#### 

Analyse van extreme waarden

ontwikkeling voorkeurswerkwijze bij waterschap Hollandse Delta

Siebe Bosch

**RAPPORT**



Siebe Bosch

Analyse van extreme waarden

ontwikkeling voorkeurswerkwijze bij waterschap Hollandse Delta

Vervaardigd in opdracht van Waterschap Hollandse Delta

Met bijdragen van Rudolf Versteeg, Bastiaan Kuijper (HKV Lijn in water) en Kees Vink (Waterschap Hollandse Delta)

**RAPPORT**

Inhoudsopgave

1 Inleiding 6

2 Leeswijzer 7

3 Onderzoekslocaties 8

4 Plotposities 11

4.1 Theorie 11

4.1.1 Weibull 13

4.1.2 Gringorten 13

4.1.3 Benard Bos-Levenbach 13

4.2 Gebruik 14

4.2.1 Visuele controle op plausibiliteit 14

4.2.2 Fitten aan een deelverzameling van de gegevens 15

5 Kansverdelingen 17

5.1 Inleiding 17

5.2 Op basis van maxima uit onderling onafhankelijke en gelijkverdeelde variabelen 18

5.2.2 GEV-kansverdeling 21

5.2.3 De Gumbel-kansverdeling 22

5.3 Op basis van overschrijdingen van drempelwaarden 23

5.3.2 De Exponentiële verdeling 25

5.3.3 De Lognormaal-verdeling 25

5.3.4 De Gegeneraliseerde Paretoverdeling 25

5.4 Quickscan onderzoeksgebied 26

5.4.1 Uitkomsten 26

5.4.2 Fitten aan een partiële dataset 27

6 Fitten van een kansverdeling 29

6.1 Fitten aan de kansdichtheidsfunctie 29

6.2 Fitten aan de plotposities 30

7 Omgaan met discontinuïteiten 31

7.1 Transformatie 31

7.2 Deelselectie 32

8 Onzekerheidsanalyse 33

8.1 Trekken uit trekkingen 34

8.2 Trekken uit de kansverdeling 34

8.3 Quickscan onderzoeksgebied 34

9 Uitkomsten en conclusies 38

9.1 Uitkomsten 38

9.2 Conclusies 38

10 Literatuur 39

BIJLAGEN 40

BIJLAGE A 40

Titel 40

10.1 Inleiding 41

10.2 Elementen 41

# Inleiding

Tijdens het overleg van de westelijke waterschappen is een indicatieve benchmark uitgevoerd van de diverse methoden die toegepast worden om kansen en herhalingstijden van waterstanden te berekenen.

De betrokken waterschappen zijn Schieland en de Krimpenerwaard, Delfland, Rijnland, Stichtse Rijnlanden en Hollandse Delta. Een memo van Waterschap Hollandse Delta van 1 juni 2016 geeft een overzicht van de onderzochte statistische methodieken. In de onderstaande tabel vatten wij dit overzicht samen.

Het memo stelt daarnaast dat bij Nederlandse waterschappen ook andere methoden worden gebruikt, zoals bijvoorbeeld de ‘methode Willems’ (P. Willems, Universiteit van Leuven), toegepast bij Waterschap Brabantse Delta en de stochastenmethode (o.a. Stichtse Rijnlanden). Sommige adviesbureaus werken daarnaast met analyse van partiële reeksen (Peaks over Threshold-aanpak), veelal in combinatie met de Paretoverdeling. Soms wordt ook de Gumbel (GEV type I) verdeling wel toegepast op partiële reeksen van maxima, in plaats van op de gehele reeks. Dit wordt gedaan om te voorkomen dat non-lineair systeemgedrag de resultaten beïnvloedt.

Het memo geeft inzage in de resultaten onder de gevolgde methodieken, maar verbindt daar geen conclusies aan voor wat betreft een voorkeursmethodiek.

Aan de hand van het voorgaande gaf Waterschap Hollandse Delta aan Hydroconsult de opdracht om onderzoek te verrichten en richtlijnen op te stellen voor de voorkeursmethodiek bij hoogwateranalyses in haar beheergebied. Het waterschap vraagt binnen deze opdracht:

* Een advies voor de meest geschikte kansverdeling
* De wijze waarop deze kansverdeling toegepast dient te worden
* Een handreiking voor het produceren van de statistische onzekerheidsband
* Advies voor het omgaan met discontinuïteiten in de overschrijdingsgrafieken

# Leeswijzer

In hoofdstuk 3 kiezen we enkele representatieve locaties uit het beheergebied waarvoor we de hoogwaterstatistiek gaan onderzoeken. In de hoofdstukken daarna werken we telkens een aspect van hoogwaterstatistiek uit:

* Hoofdstuk 4: plotposities
* Hoofdstuk 5: inventarisatie van de soorten kansverdelingen
* Hoofdstuk 6: fitten van een kansverdeling
* Hoofdstuk 7: omgaan met discontinuïteiten
* Hoofdstuk 8: onzekerheidsanalyse

In elk van die hoofdstukken komen relevante uitkomsten van de gekozen locaties aan bod. In het hoofdstuk 9 combineren we vervolgens alle kennis uit de voorgaande hoofdstukken en gaan we dieper in op het studiegebied en de fenomenen die daar optreden. In hoofdstuk 10 trekken we conclusies voor het beheergebied van Waterschap Hollandse Delta.

Voor getallen hanteren wij in dit rapport de Angelsaksische notatie, namelijk met de punt (.) als decimaalscheider en de komma (,) als duizendtalscheider.

# Onderzoekslocaties

In dit hoofdstuk kiezen we een achttal representatieve locaties uit het beheergebied van Waterschap Hollandse Delta waarvoor we de hoogwaterstatistiek gaan onderzoeken.

Datasets van twee gebieden staan tot onze beschikking ten behoeve van deze studie:

* Bemalingsgebied De Bommelse Polders (Goeree-Overflakkee)
* Hoogvliet (regio IJsselmonde)

Voor beide gebieden is een reeks aan 323 met SOBEK berekende maximale waterstanden beschikbaar. Deze waterstanden zijn berekend door de modelschematisaties te onderwerpen aan een langjarige toetsingsreeks voor het huidige klimaat (Lit.1). Voor De Bommelse Polders hebben we bovendien de beschikking over de bergingscurves van het watersysteem.

De selectie van representatieve locaties maken we op basis van de zogenoemde plotposities van de berekende maxima (zie hoofdstuk 4). Door die grafisch weer te geven worden uitbijters en discontinuïteiten in de gegevens snel inzichtelijk gemaakt.

De selectie van representatieve punten is daarop gebaseerd dat we zo divers mogelijk systeemgedrag willen dekken. Dit doen we aan de hand van de plotposities (andermaal: hoofdstuk 4). We zoeken locaties die een glad verloop vertonen (lineair en gebogen) en locaties die veel verspringingen laten zien; zowel bij kleine als bij grote herhalingstijden.

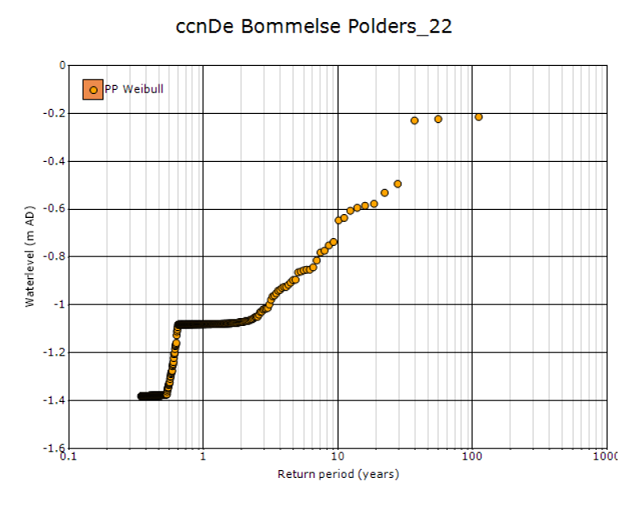
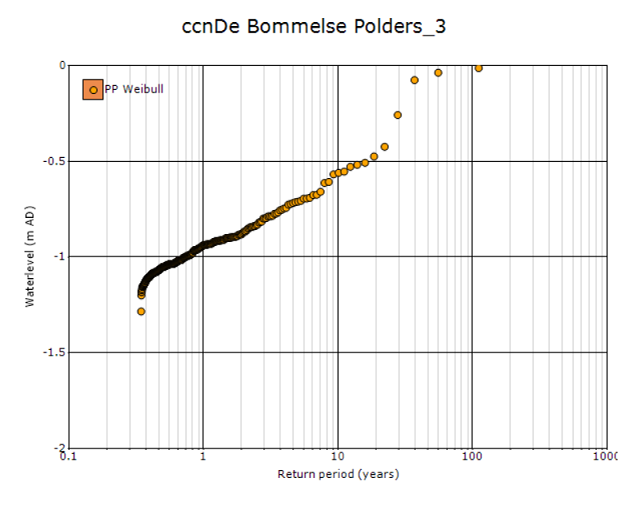
Uit de dataset van De Bommelse Polders kiezen we op basis van de bovengenoemde criteria de volgende rekenpunten:

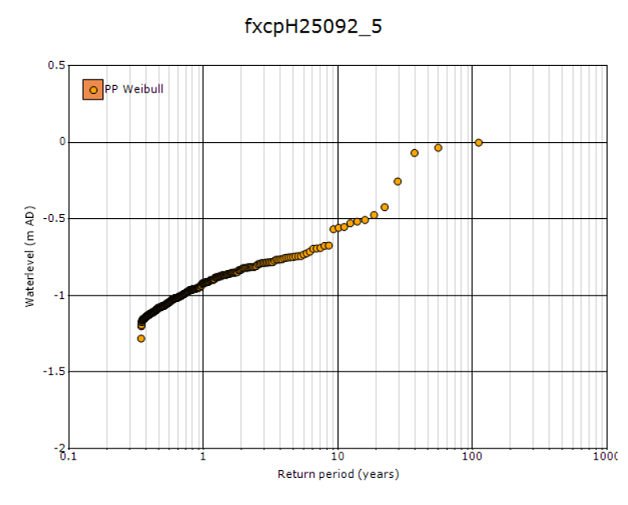
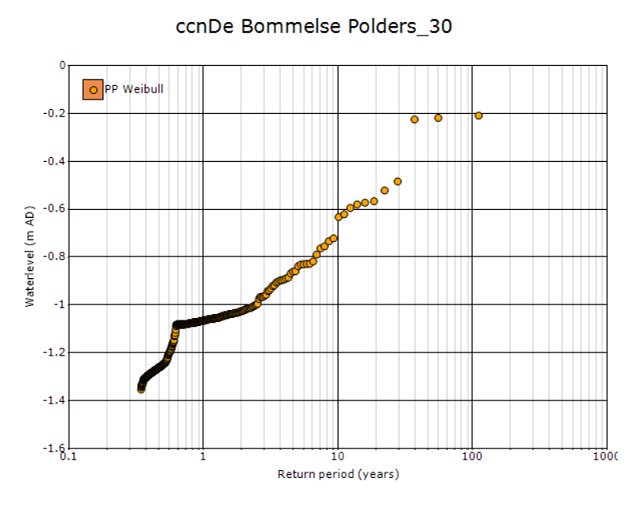
1. *ccnDeBommelsePolders\_22*, hierna te noemen: locatie A  
     
   duidelijk afwijkend systeemgedrag bij kleine herhalingstijden; mogelijk veroorzaakt doordat nog geen peilstijgingen t.o.v. streefpeil plaatsvinden. Bovenin drie uitbijters.
2. *ccnDeBommelsePolders\_3*, hierna te noemen: locatie B  
     
   bij kleine herhalingstijden een mooi recht verloop van de plotposities op enkellogaritmisch-papier. Bij de kleinste herhalingstijden een steil verticaal verloop, wat duidt op weinig waterbergend volume rond streefpeil.
3. *ccnDeBommelsePolders\_30*, hierna te noemen: locatie C  
     
   wat betreft het verloop van de plotposities houdt deze locatie het midden tussen de twee hierboven genoemde locaties.
4. *fxcpH25092\_5*, hierna te noemen: locatie D  
     
   op deze locatie lijkt zich een discontinuïteit voor te doen rond de geschatte herhalingstijd T=10, te zien aan de verspringing in de grafiek.

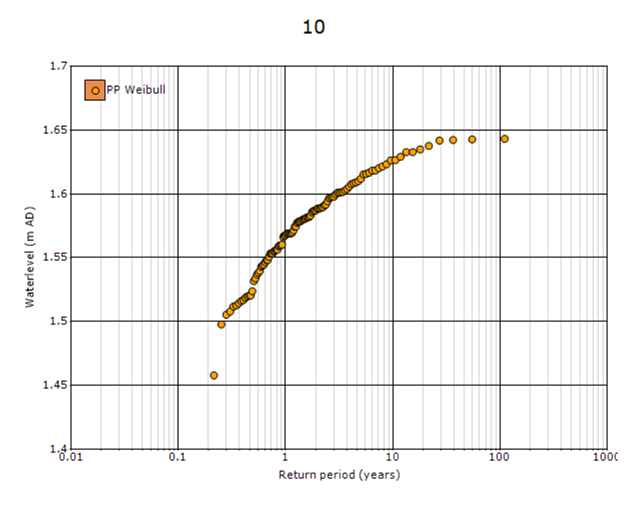
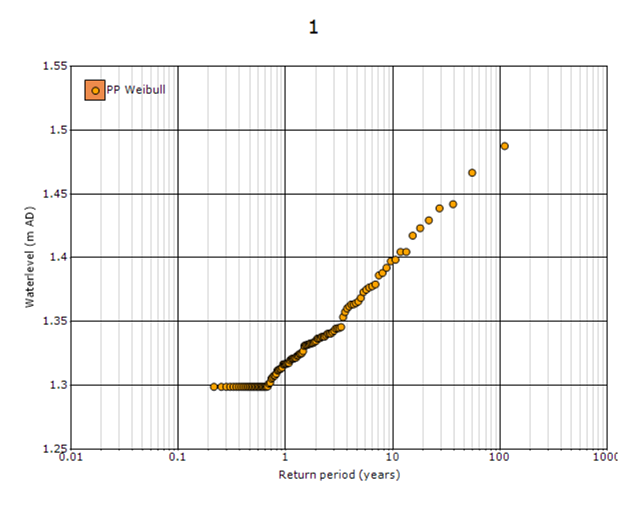
Uit de dataset van Hoogvliet kiezen we de volgende rekenpunten:

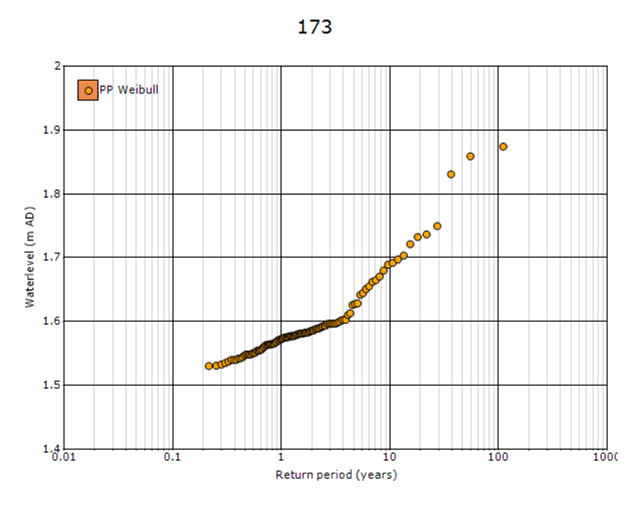
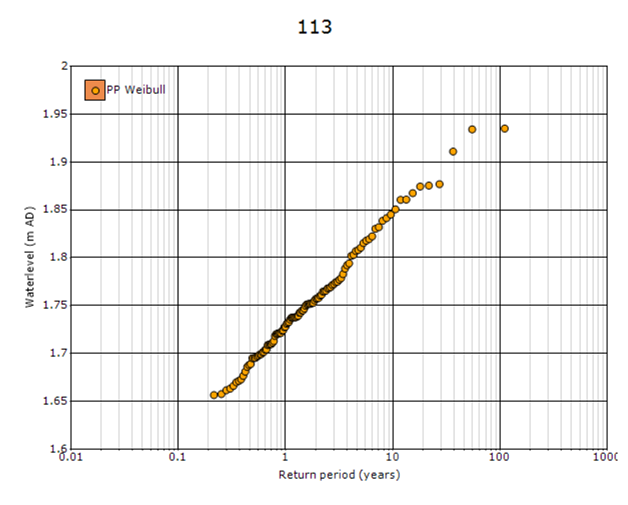
1. *1*, hierna te noemen: locatie E  
     
   Een vrijwel rechte lijn op logaritmisch papier, maar met tot ca T=0.7 een constante waarde, wat duidt op streefpeil.
2. *10*, hierna te noemen: locatie F  
     
   op enkel-logaritmisch papier een sterk gekromd verloop van de plotposities met nauwelijks discontinuïteiten
3. *113*, hierna te noemen: locatie G  
     
   Een vrijwel lineair verloop van de plotposities op enkellogaritmisch papier, met nauwelijks discontinuïteiten
4. *173*, hierna te noemen: locatie H  
     
   Een reeks die twee soorten systeemgedrag lijkt te bevatten, te oordelen aan de knik rond T=4.

Op de volgende pagina tonen we voor elk van de genoemde locaties de plotposities volgens Weibull voor elk van de 323 maxima uit de reeks.







Figuur vlnr. de plotposities volgens Weibull voor locaties A t/m H.

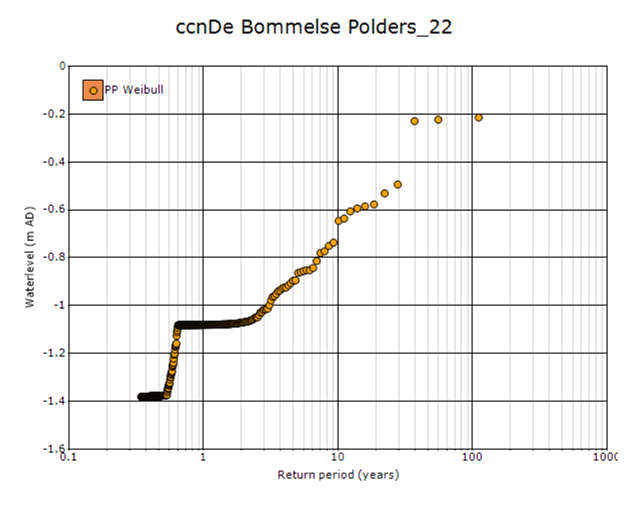
# Plotposities

## Theorie

Hier bespreken we de zogenoemde plotposities (Engels: *Plotting Positions*), zoals reeds getoond in het voorgaande hoofdstuk.

Plotposities geven een *eerste inschatting* van terugkeertijden van extremen. Dit doen ze aan de hand van het rangnummer van ieder maximum binnen zijn reeks.

Figuur 2 toont een visuele weergave van plotposities voor een willekeurig gekozen locatie in het beheergebied.



Figuur Voorbeeld van een grafiek met plotposities.

