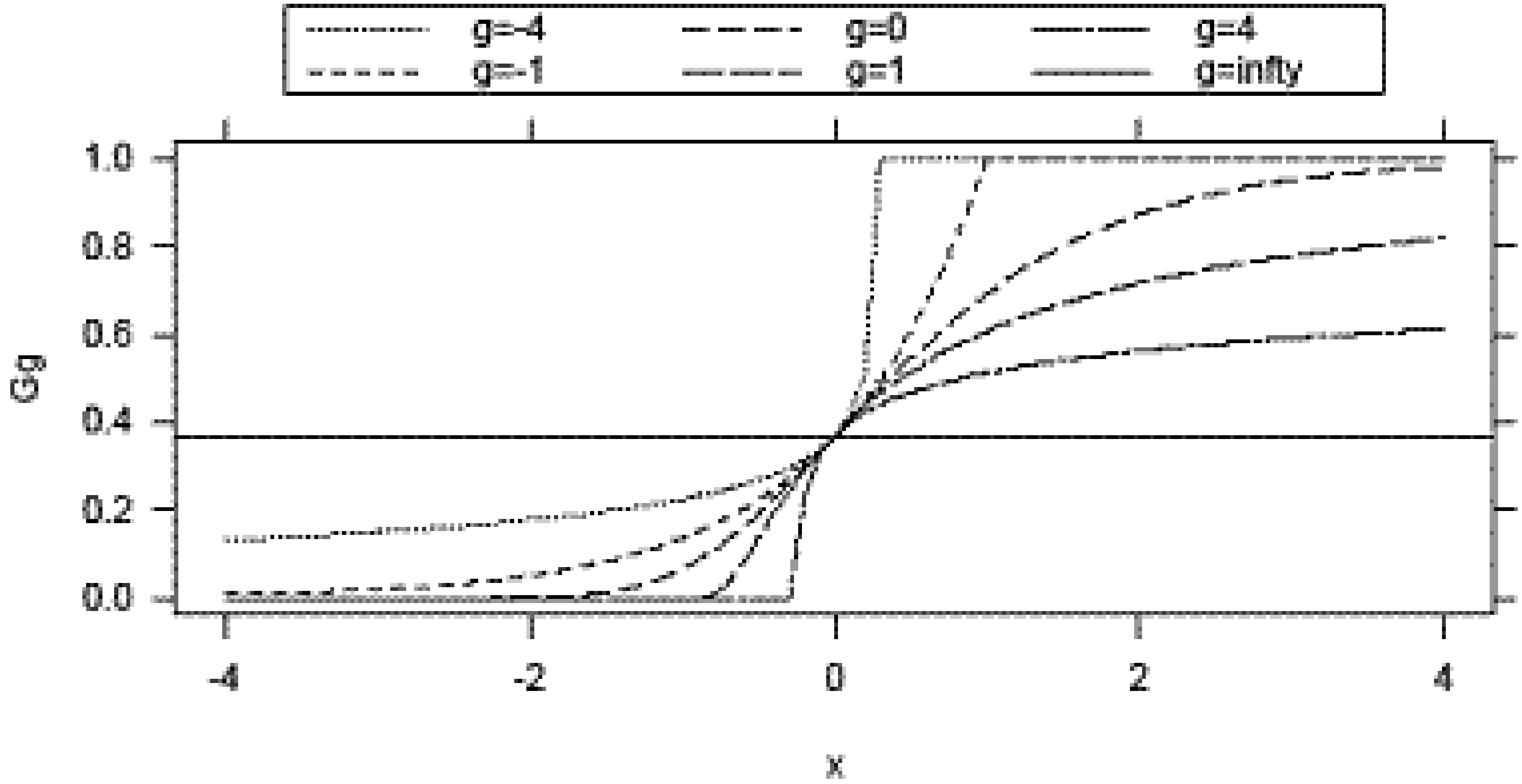
Theorem 1.1.2 *Equivalent are:*

1) $$\lim_{n\to\infty} F^n(a_n x + b_n) = \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\} .$$

2) $$\lim_{t\to\infty} t\left(1 - F\left(b(t) + x a(t)\right)\right) = (1+\gamma x)^{-1/\gamma} .$$

3) $$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} .$$

with $U(t) := F^{\leftarrow}\left(1 - \frac{1}{t}\right)$ *quantile function* .


g=-4
g=0
g=4
g=-1
g=1
g=infty
1.0
0.8
0.6
0.4
0.2
0.0
Gg
-4
-2
0
2
4
x

Now we want to estimate $F^{\leftarrow}(1-p)$ for small p namely $p<\frac{1}{n}$, $n=$ number of observations (extrapolation).

For this problem we use part 3) of the Theorem.

Start from the limit relation

$$\lim_{t\to\infty}\frac{U(tx)-U(t)}{a(t)}=\frac{x^{\gamma}-1}{\gamma}$$

where

$$U(t):=\left(\frac{1}{1-F}\right)^{\leftarrow}(t)=F^{\leftarrow}\left(1-\frac{1}{t}\right).$$

We want to estimate $F^{\leftarrow}(1-p)$ for p small

(i.e. $F^{\leftarrow}(1-p)$ is a high number) .

From limit relation:

$$U(tx) \approx U(t) + a(t)\frac{x^{\gamma}-1}{\gamma}$$

or, (for large y and t, generally we use it for $y > t$)

$$U(y) \approx U(t) + a(t)\frac{(y/t)^{\gamma}-1}{\gamma}\,.$$

In the relation

$$U(y) \approx U(t) + a(t)\frac{(y/t)^{\gamma} - 1}{\gamma}$$

we choose $y := \frac{1}{p}$ and $t = \frac{n}{k}$.

We get :

$$F^{\leftarrow}(1-p) \approx U\left(\frac{n}{k}\right) + a\left(\frac{n}{k}\right) \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma}.$$

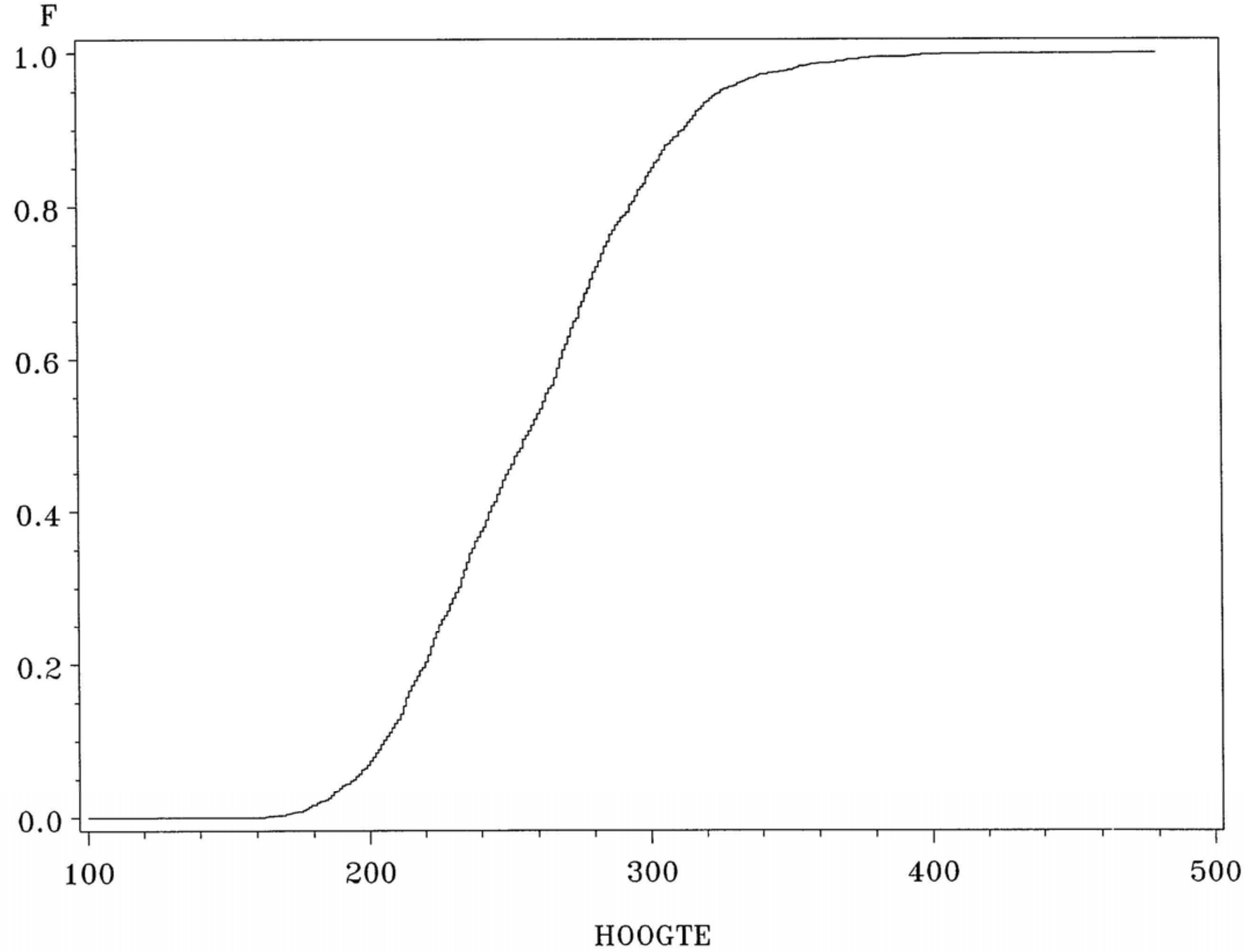
Gives estimator for $F^{\leftarrow}(1-p)$ by extrapolation, since $F^{\leftarrow}(1-p)$ is a quantile **outside** the sample and $U\left(\frac{n}{k}\right)=F^{\leftarrow}\left(1-\frac{k}{n}\right)$ is a quantile **inside** the sample.

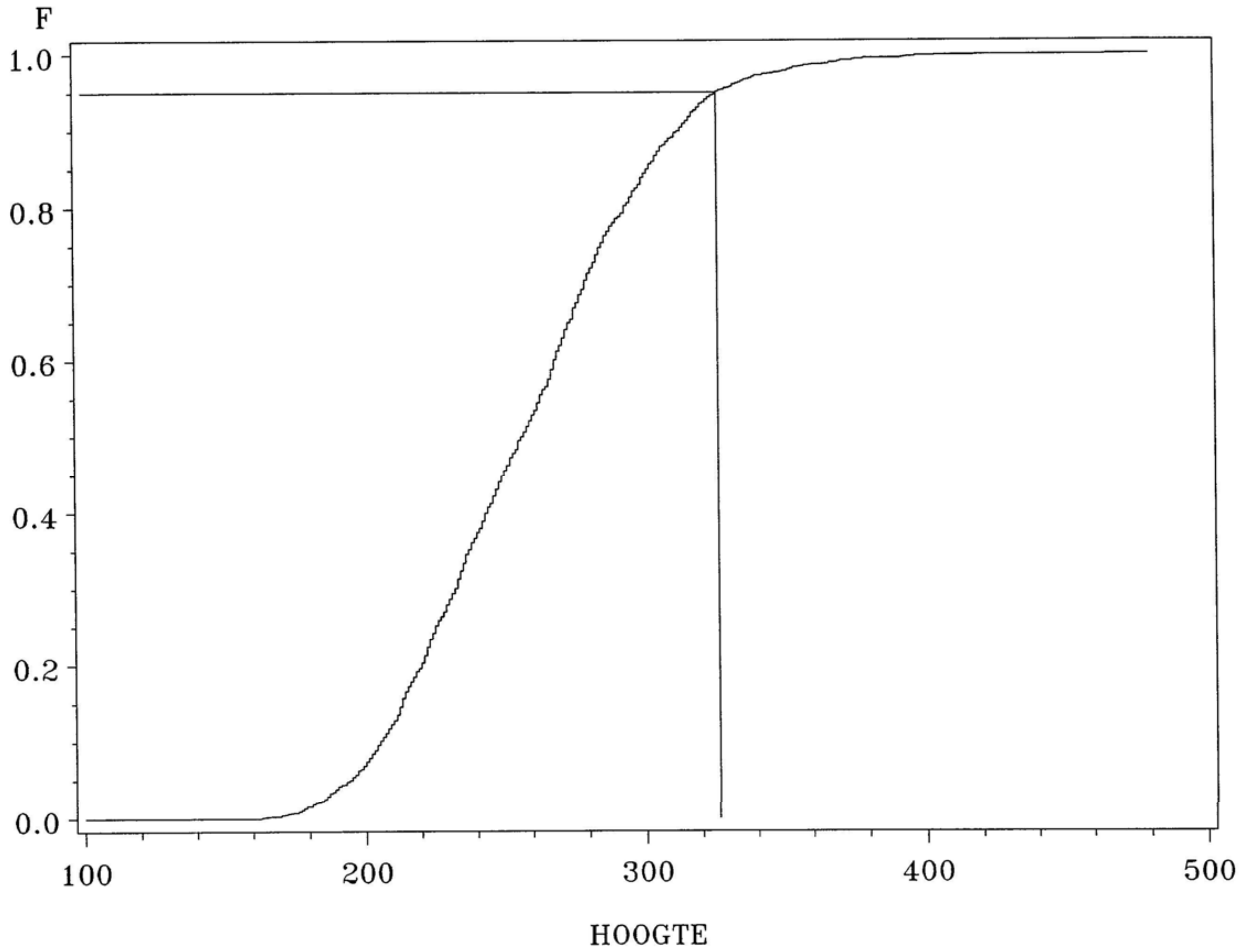
Need estimators for $U\left(\frac{n}{k}\right)$, $a\left(\frac{n}{k}\right)$ and γ .

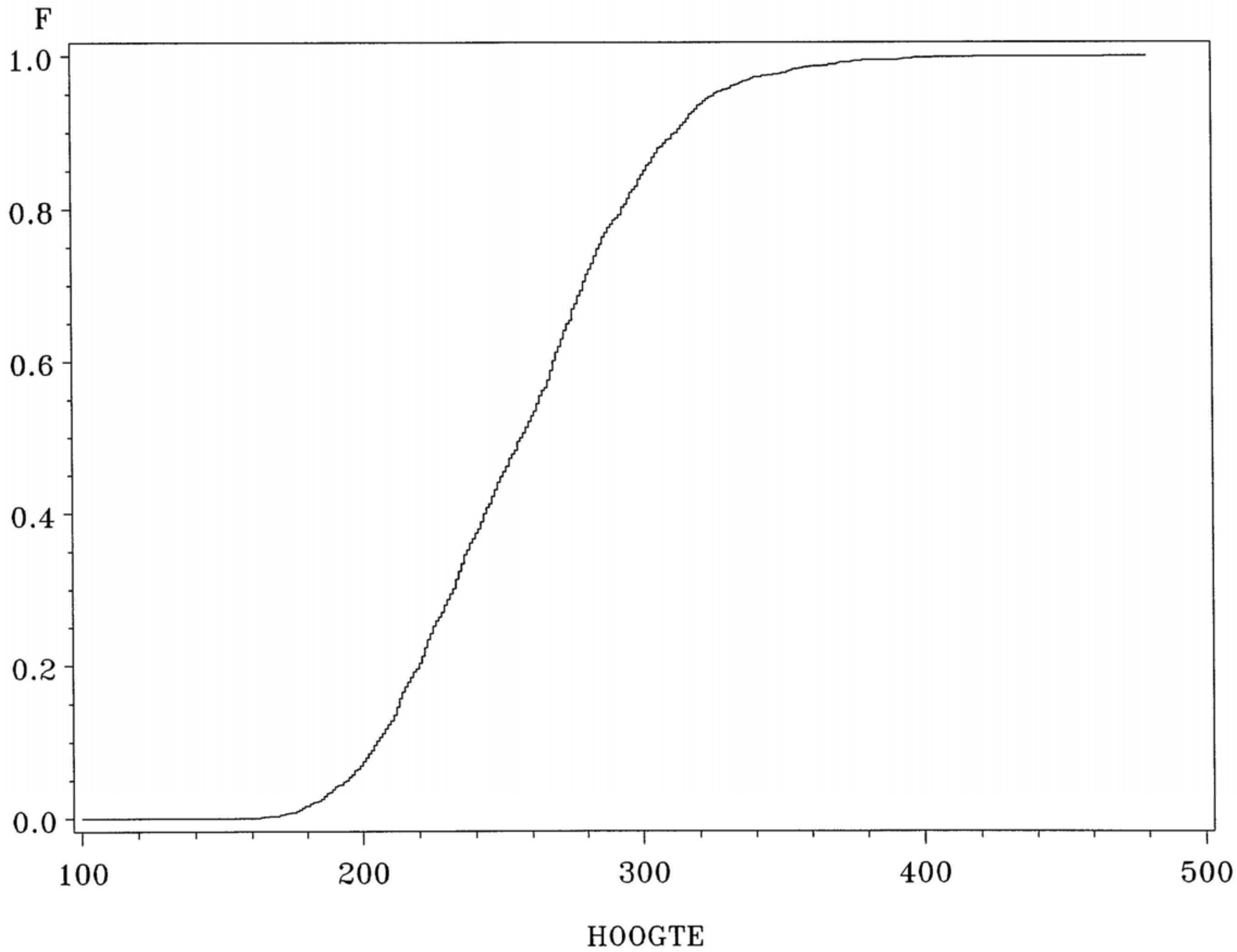
Estimator of quantile $U\left(\frac{n}{k}\right)=F^{\leftarrow}\left(1-\frac{k}{n}\right)$ is the order statistic $X_{n-k,n}$ (k –th largest observation).

I do not discuss the estimators of $a\left(\frac{n}{k}\right)$ and γ .

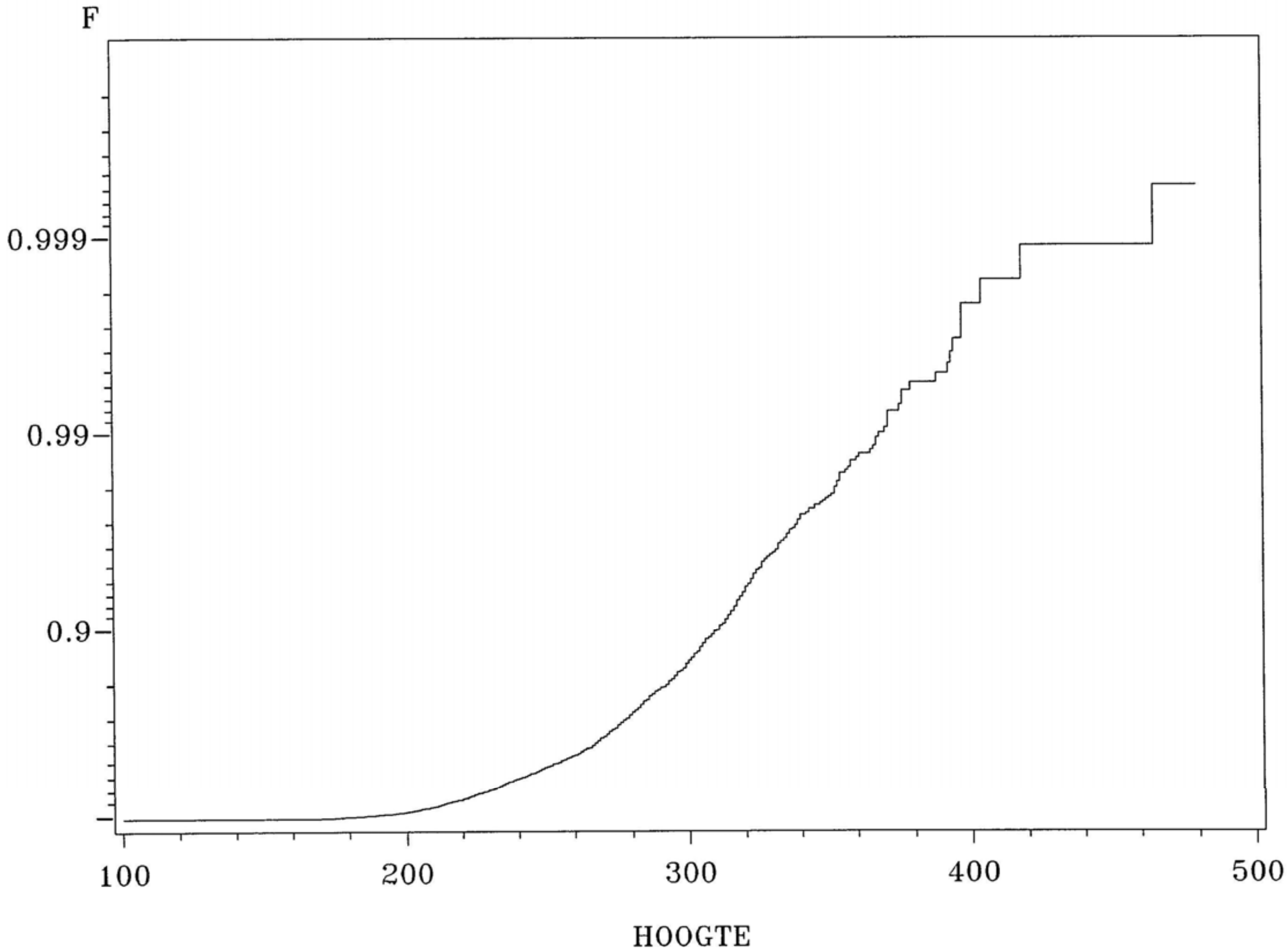
Empirical distribution function F_n of observed water levels