In de bovenstaande figuur valt het horizontale verloop rond -1.40 m + NAP direct op. Datzelfde geldt voor het gedrag rond -1.10 m + NAP. Dit betreft resp. winter- en zomerstreefpeil. Daarnaast zien we dat de hoogste drie waterstanden eruit springen. Dit laatste fenomeen kan verschillende oorzaken hebben: het kan gaan om inundatie van het laagste maaiveld of om werkelijke uitbijters in de statistiek van de onderliggende neerslagvolumes.

De grondgedachte achter de plotposities is dat het hoogste maximum uit een reeks met n jaren een overschrijdingskans heeft van naar schatting eens per n jaar. De op een na hoogste waarde een overschrijdingskans van 2 keer per n jaar, de op twee na hoogste waarde 3 keer per n jaar etc.

In de praktijk is de lengte van de beschikbare reeks echter beperkt en dit limiteert de toepasbaarheid van het bovenstaande. Bij de gebeurtenis met rangnummer n uit n waarnemingen zouden we bijvoorbeeld een overschrijdingskans van n/n = 1 krijgen, wat aantoonbaar onjuist is voor alle n < ∞. Vandaar dat er een variant bestaat (Weibull): i/(n+1). Daarnaast geldt dat 1/n niet bij alle steekproefgroottes de beste schatting is voor de overschrijdingskans van het hoogste maximum.

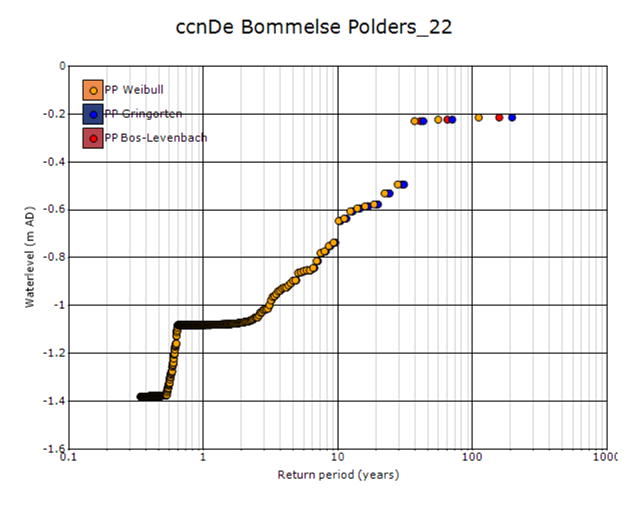
Voor gebruik in de praktijk -met reeksen van beperkte duur- zijn daarom in de literatuur tientallen varianten te vinden voor het bepalen van de plotposities. Al deze varianten beogen gegeven de steekproef-omvang @@@en de wijze waarop de data verdeeld is, Bastiaan?@@@ een iets betere schatting te geven voor de terugkeertijd van maxima; met name voor de hoogste waarden uit de reeks.

In dit rapport gaan we dieper in op de plotposities die in de hydrologie het meest worden gebruikt: Weibull, Gringorten en Benard Bos-Levenbach.

* Weibull (i/n+1)
* Gringorten (i-0.44)/(n+0.12)
* Benard Bos-Levenbach (i-0.3)/(n+0.4)

De grootheid n hierin hoeft overigens niet per sé het aantal maxima te vertegenwoordigen. n kan ook staan voor het aantal jaren waarvoor maxima beschikbaar zijn. In dat geval mag de geschatte overschrijdingskans direct worden omgezet naar een schatting van de terugkeertijd, namelijk door de reciproke te nemen: T = 1/p.

Figuur 3 toont voor de drie genoemde varianten de plotposities. Uit het resultaat wordt duidelijk dat de grootste verschillen zich rond de hoogste waterstanden voordoen. In het voorbeeld varieert de geschatte herhalingstijd van het hoogste waarde tussen 110 jaar (Weibull) en 195 jaar (Gringorten).



Figuur Drie varianten van plotposities, toegepast op dezelfde reeks.

### Weibull

De plotposities p volgens Weibull (lit @@@) worden berekend aan de hand van de volgende vergelijking:

p = i/(n+1).

<Uitleg achtergrond>

### Gringorten

De plotposities p volgens Gringorten (lit @@@) worden berekend aan de hand van de volgende vergelijking:

p = (i-0.44)/(n+0.12)

<Uitleg achtergrond>

### Benard Bos-Levenbach

De plotposities p volgens Benard Bos-Levenbach (lit @@@) worden berekend aan de hand van de volgende vergelijking:

p = (i-0.3)/n(+0.4)

<Uitleg achtergrond>

## Gebruik

Plotposities worden ingezet voor twee doeleinden:

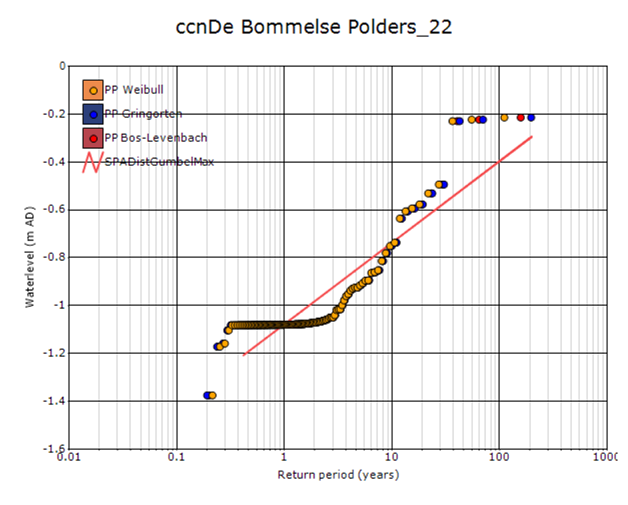
1. Als visueel controlemiddel om de plausibiliteit van een gekozen en gefitte kansverdeling te bepalen
2. Fitten van een deelverzameling van waarden aan een kansverdeling.

### Visuele controle op plausibiliteit

De belangrijkste functie van plotposities is dat ze de gebruiker direct een visuele terugkoppeling geven over de plausibiliteit van de gekozen kansverdeling en hoe goed die past bij de gegeven waarden. Daarnaast geven ze snel inzicht in mogelijke discontinuïteiten in de gegevensreeks.

Het fitten zelf gebeurt doorgaans zonder een waardeoordeel te geven aan het resultaat. Het algoritme kijkt uitsluitend naar de opgetreden waarden (waterhoogtes in dit geval), verdeelt die in klassen (zogeheten *bins*) en zet ze in een histogram om te onderzoeken hoe ze verdeeld zijn.

Daarbij zoekt het de optimale instelling van de kansparameters, gegeven de gevraagde kansverdeling. Het beoordeelt echter niet of de afgeleide kansverdeling ook daadwerkelijk hout snijdt.



Figuur Een mismatch tussen plotposities en gefitte kansverdeling.

Figuur 5 toont een situatie waarbij de gekozen kansverdeling leidt tot grote verschillen tussen de herhalingstijden zoals die volgen uit de kansverdeling volgen in vergelijking met de geschatte herhalingstijden volgens de plotposities (Weibull, Gringorten, Bos-Levenbach).

In dit geval ligt de oorzaak in een discontinuïteit in het systeemgedrag: maxima rond de -1.1 m + NAP overstijgen het streefpeil niet. Dit niet-natuurlijke systeemgedrag is het gevolg van de vaste gemaalcapaciteit. Zolang de aanvoer van water de gemaalcapaciteit niet overstijgt, zal geen peilstijging plaatsvinden.

### Fitten aan een deelverzameling van de gegevens

In veel gevallen zal uit de plotposities blijken dat er een discontinuïteit in de onderliggende werking van het watersysteem zit. Vaak betreft dit een ‘knik’ in het verloop van de waarden.

In Figuur 5 zagen we al zo’n discontinuïteit. Rond ca. T=3 vertoont het verloop van de plotposities een plotse knik naar boven. Fysisch kunnen we dit relateren aan het zomerstreefpeil en de gemaalcapaciteit.

Andere veroorzakers van discontinuïteiten zijn:

* Het inunderen van de laagste maaivelddelen
* Plotse extra instromingen als gevolg van riooloverstorten bij hevige neerslag
* Het bedienen van kunstwerken

In dergelijke situaties moet worden beoordeeld of de discontinuïteit van dien aard is dat alleen nog aan een deelselectie van de gegevens mag worden gefit. Wanneer dit het geval is, kunnen de plotposities een centrale rol krijgen bij het berekenen van terugkeertijden.