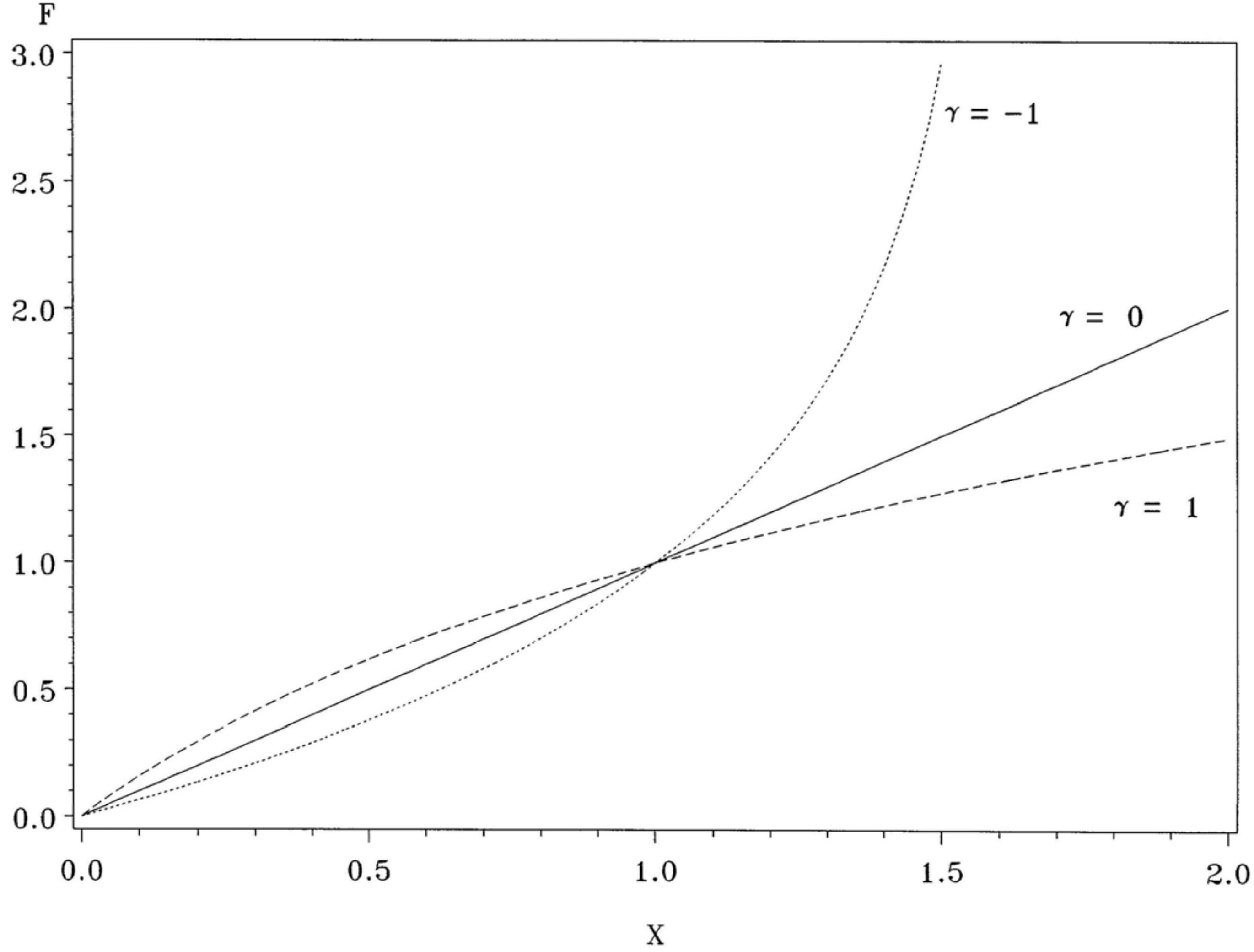

F
1.0
0.8
0.6
0.4
0.2
0.0
100
200
300
400
500
HOOGTE

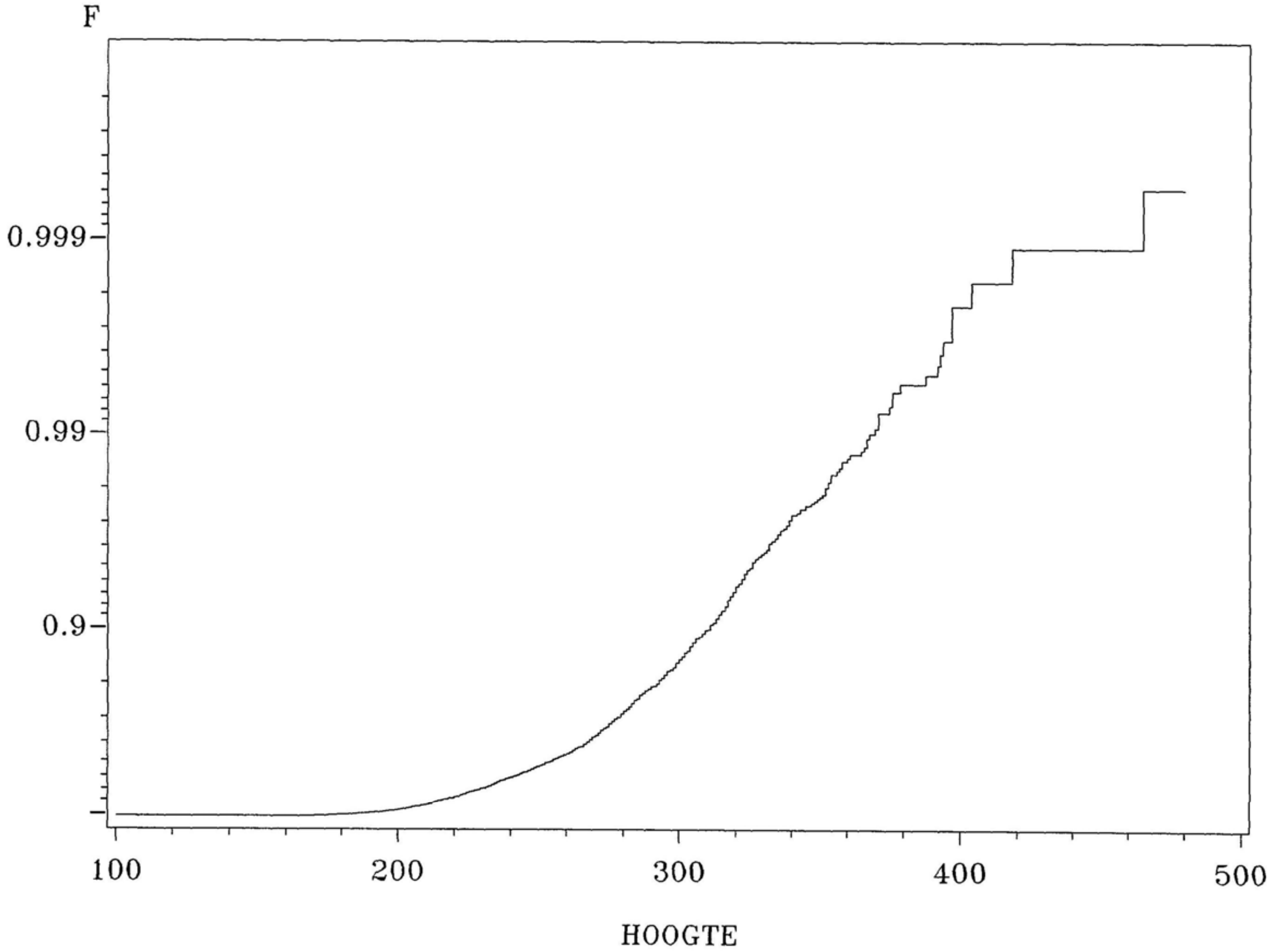

F
1.0
0.8
0.6
0.4
0.2
0.0
100
200
300
400
500
HOOGTE

$$\log \frac{1}{1-F_n}$$

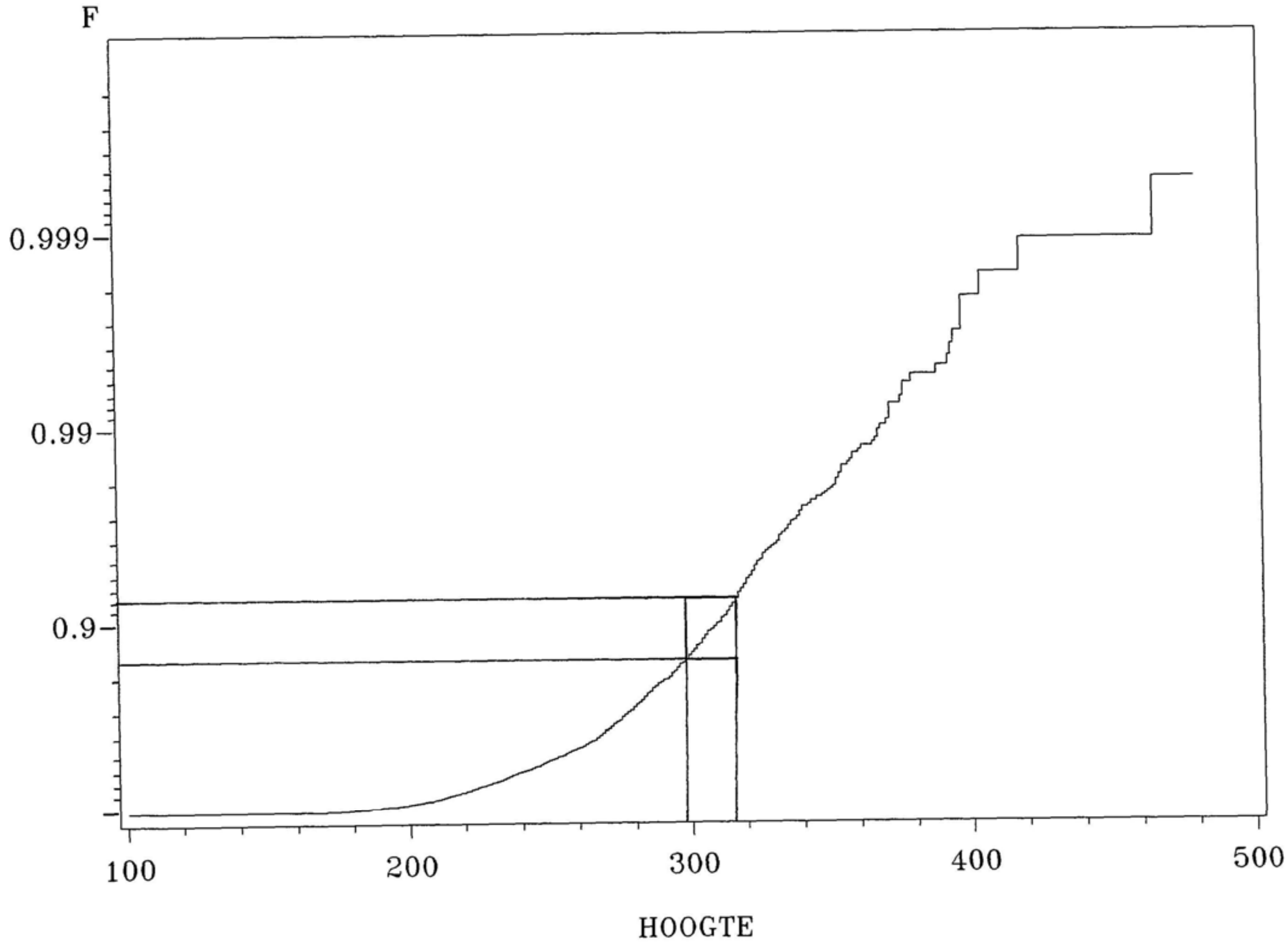


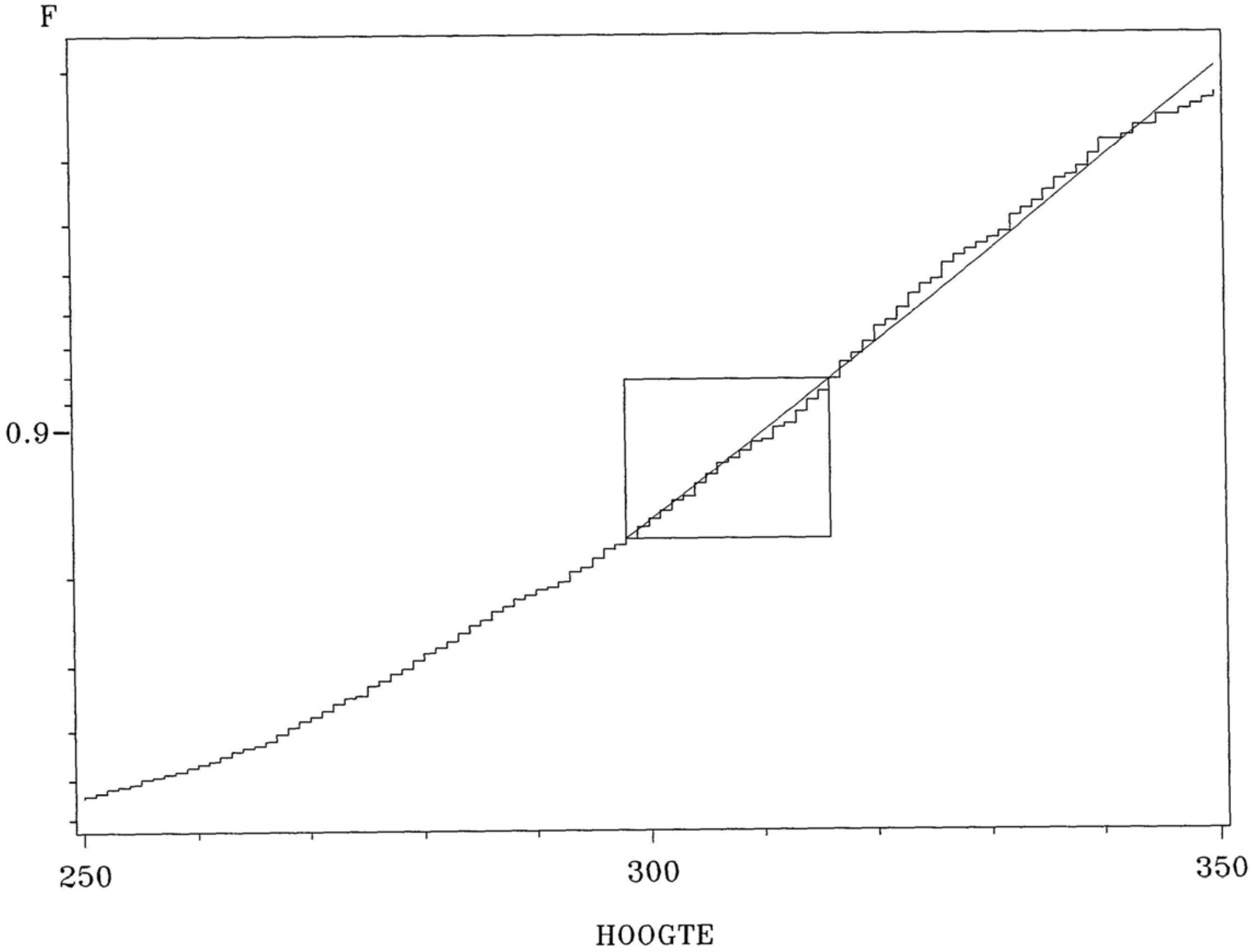
Limit functions from the Theorem

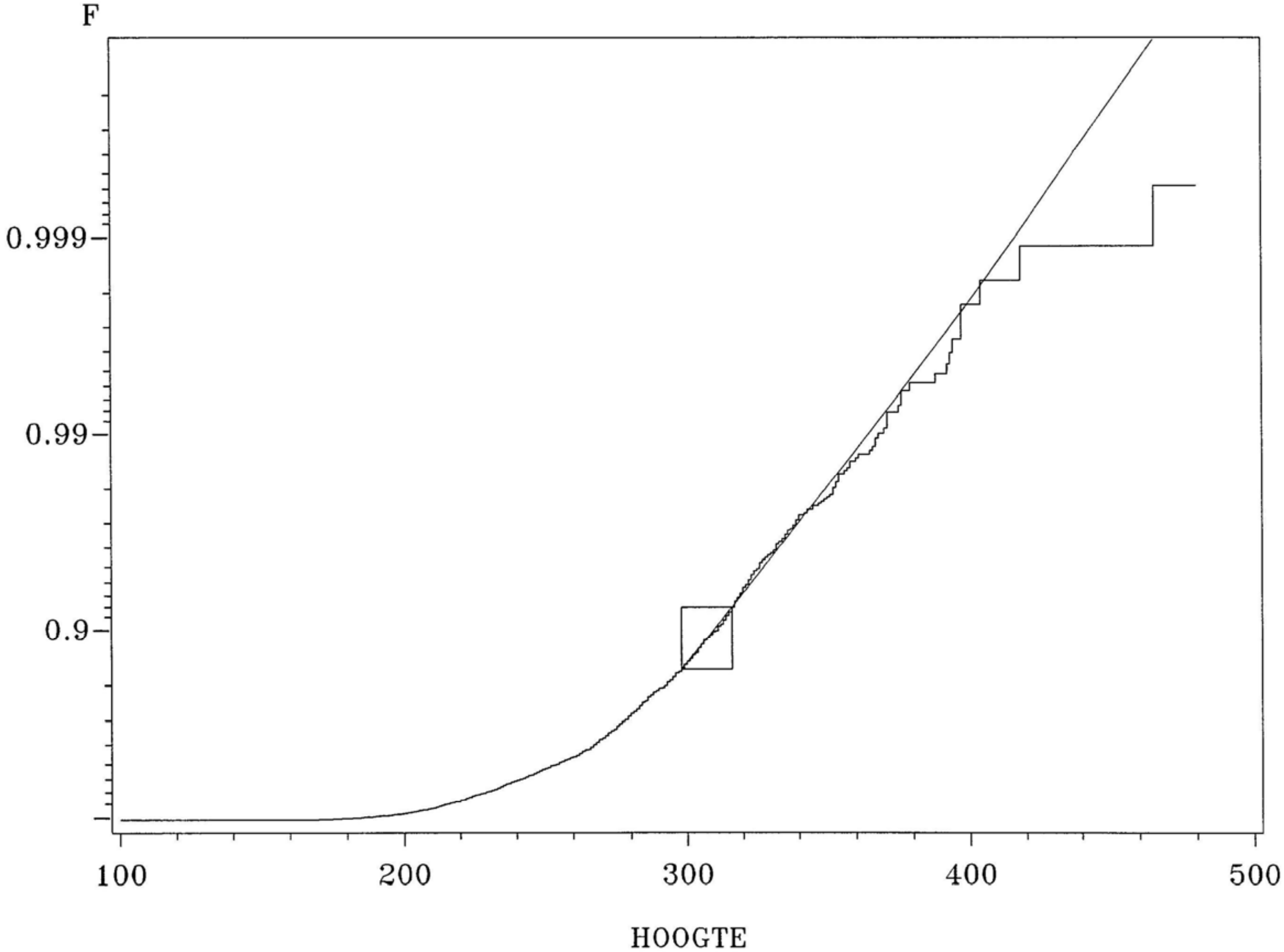


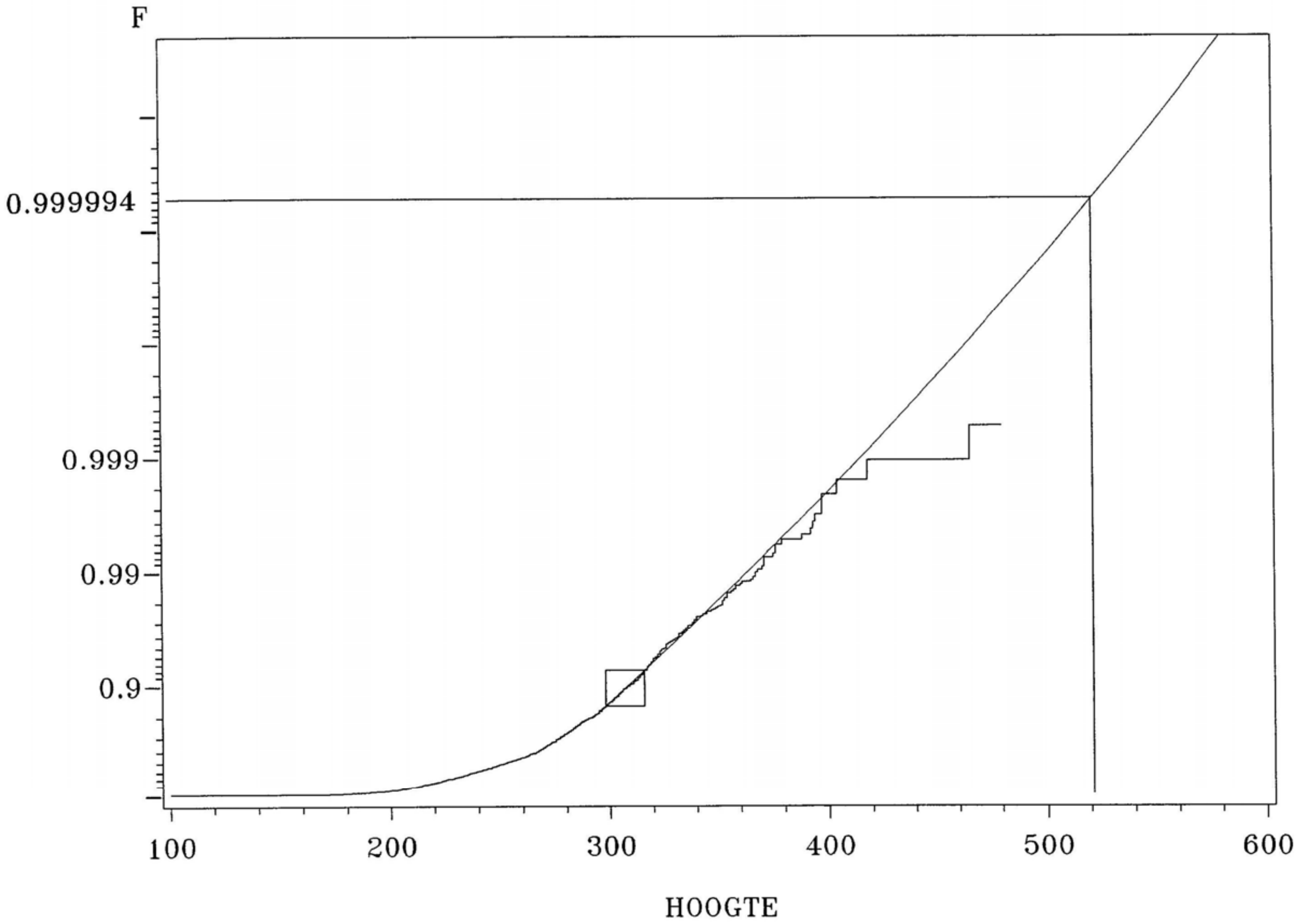

F
0.999
0.99
0.9
100
200
300
400
500
HOOGTE

How to connect to empirical distribution function

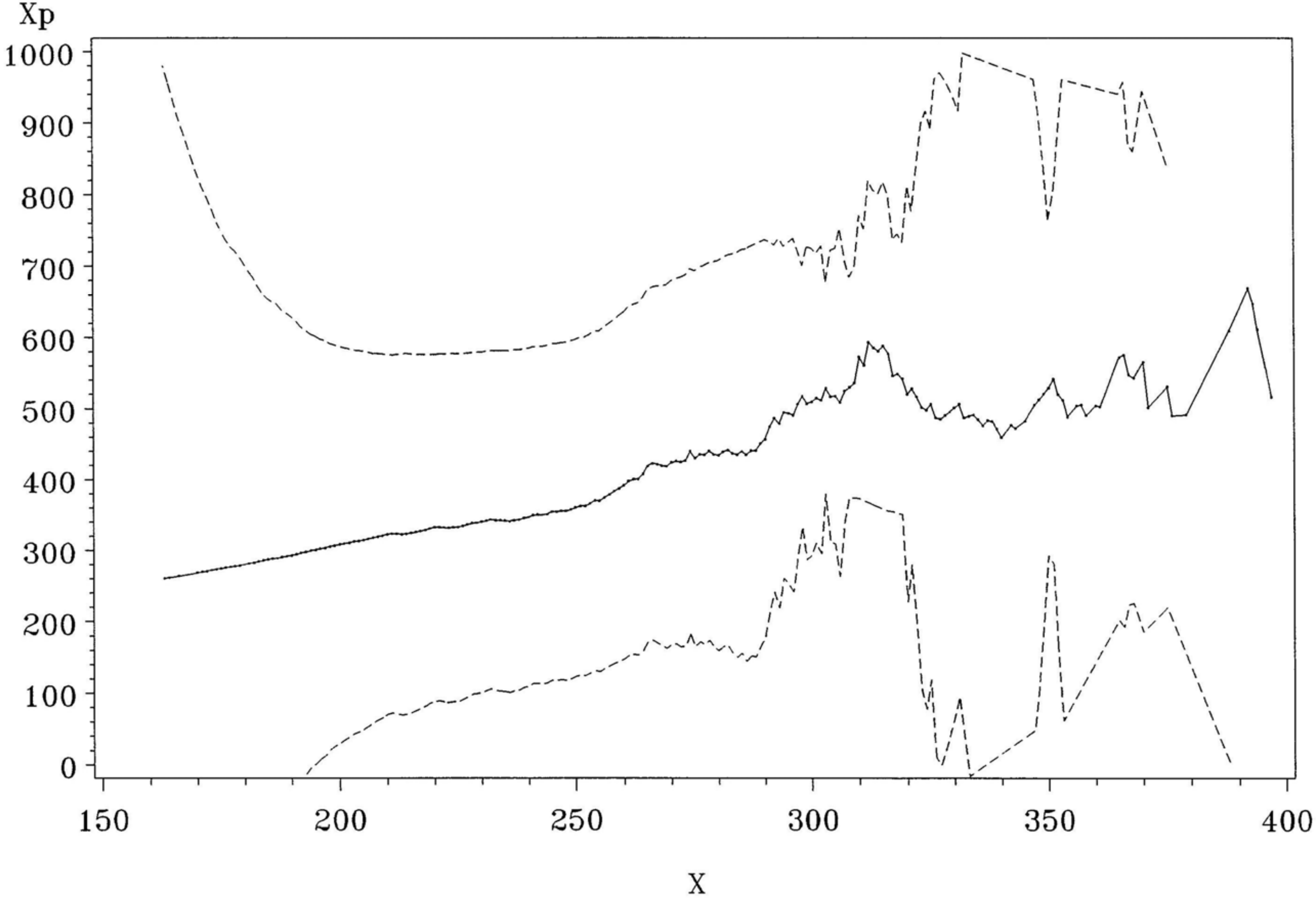



F
0.9
250
300
350
HOOGTE


F
0.999
0.99
0.9
100
200
300
400
500
HOOGTE


F
0.999994
0.999
0.99
0.9
100
200
300
400