# Kansverdelingen

## Inleiding

In de statistiek van hoogwater zijn er twee hoofdstromingen van kansverdelingen te onderscheiden (Lit. @@@):

1. Kansverdelingen op basis van maxima uit onderling onafhankelijke en gelijkverdeelde variabelen
2. Kansverdelingen op basis van overschrijdingen van een drempelwaarde (Peaks over Threshold; ‘POT’)

In de volgende paragrafen werken wij voor beide hoofdstromingen de meestgebruikte kansverdelingen uit:

Categorie 1 (maxima uit onderling onafhankelijke en gelijkverdeelde variabelen):

* De GEV-kansverdeling (Gegeneraliseerde Extreme-waarden)
* De Gumbel-kansverdeling

Categorie 2 (Peaks over Threshold):

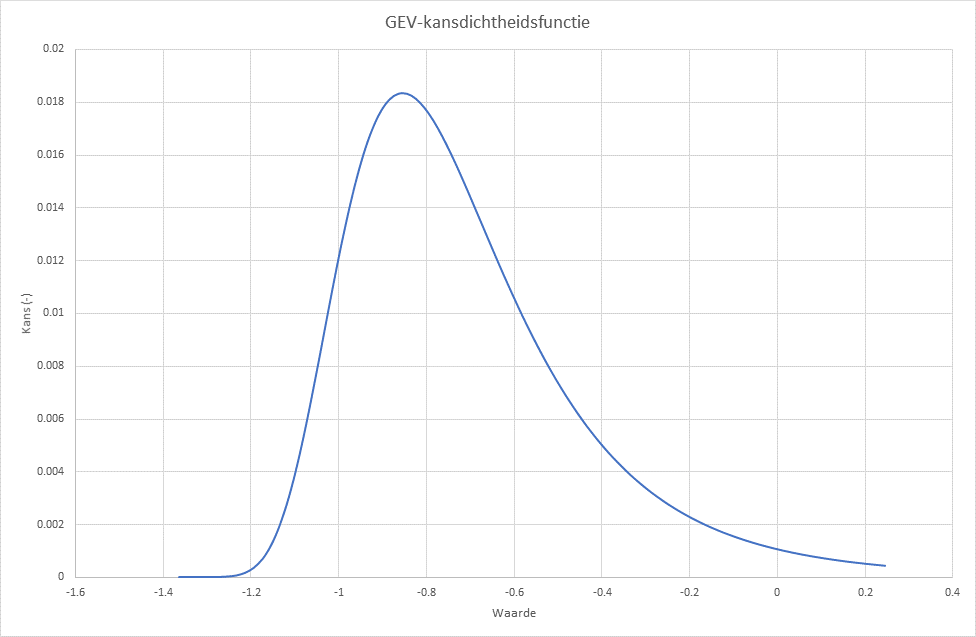
* de Exponentiële verdeling
* GPV (Gegeneraliseerde Paretoverdeling)
* lognormaalverdeling.

In de volgende secties bespreken we de karakteristieken van elk van deze hoofdstromingen en onderliggende kansverdelingen. Daarnaast benoemen we de consequenties van en de voorwaarden voor het gebruik ervan.

## Op basis van maxima uit onderling onafhankelijke en gelijkverdeelde variabelen

Dit type kansverdeling staat ook wel bekend als de familie van GEV-kansverdelingen. GEV staat voor *Generalized Extreme Values*. Deze kansverdeling behoort tot de hoofstroming “*maxima uit onderling onafhankelijke en gelijkverdeelde variabelen*”. Doorgaans wordt dit criterium in de hydrologie vertaald naar “jaarmaxima”.

Vergeleken met de welbekende normale verdeling hebben ze een ‘zwaardere staart’. Extremen aan de rechter zijde van de verdeling hebben een relatief grotere kans op voorkomen dan extremen aan de linker zijde.



Figuur Kansdichtheidsfunctie van de GEV-kansverdeling.

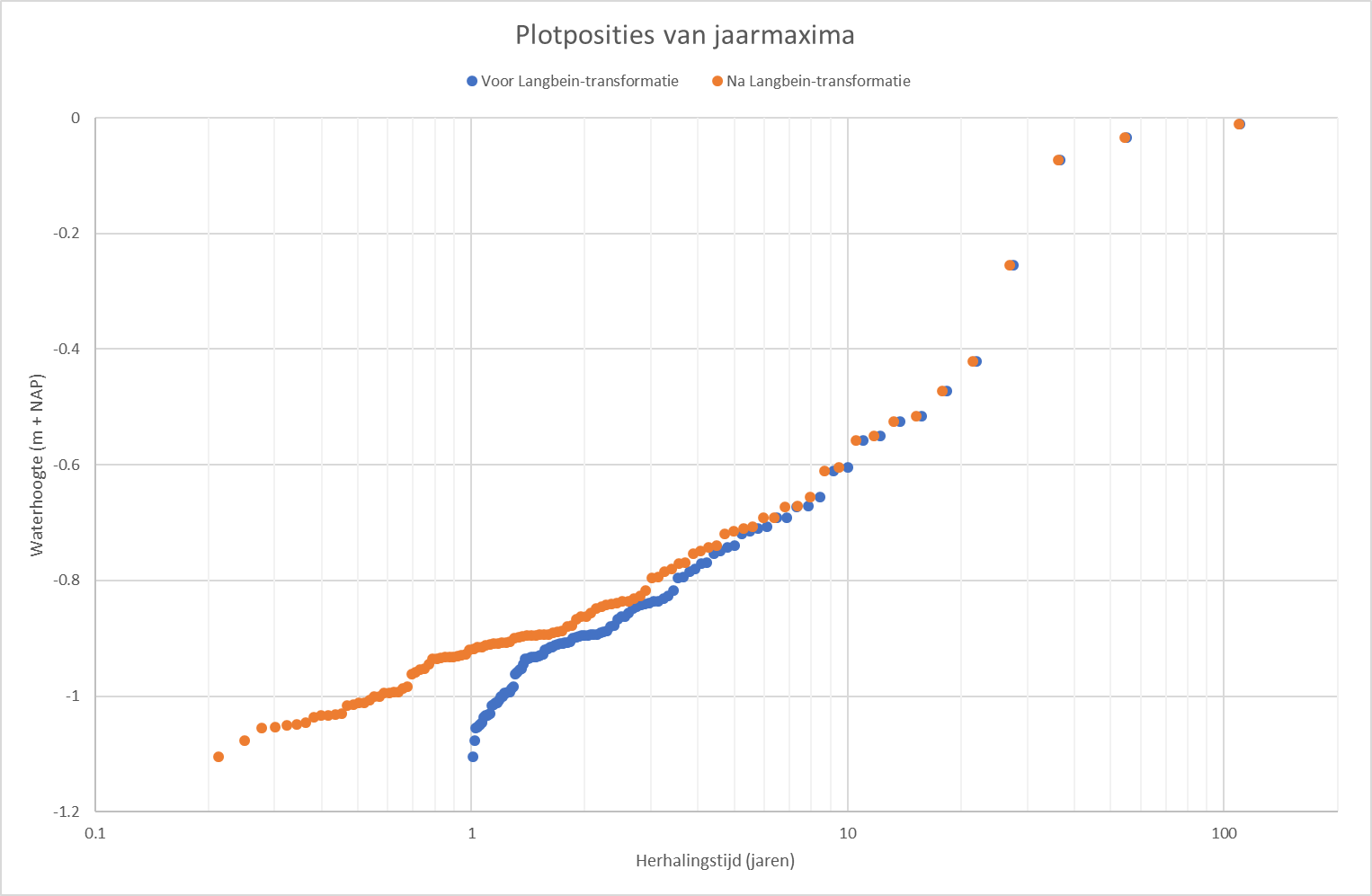
##### Consequenties van het gebruik

Kiezen voor de GEV-kansverdeling betekent per definitie dat secundaire en tertiaire hoogwaterpieken binnen dezelfde periode (meestal 1 jaar) buiten beschouwing blijven. Van ieder jaar wordt uitsluitend de hoogste waarde gebruikt. Daardoor wordt voorbijgegaan aan het feit dat in datzelfde jaar nog een tweede piek zou kunnen voorkomen die hoger is dan het maximum uit een ander jaar. Met name bij hoogfrequenter gebeurtenissen leidt dit tot een structurele onderschatting.

Aan de hand van GEV in zijn pure vorm kunnen geen uitspraken worden gedaan over waarden met een grotere kans op voorkomen dan eens per gekozen interval (meestal 1 jaar). Bovendien worden waarden bij kleine herhalingstijden structureel onderschat. In de praktijk is dat geen belemmering omdat voor wateroverlast omdat in het algemeen wordt gekeken naar herhalingstijden met een grootteorde van 10 – 100 jaar.

<Afbeelding secundaire pieken>

Om te corrigeren voor deze fout is het evenwel mogelijk om de zogenoemde Langbein-transformatie toe te passen op de berekende overschrijdingsfrequenties: T=1/-Ln(1-F) in plaats van T=1/F. T staat hierin voor de herhalingstijd in jaren en F voor de overschrijdingsfrequentie. Figuur 3 toont het effect van deze transformatie op de plotposities van een reeks met jaarmaxima.



Figuur Plotposities van de jaarmaxima uit een reeks, voor en na transformatie met Langbein.

Ondanks dat de Langbein-transformatie het mogelijk maakt om ook uitspraken te doen over hoogfrequenter gebeurtenissen is het van belang om te beseffen dat deze transformatie slechts een ‘model’ is, en daarmee een benadering van de werkelijkheid.

##### Voorwaarden voor gebruik

De GEV-kansverdeling mag alleen worden toegepast op maxima uit een aaneengesloten set van onderling onafhankelijke en equidistante perioden. In de hydrologie wordt doorgaans aangenomen dat het gebruik van jaarmaxima beantwoordt aan deze criteria.

In het geval van discontinuïteiten in het systeemgedrag is deze kansverdeling alleen bruikbaar met enkele substantiële beperkingen. Voorbeelden van zulke discontinuïteiten zijn:

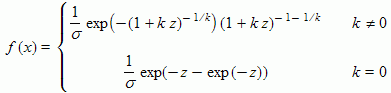
* Bij lage terugkeertijden een knik in de overschrijdingsgrafiek als gevolg van het handhaven van streefpeilen door een gemaal
* Bij hoge terugkeertijden een knik in de overschrijdingsgrafiek als mogelijk gevolg van het inunderen van maaiveld

In dergelijke gevallen mag de kansverdeling uitsluitend aan een deelverzameling van de gegevens worden gefit. De consequentie is echter dat niet langer rechtstreeks kan worden gefit aan de kansdichtheidsfunctie. De gebruiker moet terugvallen op het fitten aan de plotposities; iets wat niet ideaal is gezien de enorme diversiteit in plotposities.

Hier gaan wij dieper op in in het hoofdstuk 6 *Fitten*.

### GEV-kansverdeling

De GEV-kansverdeling in algemene zin heeft de volgende kansdichtheidsfunctie:

  
<nog vervangen door definitieve vorm.>

Vergelijking kansverdelingsfunctie voor de GEV-verdeling.