500
600
HOOGTE

Estimated height of sea wall (centimetres)



Multidimensional extension

A simple example

- Take r.v.'s (R, Φ), independent, $\begin{cases} \Phi \in \left[0, \frac{\pi}{2}\right] \\ P\{R > r = 1/r, r > 1\} \end{cases}$

  and $(X,Y) := (R \cos \Phi, R \sin \Phi)$.

- Take a Borel set $A \subset \mathbb{R}_+^2$ with positive distance to the origin.

- Write $aA := \{a\mathbf{x} : \mathbf{x} \in A\}$.

- Clearly $(r = ar')$

$$P\{aA\} = \int_{(r\cos\phi, r\sin\phi)\in aA} f(\phi)\,d\phi\,\frac{dr}{r^2}$$

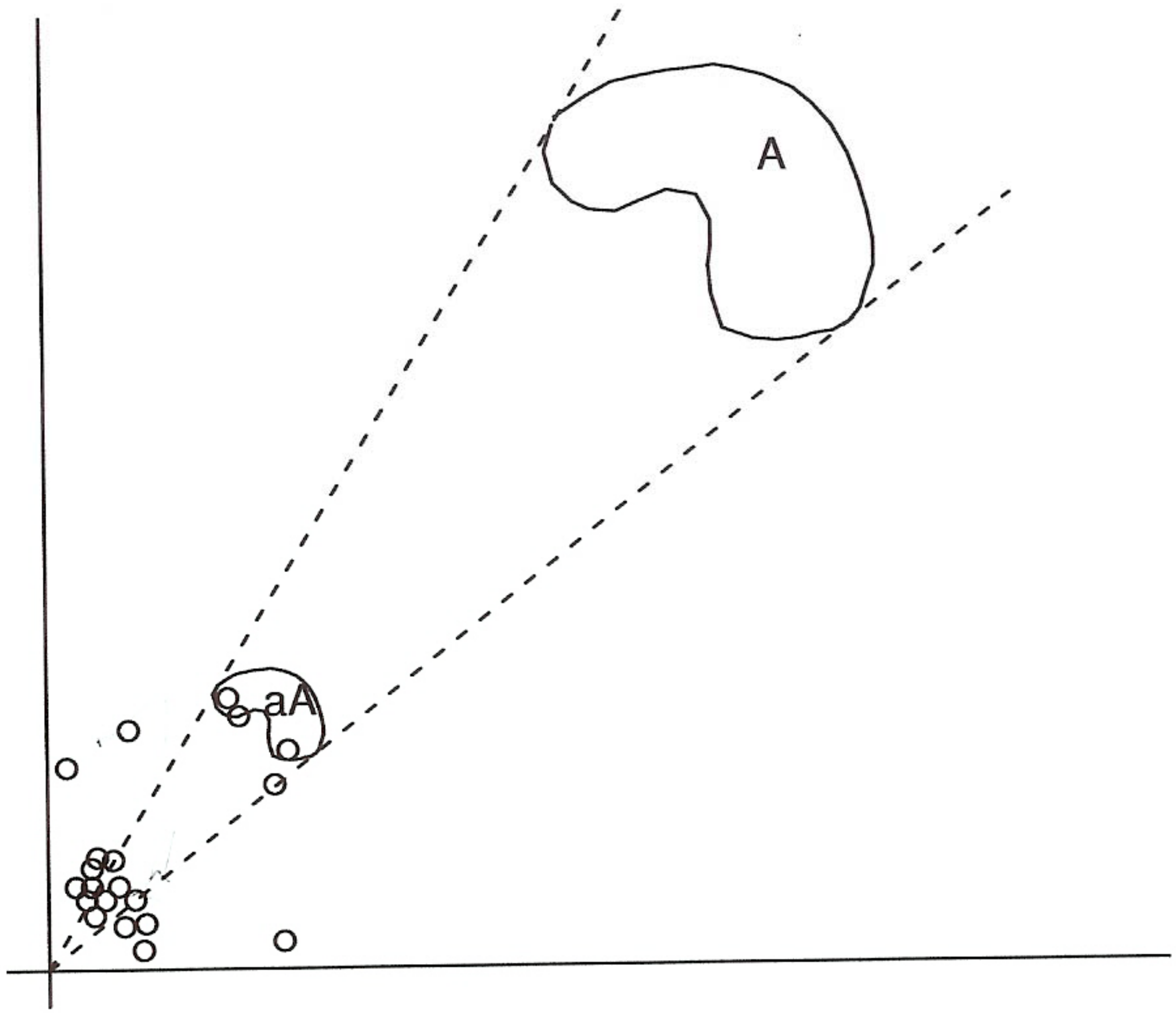
$$= \frac{1}{a}\int_{(r'\cos\phi, r'\sin\phi)\in A} f(\phi)\,d\phi\,\frac{dr'}{(r')^2} = \frac{1}{a}P\{A\}.$$