Hij wordt gedefinieerd door drie variabelen:

* Locatieparameter
* Schaalparameter
* Vormparameter

De situatie waarbij geldt: k > 0 staat ook wel bekend als de Weibull-kansverdeling. Bij k < 0 heet hij de Frèchet-kansverdeling en bij k=0 spreken we van Gumbel. Omdat die laatste variant voor de hydrologie enkele bijzondere eigenschappen heeft, bespreken we Gumbel in een aparte paragraaf.

### De Gumbel-kansverdeling

De Gumbel-kansverdeling is een speciaal geval van de hierboven besproken GEV-verdeling. Ook Gumbel behoort tot de hoofstroming “maxima uit onderling onafhankelijke en gelijkverdeelde variabelen”.

Wanneer de vormparameter k uit de kansverdelingsfunctie van de GEV-verdeling op 0 wordt gezet, resulteert dit in de Gumbel-kansverdeling.

Gumbel Distribution PDF <nog vervangen door def.>

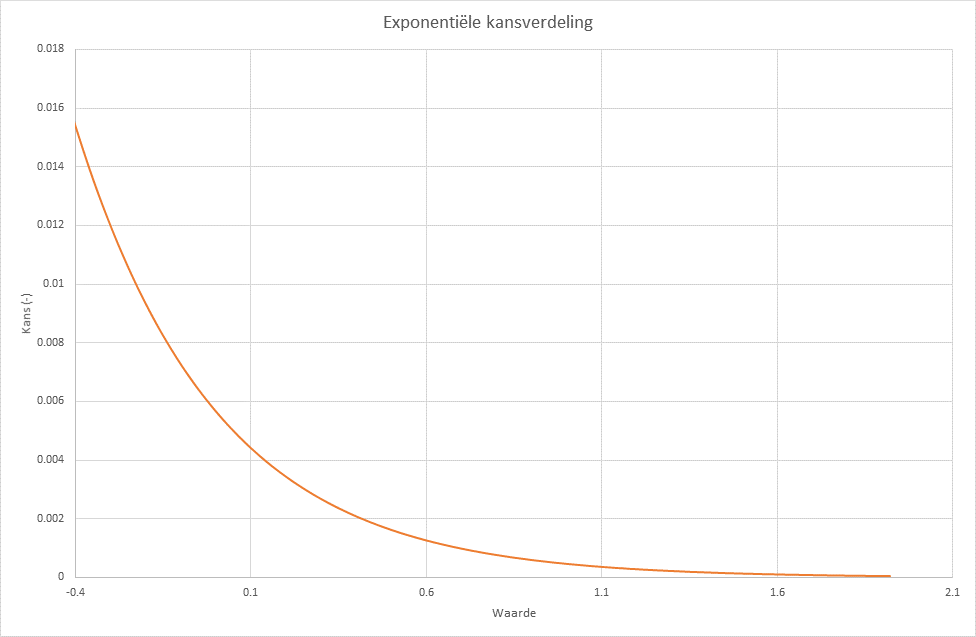
De Gumbel-kansverdeling is een afgeleide van de GEV-verdeling, maar met één groot verschil: Gumbel wordt gedefinieerd door twee in plaats van drie parameters:

* De locatieparameter
* De schaalparameter

Doordat de vormparameter ontbreekt ligt de ‘scheefstand’ van de verdeling op voorhand vast. Bij het fitten aan bepaalde gegevens kan dit door de gebruiker worden ervaren als een handicap, maar een groot voordeel van Gumbel boven de overige twee GEV-verdelingen is dat de onzekerheidsband een stuk smaller is. Hierop gaan we dieper in in hoofdstuk 8: Onzekerheidsanalyse.

## Op basis van overschrijdingen van drempelwaarden

Alle kansverdelingen die werken op basis van een drempeloverschrijding (in het Engels Peaks Over Threshold) beschrijven in essentie alleen het *staartverloop* van de kansverdelingen die op basis van jaarmaxima werken. Ter illustratie tonen we onderstaand de kansdichtheidsfunctie van de exponentiële verdeling.



Figuur Voorbeeld van de exponentiële kansverdeling.

##### Consequenties van het gebruik

Kansverdelingen op basis van POT-waarden beschrijven slechts een gedeelte van reeks met gegevens. We hebben niet langer gemiddeld één waarde per jaar, en daarom is het noodzakelijk om de berekende kansen te corrigeren voor het ‘ontbrekende’ gedeelte.

Op basis van een gegeven kans berekenen we de herhalingstijd als volgt:

p = 1/ T \* j/n

Waarbij:

p de overschrijdingskans

T de gevraagde herhalingstijd [jaren]

j het aantal jaren waarvoor maxima beschikbaar zijn

n het aantal maxima waaraan de kansverdeling is gefit

Een groot voordeel van deze methodiek is dat de verdelingen gebruik maken van zo mogelijk alle drempeloverschrijdingen, en zich dus niet beperken tot jaarmaxima. Hierdoor bestaat er geen groot risico dat er pieken worden ‘gemist’. Een correctie zoals Langbein is voor deze verdeling daarom per definitie niet aan de orde.

##### Voorwaarden voor het gebruik

De gebruikte maxima dienen statistisch onafhankelijk te zijn. Op dit punt doen we altijd een aanname want 100% onafhankelijk zullen ze nooit zijn. Denk hierbij aan grootschalige weersystemen zoals El Niño, die weersystemen over lange duur kunnen beïnvloeden. De essentie is dat gebeurtenissen kiezen waarvoor wij de aanname van onafhankelijkheid verdedigbaar vinden.

### De Exponentiële verdeling

De exponentiele verdeling is verwant aan de GEV-verdeling in die zin dat hij alleen het staartverloop van die laatstgenoemde beschrijft:

De kansdichtheidsfunctie luidt:

Hierbij staat x voor de overschrijding van de gekozen drempelwaarde.

### De Lognormaal-verdeling

De lognormaal-kansverdeling beschrijft variabelen waarvan de logaritme normaal verdeeld is. In andere woorden: wanneer variabele X lognormaal verdeeld is, is zijn logaritme Y = ln(X) normaal verdeeld.

### De Gegeneraliseerde Paretoverdeling

## Quickscan onderzoeksgebied

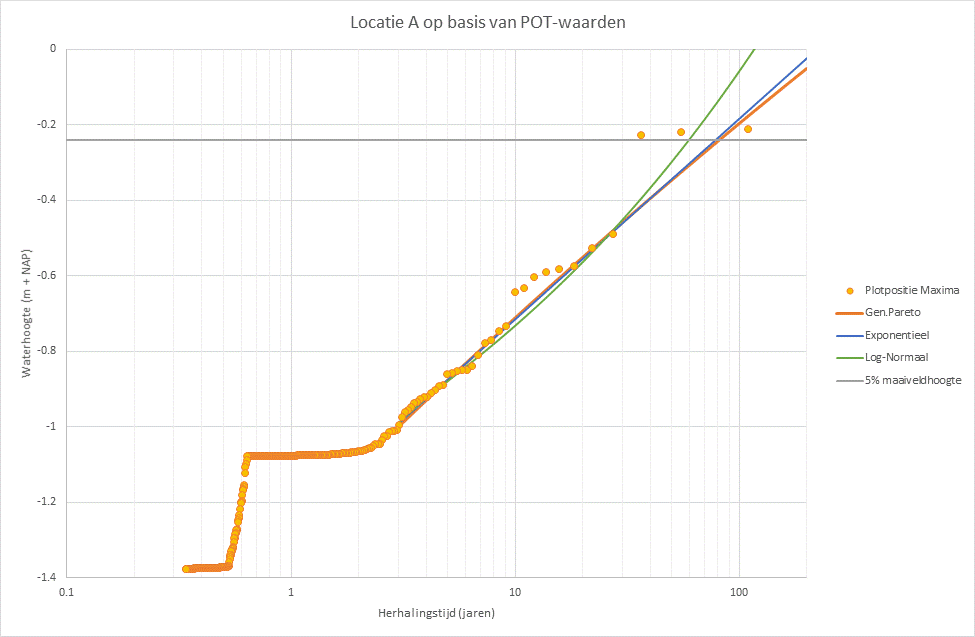
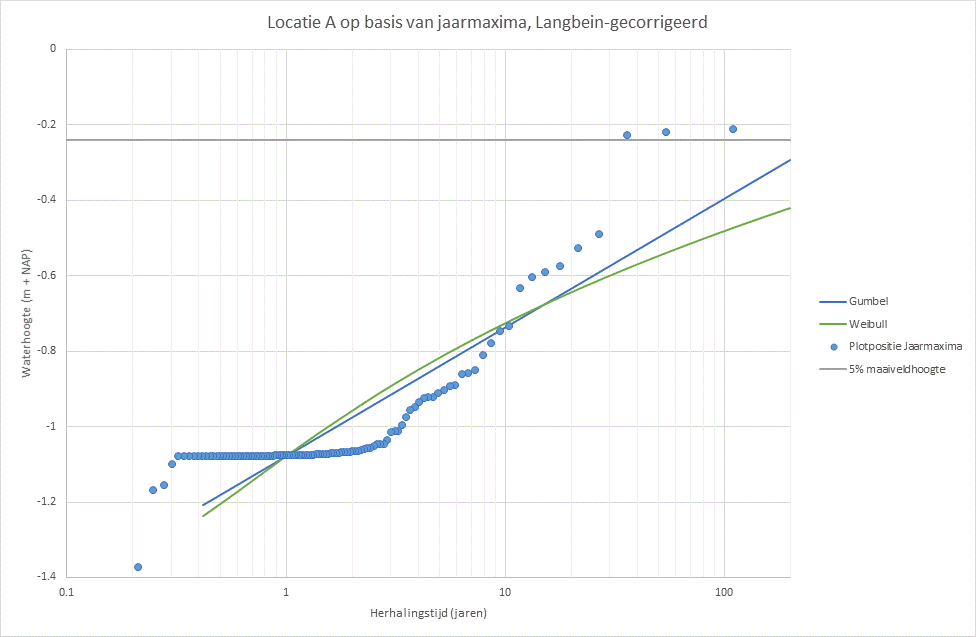
In deze paragraaf doen we een snelle eerste analyse van de gegevensreeksen uit het studiegebied aan de verschillende soorten kansverdelingen. Als referentie voegen we de plotposities van de beschikbare extremen toe. Een goede fit zal het verloop van de plotposities, zeker bij kleinere herhalingstijden, goed volgen.

Voor wat betreft de kansverdelingen op basis van POT-waarden moeten we vooraf een keuze maken voor een drempelwaarde waarboven gefit zal worden. Op basis van de plotposities schatten we in dat de hoogste 50 waarnemingen voor iedere locatie hiervoor geschikt zijn.

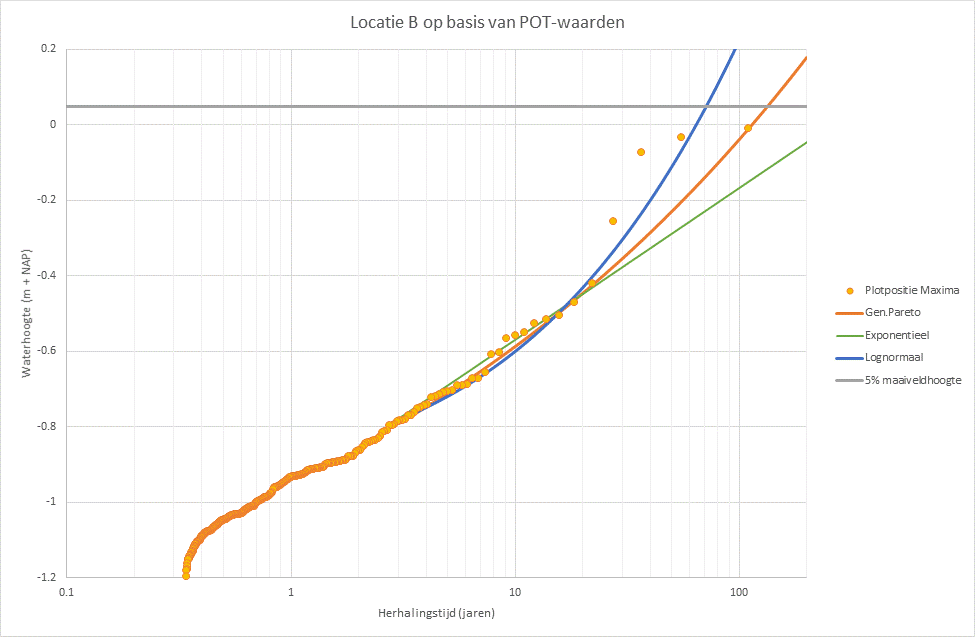
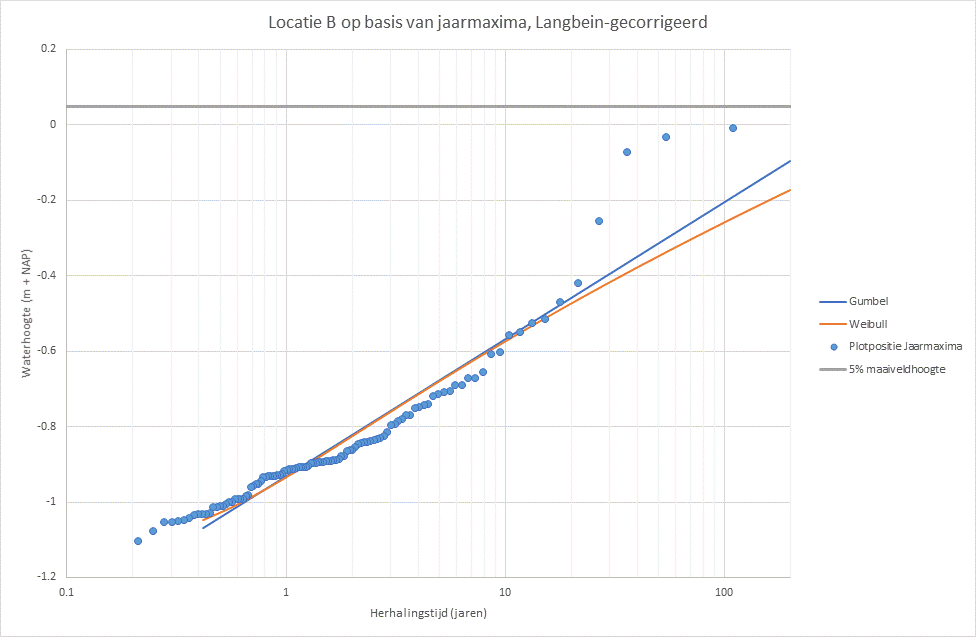
### Uitkomsten

Figuur 8 en Figuur 9 tonen de gefitte kansverdelingen voor resp. locatie A en B wanneer wordt gefit aan alle beschikbare maxima.

Als plotpositie hebben we die van Weibull gebruikt (i/(n+1)). Aan de grafiek hebben wij ter referentie toegevoegd: het 5%-maaiveldniveau van de onderhavige afwaterende eenheid. Merk op dat de grafieken links telkens slechts 109 plotposities tonen (= jaarmaxima) en de grafieken rechts 323 (= POT-waarden).



Figuur Kansverdelingen voor locatie A, gefit aan de volledige set aan maxima



Figuur Kansverdelingen voor locatie B, gefit aan de volledige set aan maxima

Het resultaat van de Jaarmaxima-kansverdelingen blijkt extreem gevoelig te zijn voor de discontinuïteiten in het systeemgedrag bij kleine herhalingstijden. Dit is te zien aan de veel minder goede fit voor locatie A. De knik te zien ter hoogte van de streefwaterstand heeft zijn weerslag op de resulterende fits.

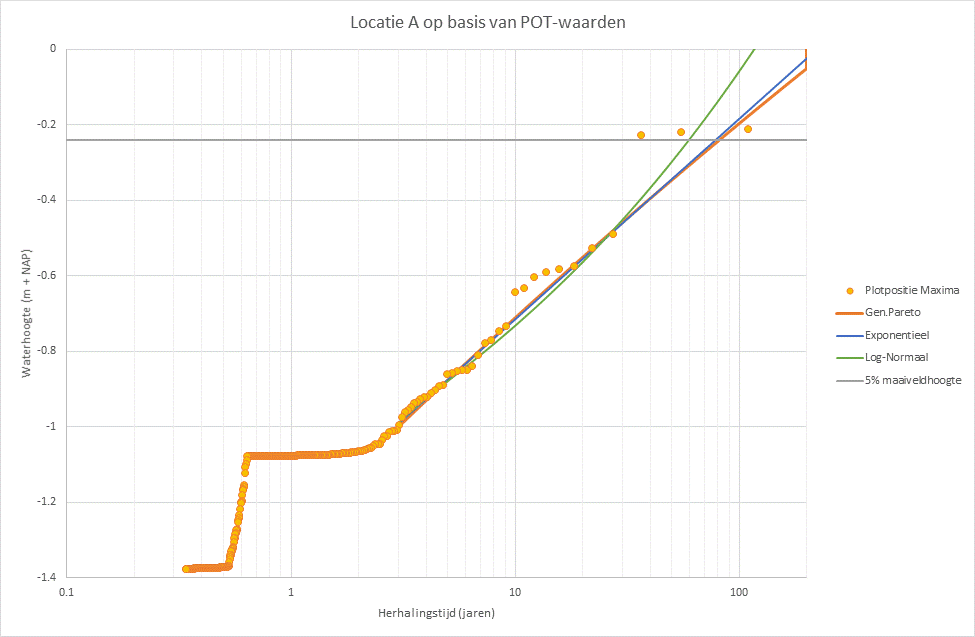
Voor locatie B zijn de resultaten een stuk beter omdat deze vrijwel geen discontinuïteit vertoont.

Voor locatie A vertonen de POT-kansverdelingen veel betere fits dan de Jaarmaxima-verdelingen. Dit is te danken aan het feit dat ze gebruik maken van een drempelwaarde. Ze hebben daardoor geen last van de discontinuïteiten ter hoogte van het streefpeil.

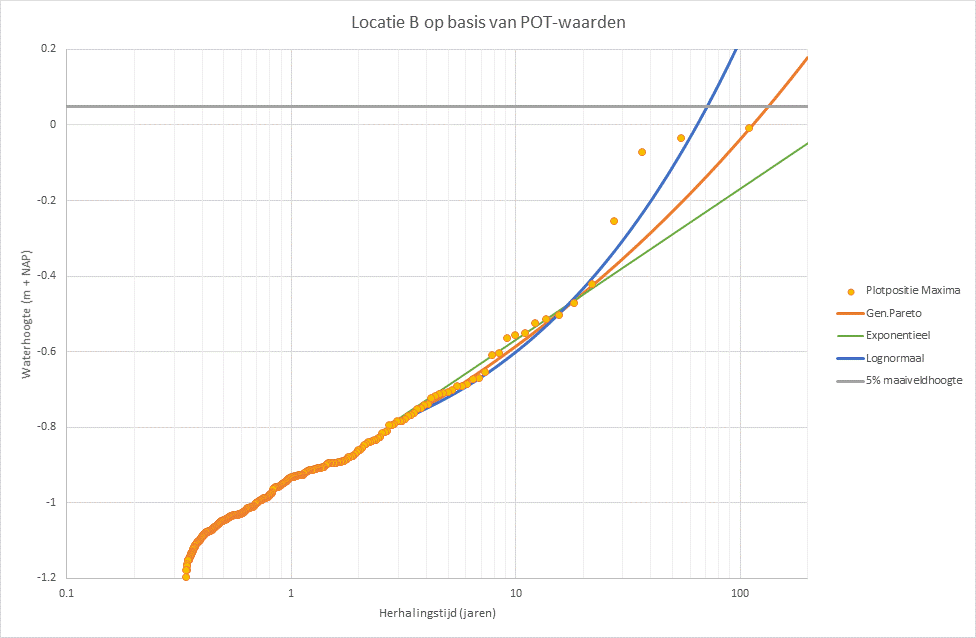
Voor locatie B geven alle kansverdelingen redelijk tot goede fits. Dit is te danken aan het feit dat hier geen grote discontinuïteiten aanwezig zijn. Grote verschillen tussen de kansverdelingen zien we daarom pas vanaf grote herhalingstijden (> 10 jaar) ontstaan. Wel is het zo dat alle POT-kansverdelingen beter de plotposities volgen.