- Suppose: probability distribution of Φ unknown.
- We have i.i.d. observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ and a failure set A away from the observations in the NE corner.
- To estimate $P\{A\}$ we may use

$$a\widehat{P}\{aA\}$$

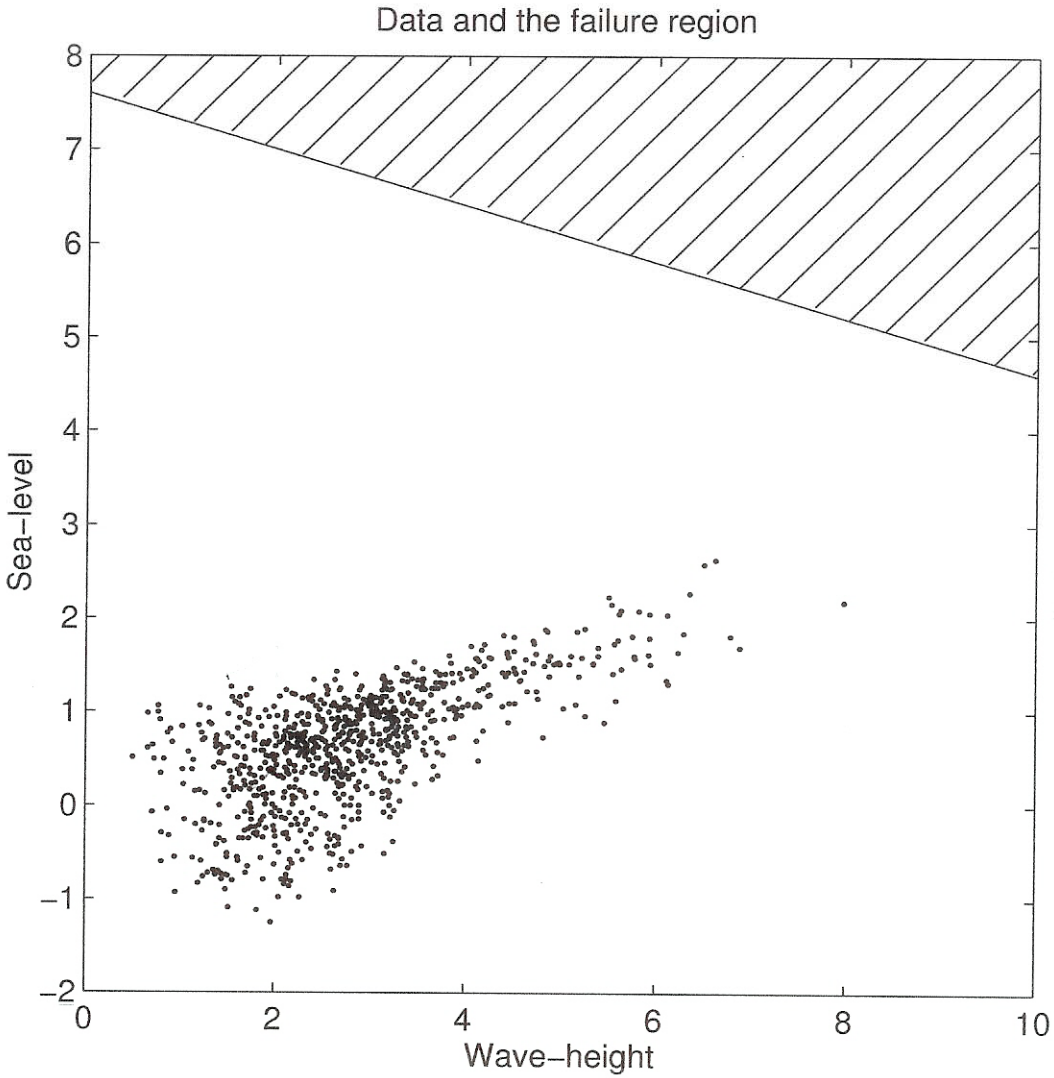
  where $\widehat{P}$ is the empirical measure.

This is the main idea of estimation of failure set probability.


A
aA

The problem:

- Some device can fail under the combined influence of extreme behaviour of two random forces X and Y. For example: wind and waves.
- “Failure set” C: if (X, Y) falls into C, then failure takes place.
- “Extreme failure set”: none of the observations we have from the past falls into C. There has never been a failure.
- Estimate the probability of “extreme failure”


Data and the failure region
Sea−level
Wave−height
8
7
6
5
4
3
2
1
0
−1
−2
0
2
4
6
8
10

A bit more formal

- Suppose we have n i.i.d. observations $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ with distribution function F and a failure set C.
- The fact " none of the n observations is in C " can be reflected in the theoretical assumption

$$P(C) < 1/n \ .$$

Hence C can not be fixed, we have

$C = C_n$ and $P(C_n) = O(1/n)$ as $n \to \infty$.

i.e. when n increases the set C moves, say, to the NE corner.

Domain of attraction condition EVT

There exist

- Functions $a_1, a_2 > 0, b_1, b_2$ real
- Parameters γ_1 and γ_2
- A measure υ on the positive quadrant such that

$$\upsilon(aA) = a^{-1}\upsilon(A) \qquad (1)$$

for each Borel set $A,$ such that

$$\lim_{t\to\infty} tP\left\{\left(\left(1+\gamma_1 \frac{X-b_1(t)}{a_1(t)}\right)^{\frac{1}{\gamma_1}}, \left(1+\gamma_2 \frac{Y-b_2(t)}{a_2(t)}\right)^{\frac{1}{\gamma_2}}\right) \in A\right\} = \upsilon(A)$$

for all such A with positive distance to the origin.

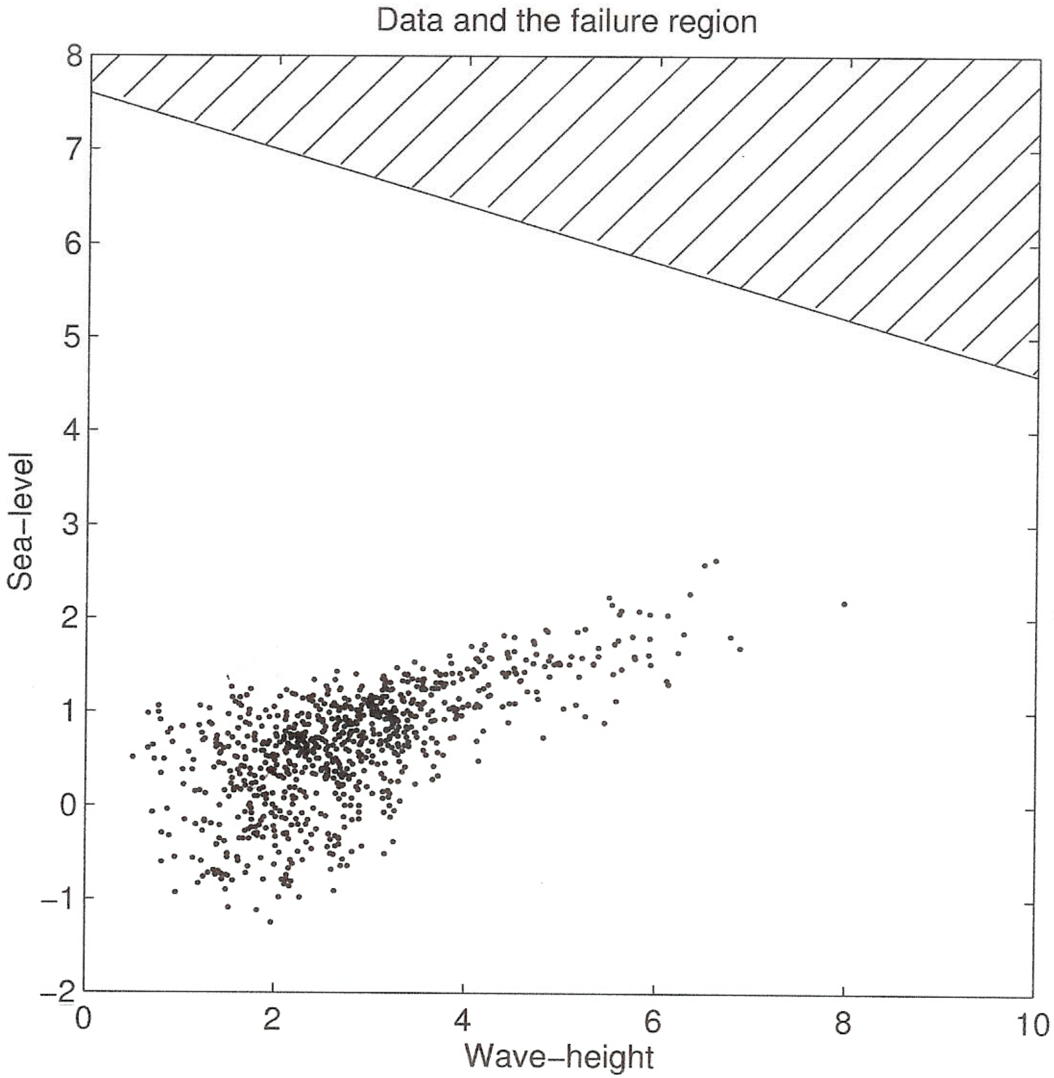
Remark

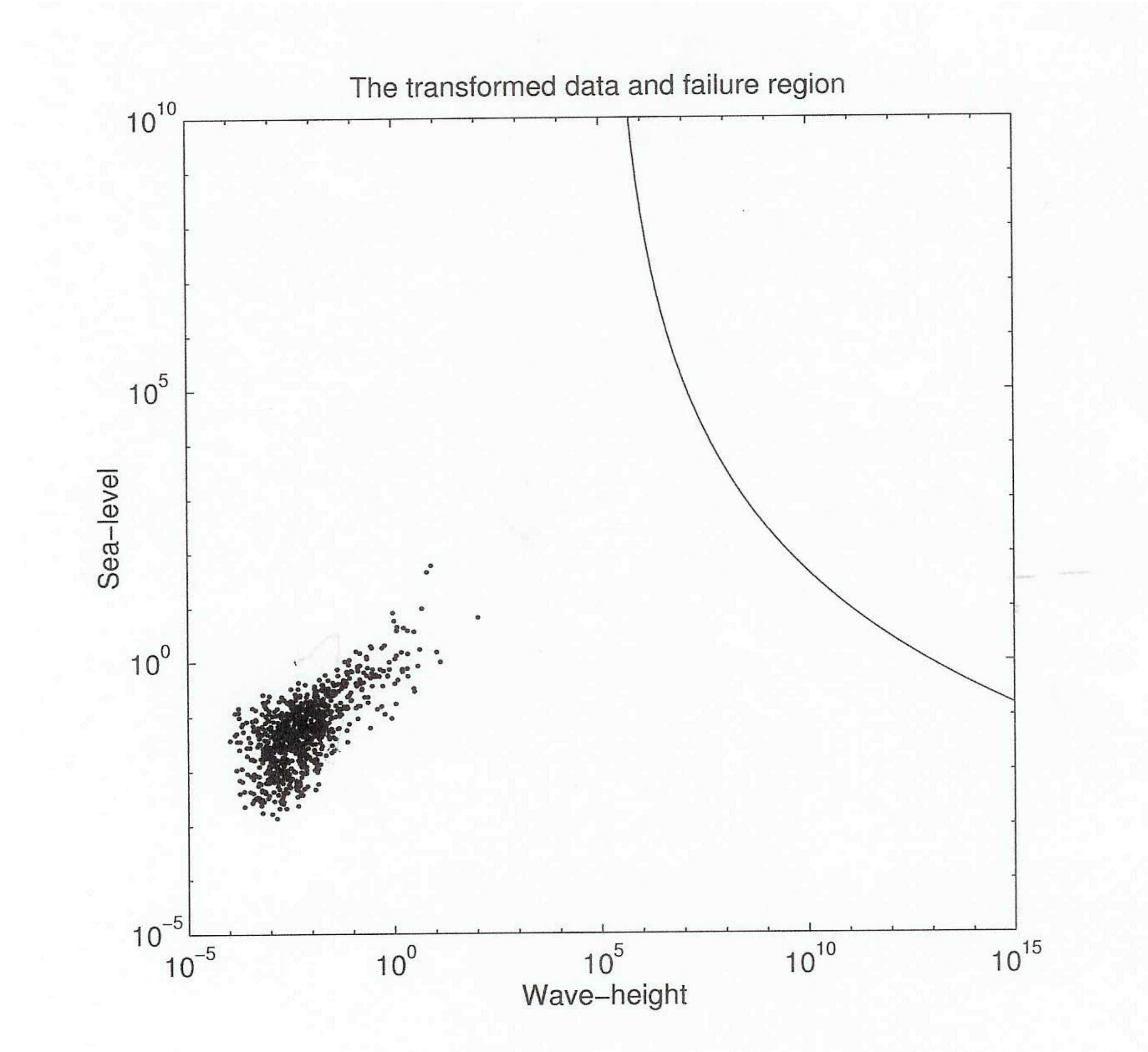
Relation (1) is as in the example (homogeneity).

But here we have the marginal transformations

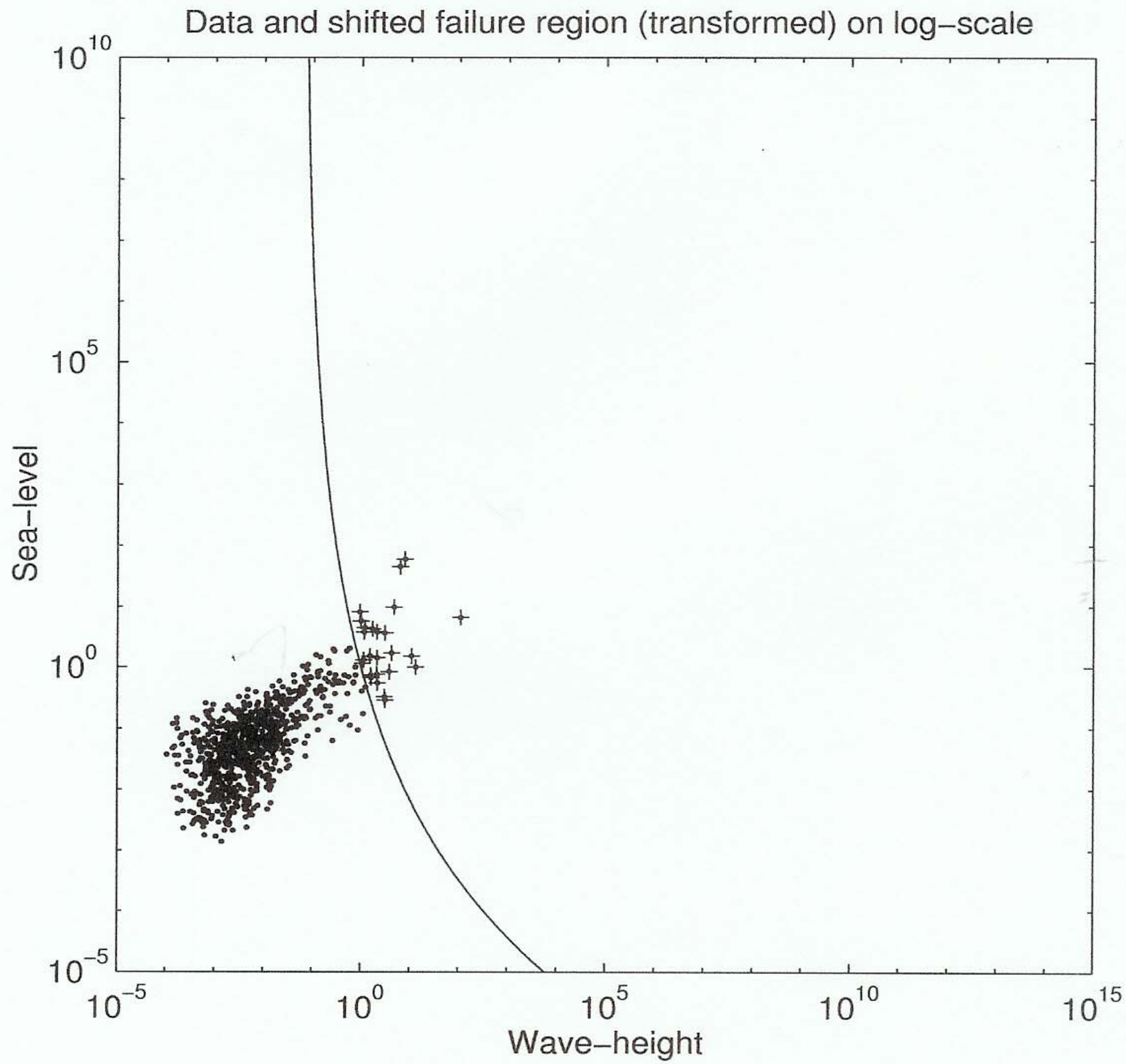
$$\left(1+\gamma_1 \frac{x-b_i}{a_i}\right)^{\frac{1}{\gamma_i}}$$

on top of that.


Data and the failure region
Sea–level
Wave–height
8
7
6
5
4
3
2
1
0
−1
−2
0
2
4
6
8
10


The transformed data and failure region
Sea–level
Wave–height
10^{10}
10^{5}
10^{0}
10^{-5}
10^{-5}
10^{0}
10^{5}
10^{10}
10^{15}

(log scale)


Data and shifted failure region (transformed) on log−scale
10^10
10^5
10^0
10^−5
Sea−level
10^−5
10^0
10^5
10^10
10^15
Wave−height

Hence three steps:

1) Transformation of marginal distributions
2) Use of homogeneity property of υ when pulling back the failure set.
3) Estimate probability of pulled back failure set by counting the number of observations in it.

Conditions

Condition 1 Domain of attraction:

$$\lim_{t\to\infty} tP\left\{\left(\left(1+\gamma_1\frac{X-b_1(t)}{a_1(t)}\right)^{\frac{1}{\gamma_1}},\left(1+\gamma_2\frac{Y-b_2(t)}{a_2(t)}\right)^{\frac{1}{\gamma_2}}\right)\in A\right\}=\upsilon(A)$$

Condition 2 We need estimators $\hat{\gamma}_i, \hat{a}_i\left(\frac{n}{k}\right), \hat{b}_i\left(\frac{n}{k}\right)$ with

$$\sqrt{k}\left(\hat{\gamma}_i-\gamma_i, \frac{\hat{a}_i\left(\frac{n}{k}\right)}{a_i\left(\frac{n}{k}\right)}-1, \frac{\hat{b}_i\left(\frac{n}{k}\right)-b_i\left(\frac{n}{k}\right)}{a_i\left(\frac{n}{k}\right)}\right)=\left(O_p(1),O_p(1),O_p(1)\right)$$

for $i=1,2$ with $k=k(n)\to\infty,\ k/n\to 0,\ n\to\infty$.

Condition 3 C_n is open and there exists $(v_n, w_n) \in \partial C_n$ such that $(x,y) \in C_n \Rightarrow x > v_n$ or $y > w_n$.

Condition 4 (stability condition on C_n). The set

$$S := \left\{ \left(\frac{1}{c_n} \left(1 + \gamma_1 \frac{x - b_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)} \right)^{\frac{1}{\gamma 1}}, \frac{1}{c_n} \left(1 + \gamma_2 \frac{y - b_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)} \right)^{\frac{1}{\gamma 2}} \right) : (x,y) \in C_n \right\}$$

(2)

in $\mathbb{R}_+^2$ does not depend on n where

$c_n := \sqrt{q_n^2 + r_n^2} \to \infty$ ($1/c_n$ is pull back factor)

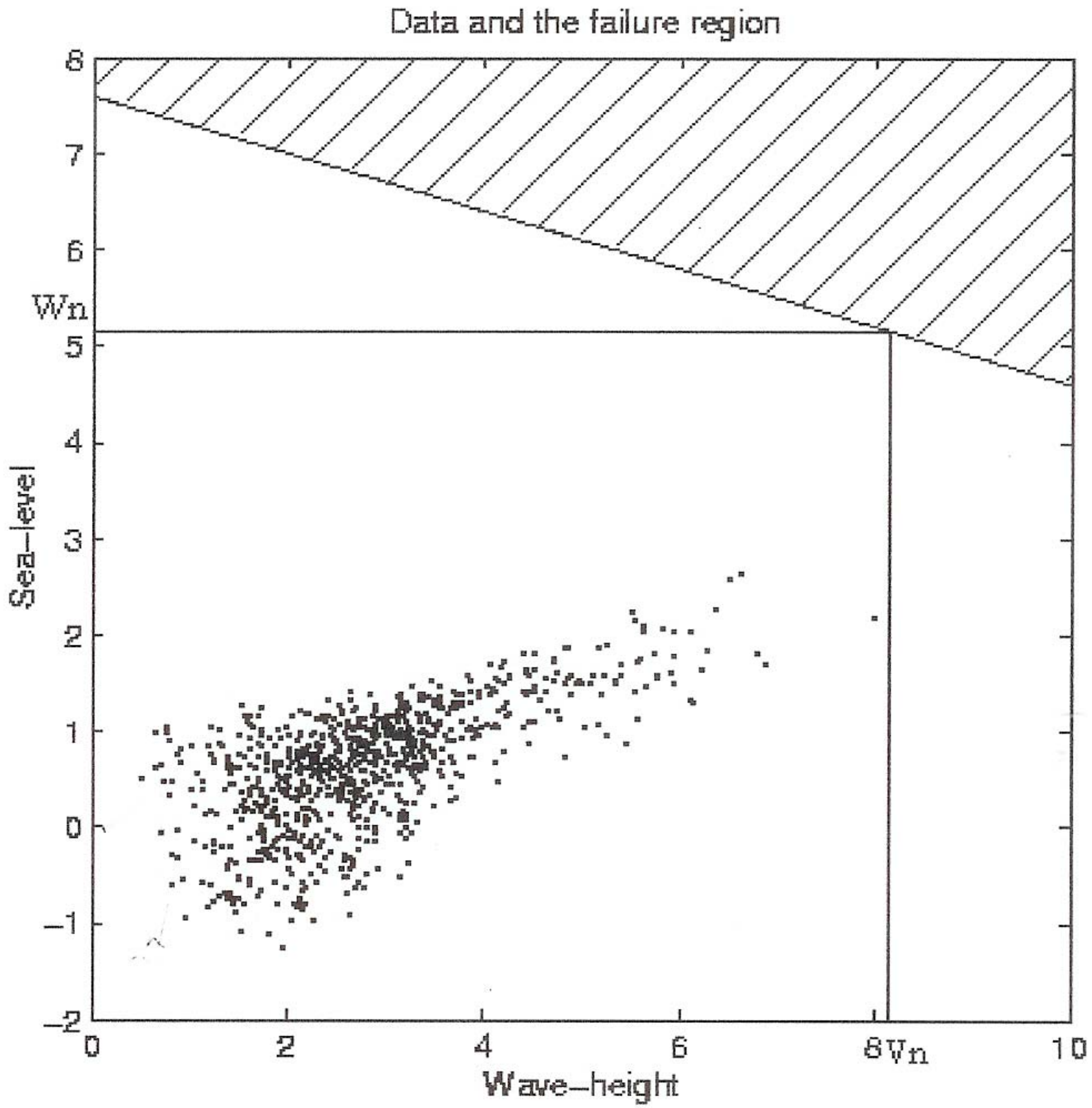
$$q_n := \left(1 + \gamma_1 \frac{v_n - b_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)} \right)^{\frac{1}{\gamma_1}} .$$