### Fitten aan een partiële dataset

De discontinuïteiten doen vermoeden dat voor de Jaarmaxima-kansverdelingen een beter resultaat kan worden verkregen wanneer we alleen fitten aan de maxima die boven streefpeil uitstijgen. Figuur 10 en Figuur 11 tonen aan dat dit inderdaad het geval is. Hierin zijn alleen de hoogste 50 POT-waarden gebruikt om te fitten en de hoogste @@@ jaarmaxima.

<afb jaarmax met left censoring> 

Figuur Kansverdelingen voor locatie A, gefit aan de hoogste 40 maxima.

<afb jaarmax met left censoring> 

Figuur Kansverdelingen voor locatie A, gefit aan de hoogste 40 maxima.

Het bovenstaande toont het belang aan van een goede identificatie van discontinuïteiten bij het fitten van een kansverdeling aan waterhoogtes.

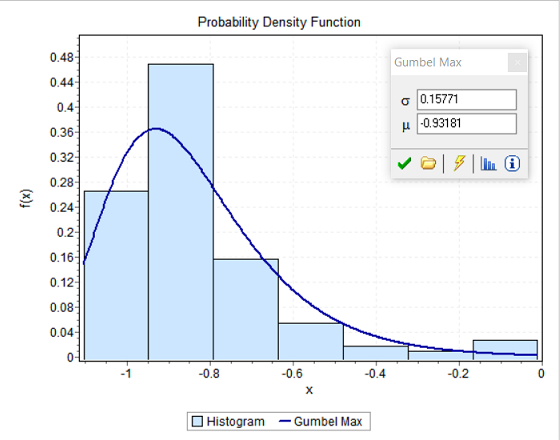
In hoofdstuk 8 *Omgaan met discontinuïteiten* gaan we daarom dieper in op de mogelijke verklaringen van de discontinuïteiten en het oplossen ervan. Maar eerst diepen we het proces van fitten verder uit in hoofdstuk 7 *Fitten van een kansverdeling*.

# Fitten van een kansverdeling

In dit hoofdstuk behandelen we de wijze waarop een kansverdeling kan worden gefit aan gemeten of berekende waarden. De werkwijze die in de literatuur ook als voorkeursmethodiek wordt aangeduid, is om de kansdichtheidsfunctie rechtstreeks te fitten aan een histogram van de waarnemingen. Een alternatief is om gebruik te maken van de plotposities. Dit heeft echter als nadeel dat daarmee een extra onzekerheid wordt geïntroduceerd.

## Fitten aan de kansdichtheidsfunctie

De meest toegepaste wijze om een kansverdeling te fitten aan waarnemingen is met de *Method of Moments* of de *Maximum Likelihood*-methode. Beide methodes delen de waarnemingen eerst op in klassen, zogeheten *bins*, die worden uitgezet als een histogram. Daarna wordt de kansdichtheidsfunctie zo afgeregeld dat die de klassenfrequenties zo goed mogelijk volgt.



Figuur Het fit-algoritme brengt de gegevens onder in een histogram, waarna de kansdichtheidsfunctie eraan wordt gefit (schermafdruk uit het programma EasyFit Professional van MathWave technologies).

Het voordeel van deze methodiek is dat hij bijzonder eenduidig is geen ruimte laat voor interpretatie. Een nadeel is dat de methodiek alleen werkt als ook daadwerkelijk alle waarnemingen waarvoor hij bedoeld is meedoen.

## Fitten aan de plotposities

Zoals in hoofdstuk 4 Plotposities al uiteengezet bestaan er talloze varianten voor het berekenen van de plotposities. Het rechtstreeks fitten van een kansverdeling aan de plotposities introduceert daarom meteen een vraagstuk: welke variant kan het beste worden gebruikt?

# Omgaan met discontinuïteiten

Een heikel punt bij de statistische analyse van extremen in het waterbeheer is het optreden van discontinuïteiten in de grafiek met terugkeertijden. Een discontinuïteit markeert de overgang tussen twee soorten systeemgedrag in het watersysteem.

Het meest bekende voorbeeld is het fenomeen dat boven een bepaalde waterhoogte het maaiveld begint te inunderen. Het gevolg hiervan is dat bij gelijke toename van het volume de waterhoogte veel minder snel stijgt.

In dit voorbeeld is het goed denkbaar dat wel dezelfde *soort* kansverdeling geldt voor de waterhoogtes onder en boven het inundatieniveau, maar dat die wel een volstrekt andere parameterisatie hebben.

In dit hoofdstuk bespreken we enkele methodes om met discontinuïteiten in hoogwaterstatistiek om te gaan. Grofweg zijn er twee manieren:

* Transformaties toepassen om de discontinuïteit te doen verdwijnen
* De kansverdeling fitten op een deelselectie van de dataset

## Transformatie

In het hiervoor genoemde voorbeeld van maaiveldinudatie is sprake van een plotse verspringing in de relatie tussen de waterhoogte en het geborgen volume. Wanneer het water zich binnen het sloottalud bevindt voldoen de maxima daardoor aan een andere kansverdeling dan wanneer ze zich boven het maaiveld bevinden.

<Afbeelding>

Een mogelijkheid om dit probleem te overkomen is om niet langer de waterhoogte te fitten aan de kansverdeling, maar in plaats daarvan het geborgen volume. De storende invloed van het maaiveld wordt op die manier uitgesloten, is de gedachte.

##### Eigenschappen

##### Consequenties van het gebruik

##### Voorwaarden voor het gebruik

## Deelselectie

Een meer gebruikelijke methode is om een deelselectie uit de reeks te kiezen, zodanig dat de maxima daaruit wel aan één kansverdeling voldoen.

##### Eigenschappen

##### Consequenties van het gebruik

Het is niet altijd mogelijk om op de klassieke wijze te ijken aan een deelselectie

##### Voorwaarden voor het gebruik

# Onzekerheidsanalyse

In dit hoofdstuk behandelen we de wijze waarop de statistische onzekerheid van een gefitte kansverdeling in beeld kan worden gebracht.

De techniek hiervoor staat bekend als *bootstrapping*. Bootstrapping staat voor het herhaaldelijk doen van trekkingen uit een kansverdeling, met teruglegging. De spreiding die in het getrokken resultaat tot uiting komt, staat bekend als de statistische onzekerheid.

We bespreken twee bootstrapping-technieken:

1. Trekken uit de oorspronkelijke reeks met trekkingen
2. Willekeurige waarden trekken uit een gefitte kansverdeling en het resultaat opnieuw fitten aan dezelfde kansverdeling

Beide methodes resulteren in een statistische onzekerheidsband rond de gefitte kansverdeling.

Bij de keuze tussen deze twee technieken spelen de volgende overwegingen een rol:

* Voordeel van trekken uit de oorspronkelijke maxima is dat er wordt gewerkt met de ‘echte’ data uit het veld. Nadeel is dat het aantal maxima vaak klein is, wat kan resulteren in een erg wijd betrouwbaarheidsinterval.
* Trekken op basis van de gefitte kansverdeling heeft als nadeel het ‘keurslijf’ van de gekozen verdelingsfunctie al op voorhand aan het systeem wordt opgedrongen.

In de volgende paragrafen werken we beide methodieken uit.

## Trekken uit trekkingen

Onze werkwijze bij trekken uit de oorspronkelijke trekkingen is als volgt:

* Uit de oorspronkelijke dataset van n waarnemingen wordt n keer met teruglegging getrokken. De uitkomsten worden weggeschreven naar een nieuwe dataset
* Het bovenstaande herhalen we vele malen (orde 1000x).
* Uit de verzameling uitkomsten zoeken we voor ieder rangnummer de 2.5- en 97.5-percentielwaarde op.
* Aan zowel de 2.5- en 97.5 percentielwaarden fitten we opnieuw een kansverdeling. Deze fits vormen tezamen het statistische betrouwbaarheidsinterval.