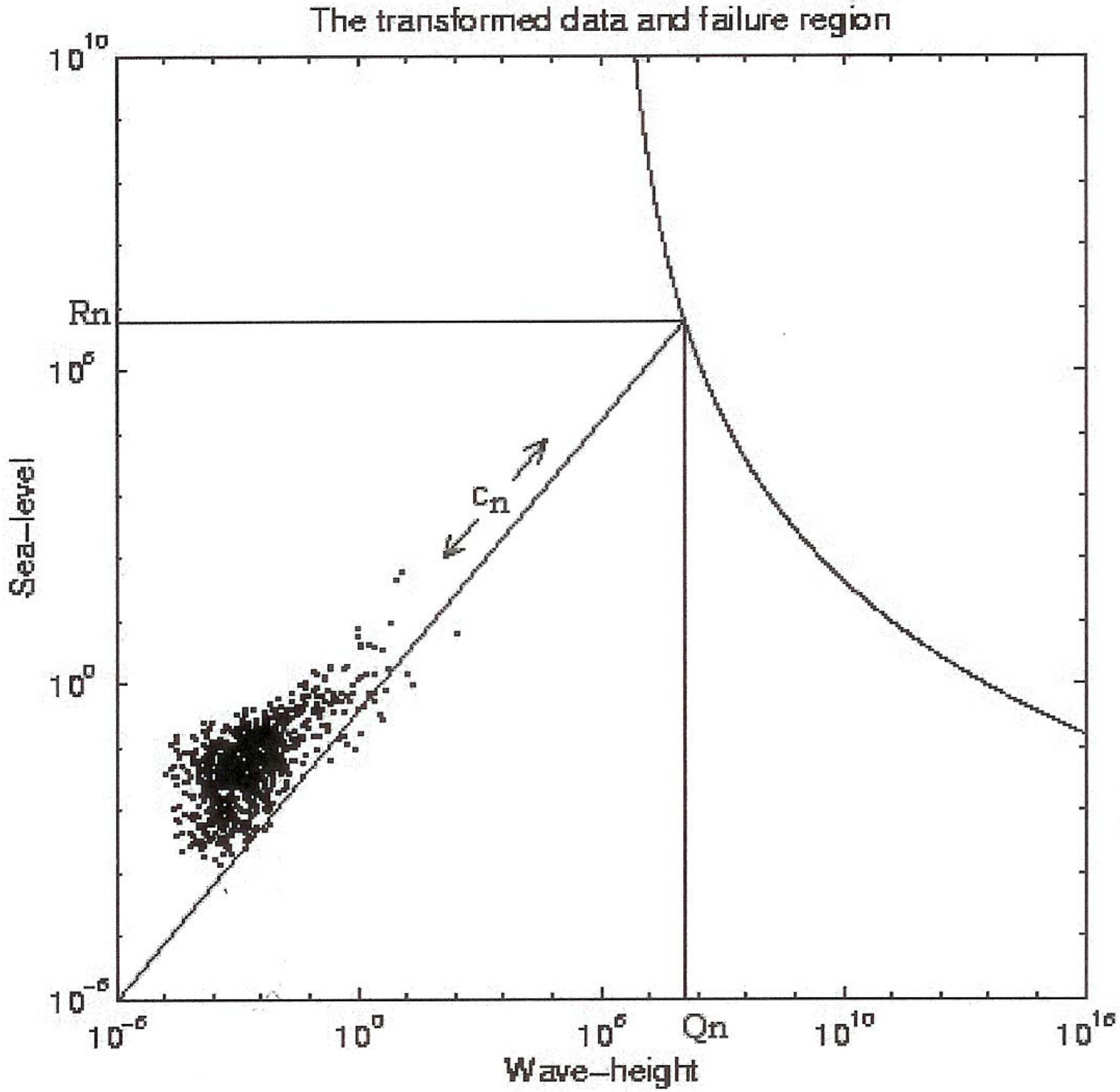
$$r_n := \left(1 + \gamma_2 \frac{w_n - b_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)} \right)^{\frac{1}{\gamma_2}} .$$

(c_n, q_n, r_n can be estimated via condition 2).

Further : S has a positive distance from the origin.

condition 3


Data and the failure region
Sea–level
Wave–height
Wn
Vn
8
7
6
5
4
3
2
1
0
–1
–2
0
2
4
6
8
10


The transformed data and failure region
10^{10}
Rn
10^{5}
10^{0}
10^{-5}
Sea–level
c_n
10^{-5}
10^{0}
10^{5}
Qn
10^{10}
10^{15}
Wave–height

Before we go on, we simplify notation:

Notation

$$R_n(x,y) := \left(\left(1 + \gamma_1 \frac{x - b_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)} \right)^{\frac{1}{\gamma_1}}, \left(1 + \gamma_2 \frac{y - b_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)} \right)^{\frac{1}{\gamma_2}} \right)$$

$$\hat{R}_n(x,y) := \left(\left(1 + \hat{\gamma}_1 \frac{x - \hat{b}_1\left(\frac{n}{k}\right)}{\hat{a}_1\left(\frac{n}{k}\right)} \right)^{\frac{1}{\hat{\gamma}_1}}, \left(1 + \hat{\gamma}_2 \frac{y - \hat{b}_2\left(\frac{n}{k}\right)}{\hat{a}_2\left(\frac{n}{k}\right)} \right)^{\frac{1}{\hat{\gamma}_2}} \right)$$

Note that:

$$R_n^{\leftarrow}(x,y)=\left(a_1\left(\frac{n}{k}\right)\frac{x^{\gamma_1}-1}{\gamma_1}+b_1\left(\frac{n}{k}\right),a_2\left(\frac{n}{k}\right)\frac{y^{\gamma_2}-1}{\gamma_2}+b_2\left(\frac{n}{k}\right)\right)$$

With this notation we can write equivalently

Cond. 1' : $\frac{n}{k}P\{R_n(X,Y)\in A\}\to\upsilon(A)$

Cond. 4' : $C_n=R_n^{\leftarrow}(c_nS)$

Condition 5 Sharpening of condition 1:

$$\frac{\frac{n}{k} P\{R_n(X,Y) \in c_n S\}}{\upsilon(c_n S)} \xrightarrow{P} 1.$$

Then :
$$\frac{nc_n}{k} P\{R_n(X,Y) \in c_n S\} \xrightarrow{P} \upsilon(S)$$

Condition 6 $\gamma_1, \gamma_2 > 1/2$ and

$$\lim_{n\to\infty} \frac{w_{\gamma_1 \wedge \gamma_2}(c_n)}{\sqrt{k}} = 0$$

for $i = 1, 2$, where $w_\gamma(x) := x^{-\gamma} \int_1^x s^{\gamma-1} \log s \, ds$.

The estimator

Note that

- $$S \overset{4}{=} \frac{1}{c_n} R_n(C_n) \quad \text{i.e.} \quad C_n = R_n^{\leftarrow}(c_n S)$$
- $$p_n = P(C_n) = P\{R_n^{\leftarrow}(c_n S)\} \overset{5}{\sim} \frac{k}{n}\upsilon(c_n S) \overset{\text{hom.}}{=} \frac{k}{nc_n}\upsilon(S).$$

Hence we propose the estimator

$$\hat{p}_n := \frac{k}{n\hat{c}_n}\hat{\upsilon}_n\left(\frac{1}{\hat{c}_n}\hat{R}_n(C_n)\right)\left(= \frac{k}{n\hat{c}_n}\hat{\upsilon}_n\left(\frac{1}{\hat{c}_n}\hat{R}_n R_n^{\leftarrow}(c_n S)\right)\right)$$

and we shall prove

$$\frac{\hat{c}_n}{c_n} \xrightarrow{P} 1 \,, \quad \hat{\upsilon}_n \xrightarrow{P} \upsilon \quad , \quad \text{"} \, \frac{1}{\hat{c}_n} \hat{R}_n R_n^{\leftarrow} c_n \xrightarrow{P} \text{identity} \text{"}.$$

Then

$$\hat{p}_n \sim \frac{k}{n c_n} \upsilon(S) \sim p_n \,.$$

More formally:

Write: $p_n := P(C_n)$. Our estimator is

$$\hat{p}_n := \frac{k}{n\hat{c}_n}\hat{\upsilon}_n\left(\frac{1}{\hat{c}_n}\hat{R}_n(C_n)\right)$$

where

$$\hat{\upsilon}_n(\cdot) := \frac{1}{k}\sum_{i=1}^{n} 1_{\left\{\hat{R}_n(X,Y)\in\,\cdot\right\}}$$

$$\hat{c}_n := \sqrt{\hat{q}_n^2 + \hat{r}_n^2}$$

$$\hat{q}_n := \left(1+\hat{\gamma}_1\frac{v_n - \hat{b}_1\left(\frac{n}{k}\right)}{\hat{a}_1\left(\frac{n}{k}\right)}\right)^{\frac{1}{\hat{\gamma}_1}}$$

$$\hat{r}_n := \left(1+\hat{\gamma}_2\frac{w_n - \hat{b}_2\left(\frac{n}{k}\right)}{\hat{a}_2\left(\frac{n}{k}\right)}\right)^{\frac{1}{\hat{\gamma}_2}}$$

Theorem

Under our conditions

$$\frac{\hat{p}_n}{p_n} \xrightarrow{P} 1$$

as $n \to \infty$ *provided* $\upsilon(S) > 0$.

For the proof note that by Condition 5

$$p_n = P(C_n) \sim \frac{k}{nc_n} \upsilon(S)$$

and

$$\hat{p}_n := \frac{k}{n\hat{c}_n} \hat{\upsilon}_n \left(\frac{1}{\hat{c}_n} \hat{R}_n C_n \right)$$

Hence it is sufficient to prove $\frac{\hat{c}_n}{c_n} \xrightarrow{P} 1$

and

$$\hat{\upsilon}_n \left(\frac{1}{\hat{c}_n} \hat{R}_n C_n \right) \xrightarrow{P} \upsilon(S).$$

For both we need the following fundamental Lemma.

Lemma

For all real γ *and* $x > 0$ *, if* $\gamma_n \to \gamma \ (n \to \infty)$ *and* $c_n \ge c > 0$*,*

$$\lim_{n\to\infty} \frac{1}{c_n}\left(1+\gamma_n\left\{\left(1+O_1(\gamma_n-\gamma)\right)\frac{(c_n x)^{\gamma}-1}{\gamma}+O_2(\gamma_n-\gamma)\right\}\right)^{\frac{1}{\gamma_n}}=x$$

provided

$$\lim_{n\to\infty}(\gamma_n-\gamma)c_n^{-\gamma}\int_1^{c_n} s^{\gamma-1}(\log s)\,ds=0 \ .$$

Proposition

$$\frac{\hat{c}_n}{c_n} \xrightarrow{P} 1$$

Proof *We first prove* $\hat{q}_n / q_n \xrightarrow{P} 1$. *Recall*

$$q_n := \left(1 + \gamma_1 \frac{v_n - b_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)} \right)^{\frac{1}{\gamma_1}} \quad \text{i.e.} \quad v_n = a_1\left(\frac{n}{k}\right) \frac{q_n^{\gamma_1} - 1}{\gamma_1} + b_1\left(\frac{n}{k}\right)$$

and

$$\hat{q}_n := \left(1 + \hat{\gamma}_1 \frac{v_n - \hat{b}_1\left(\frac{n}{k}\right)}{\hat{a}_1\left(\frac{n}{k}\right)} \right)^{\frac{1}{\hat{\gamma}_1}} .$$

Combining the two we get

$$\frac{\hat{q}_n}{q_n} = \frac{1}{q_n}\left(1+\hat{\gamma}_1\left\{\frac{a_1\left(\frac{n}{k}\right)}{\hat{a}_1\left(\frac{n}{k}\right)}\,\frac{q_n^{\gamma_1}-1}{\gamma_1}+\frac{b_1\left(\frac{n}{k}\right)-\hat{b}_1\left(\frac{n}{k}\right)}{\hat{a}_1\left(\frac{n}{k}\right)}\right\}\right)^{\frac{1}{\hat{\gamma}_1}}.$$

The Lemma gives

$$\frac{\hat{q}_n}{q_n} \xrightarrow{P} 1 .$$

Similarly

$$\frac{\hat{r}_n}{r_n} \xrightarrow{P} 1 .$$

hence

$$\frac{\hat{c}_n}{c_n} = \frac{\sqrt{\hat{q}_n^2 + \hat{r}_n^2}}{q_n^2 + r_n^2} \xrightarrow{P} 1 . \quad \square$$

Finally we need to prove

$$\hat{\upsilon}_n\left(\frac{1}{\hat{c}_n}\hat{R}_n R_n^{\leftarrow}(c_n S)\right)\xrightarrow{P}\upsilon(S).$$

We do this in 3 steps:

Proposition 1

Define

$$\tilde{\upsilon}_n(S) := \frac{1}{k}\sum_{i=1}^{n} 1_{\{R_n(X,Y)\in \mathbf{S}\}}$$

we have

$$\tilde{\upsilon}_n(S) \xrightarrow{P} \upsilon(S).$$

Proof

Just calculate the characteristic function and apply Condition 1.

Proposition 2

Define

$$\hat{\upsilon}_n(S) := \frac{1}{k}\sum_{i=1}^{n} 1_{\left\{\hat{R}_n(X,Y)\in \mathbf{S}\right\}}$$

we have

$$\hat{\upsilon}_n(S) \xrightarrow{P} \upsilon(S).$$

Proof

$$\hat{\upsilon}_n(S) = \tilde{\upsilon}_n(S)\left\{R_n^{\leftarrow}\hat{R}_n(S)\right\}$$

By the Lemma $R_n^{\leftarrow}\hat{R}_n \to$ *identity*.

Next apply Lebesgue's dominated convergence Theorem.

Proposition 3

$$\hat{\upsilon}_n\left(\frac{1}{\hat{c}_n}\widehat{R}_n\left(C_n\right)\right)\xrightarrow{P}\upsilon(S).$$

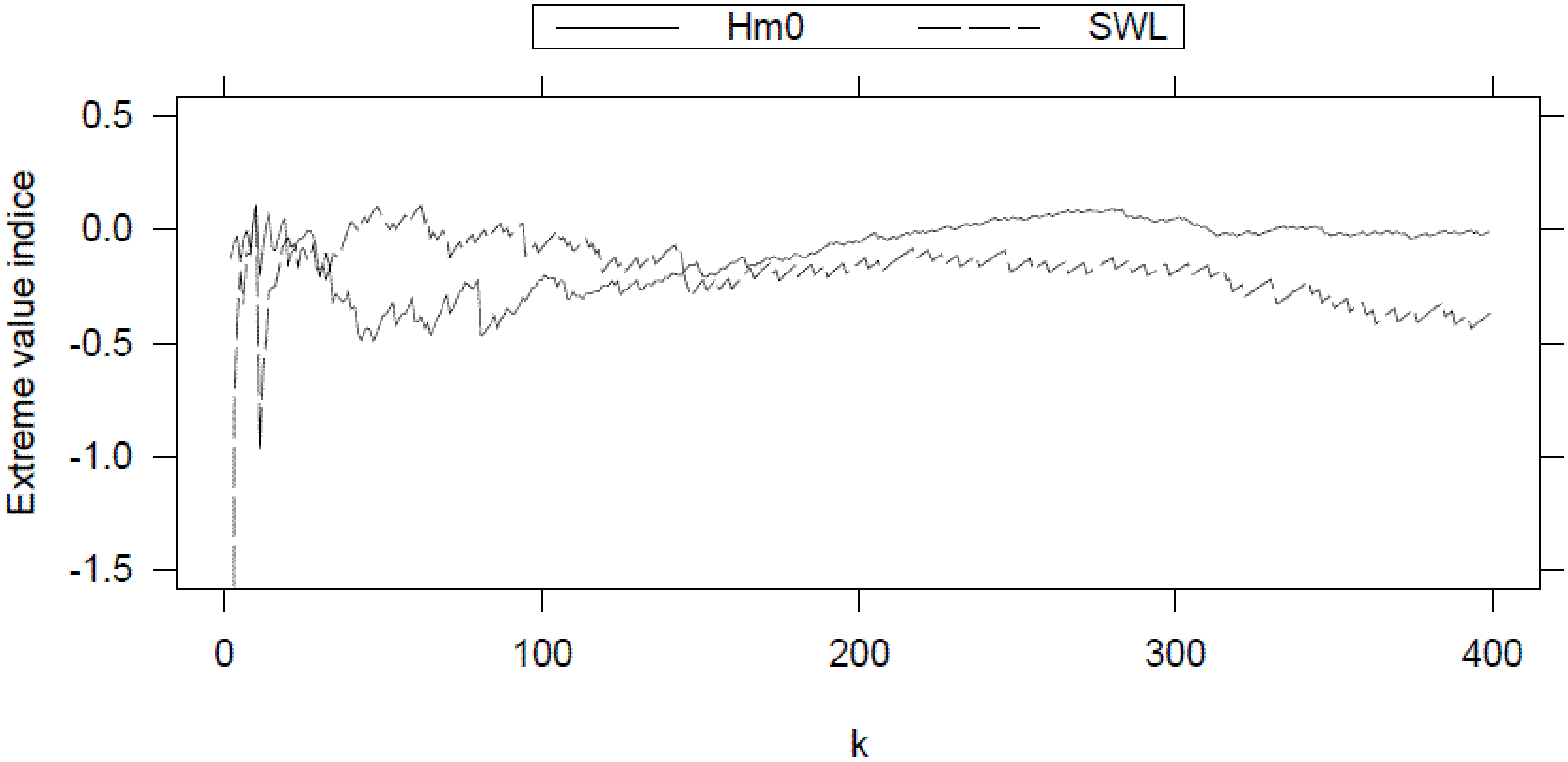
Proof

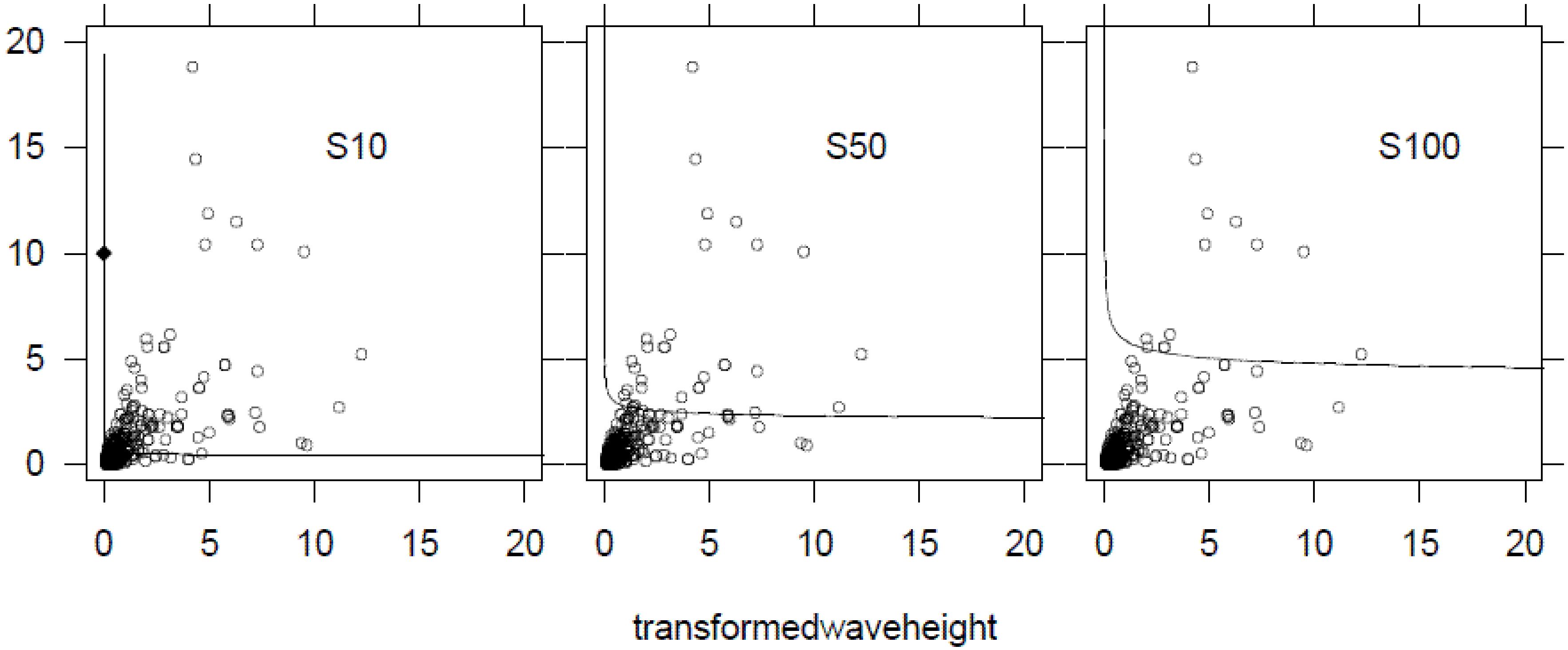
The left hand side is

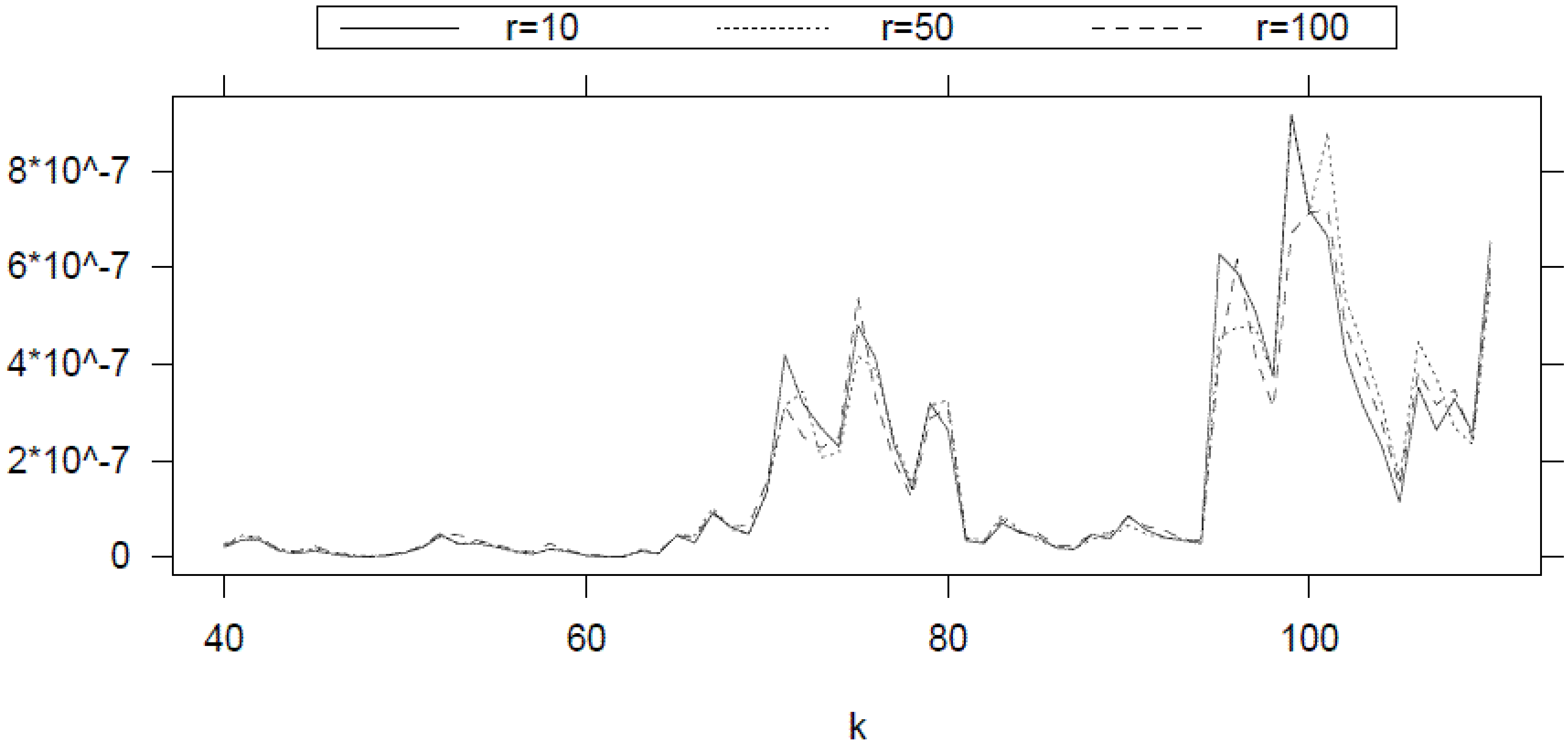
$$\hat{\upsilon}_n=\left(\hat{c}_n^{-1}\widehat{R}_n R_n^{\leftarrow}\left(c_n S\right)\right)$$

By the Lemma $\hat{c}_n^{-1}\widehat{R}_n R_n^{\leftarrow}c_n \to$ *identity.*

The result follows by using statement and proof of Proposition 2.


Hm0
SWL
Extreme value indice
0.5
0.0
-0.5
-1.0
-1.5
0
100
200
300
400
k


S10
S50
S100
0
5
10
15
20
transformedwaveheight


r=10
r=50
r=100
8*10^-7
6*10^-7
4*10^-7
2*10^-7
0
40
60
80
100
k

The 8th Conference on Extreme Value Analysis
July 8-12, 2013
Fudan University, Shanghai, China

We are pleased to announce that the 8th Conference on Extreme Value Analysis will take place from July 8 to 12, 2013 at Fudan University, Shanghai, China.

Organizers: Deyuan Li, Liang Peng,
Zhengjun Zhang, Ming Zheng

Email: eva2013sh@yahoo.com
Website: http://eva.fudan.edu.cn

Topics:
- Univariate, multivariate, infinite dimensional extreme value theory
- Order statistics and records
- Rare events and risk analysis
- Spatial/spatio-temporal extremes
- Heavy tails in actuarial sciences
- Other related applications

History: Previous EVA conferences have been held in Leuven, Belgium (2001), Lyon, France (2011), Vimeiro, Portugal (1983), Aveiro, Portugal (2004), twice in Gothenburg, Sweden (1998 and 2005), Bern, Switzerland (2007), Fort Collins, USA (2009).