## Trekken uit de kansverdeling

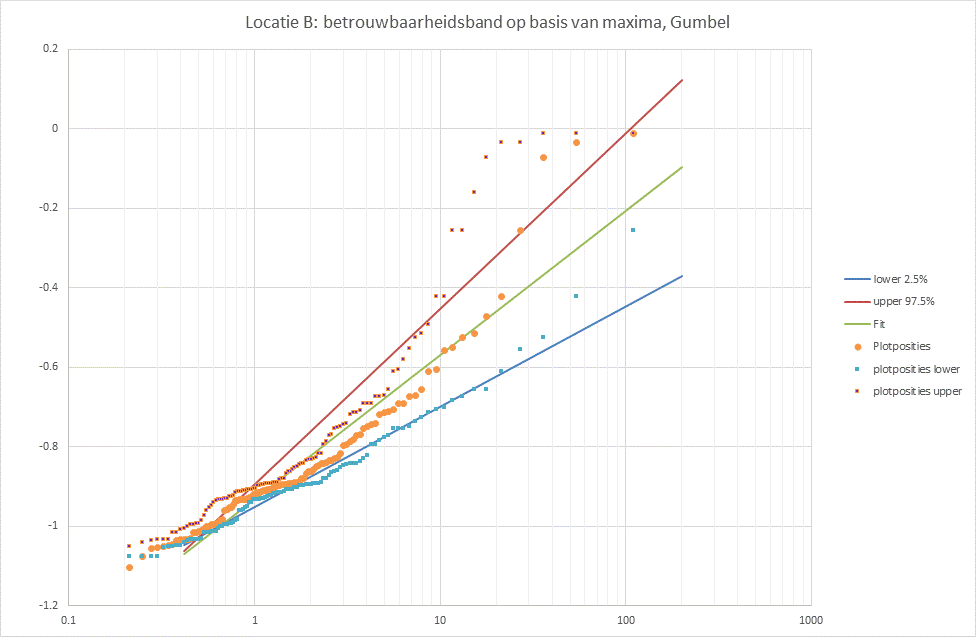
De werkwijze bij het trekken uit een kansverdeling is als volgt:

* De kansverdeling wordt gefit aan de beschikbare dataset met n waarnemingen per locatie; eventueel met uitsluiting van gegevens onder een drempelwaarde (POT).
* Uit de gefitte cumulatieve kansverdeling worden willekeurige waarden tussen 0 en 1 getrokken. De uitkomst wordt weggeschreven naar een tabel. Dit gebeurt net zo vaak als er aan gegevenspunten uit de oorspronkelijke dataset gefit werd.
* Het punt hierboven herhalen we vele malen (orde 1000x).
* Uit de verzameling uitkomsten zoeken we voor ieder rangnummer de 2.5- en 97.5-percentielwaarde op.
* Aan zowel de 2.5- en 97.5 percentielwaarden fitten we opnieuw een kansverdeling. Deze fits vormen tezamen het statistische betrouwbaarheidsinterval.

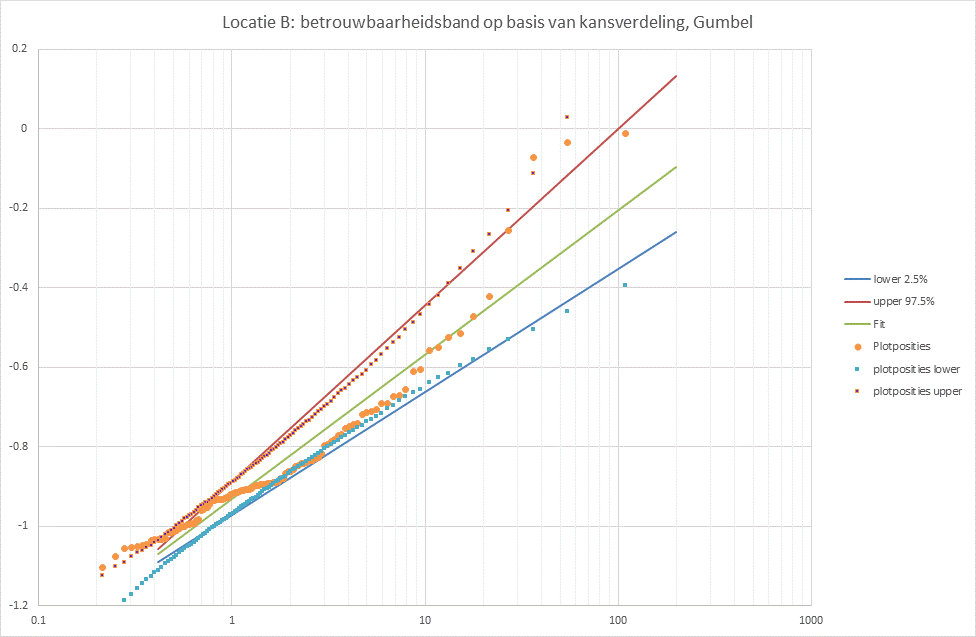
## Quickscan onderzoeksgebied

Voor locatie B onderzoeken we de het betrouwbaarheidsinterval voor elk van de kansverdelingen én de hier bovengenoemde methodieken. De keuze voor deze locatie is gebaseerd op de constatering dat hij de minste discontinuïteiten kent en daardoor een meer objectieve vergelijking tussen de kansverdelingen toestaat.

In de onderstaande figuren tonen we de onbetrouwbaarheidsband voor een aantal kansverdelingen en werkwijzen. We vatten het resultaat samen in een tabel.

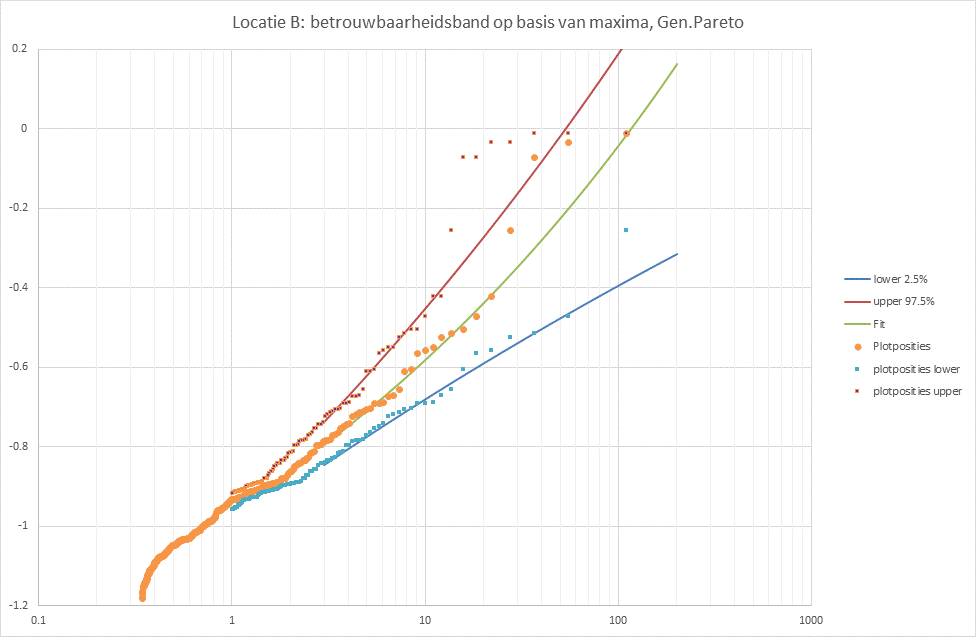


Figuur Betrouwbaarheidsband voor de Gumbel-verdeling, afgeleid door te trekken uit de oorspronkelijke waarnemingen.

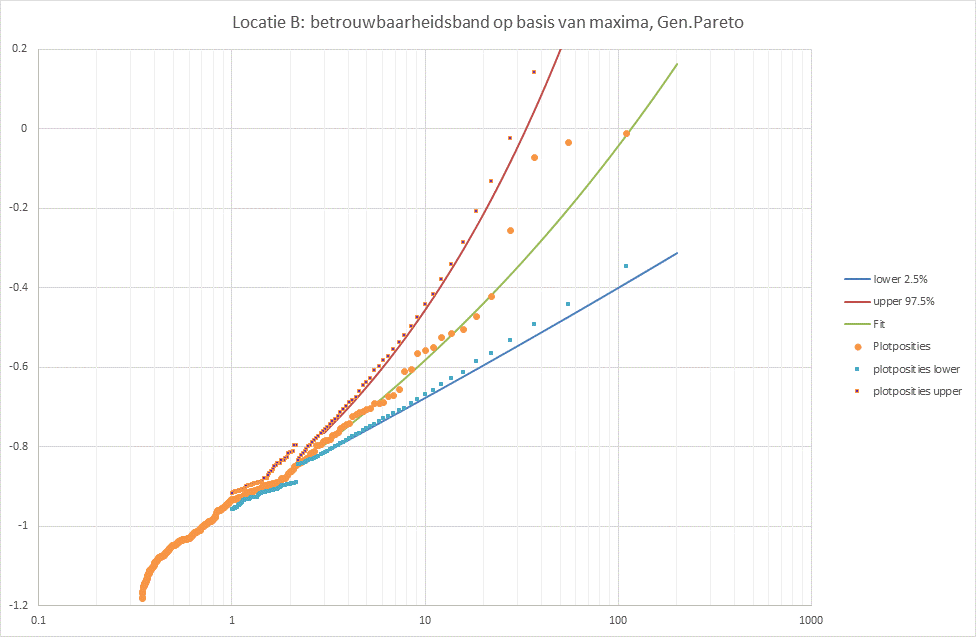


Figuur Betrouwbaarheidsband voor de Gumbel-verdeling, afgeleid door te trekken uit de gefitte kansverdeling.

In de bovenstaande vergelijking vallen meteen twee zaken op: het betrouwbaarheidsinterval op basis van de kansverdeling is smaller en de onderliggende plotposities liggen meer op een nette rechte lijn.



Figuur Betrouwbaarheidsband voor de Gen.Pareto-verdeling, afgeleid door te trekken uit de oorspronkelijke waarnemingen.



Figuur Betrouwbaarheidsband voor de Gen.Pareto-verdeling, afgeleid door te trekken uit de gefitte kansverdeling.

Ook in deze vergelijking liggen de plotposities op basis van de kansverdeling op een nette lijn. Hier is echter de bandbreedte van het betrouwbaarheidsinterval op basis van de oorspronkelijke maxima smaller.

In de onderstaande tabel vatten we de uitkomsten van alle vergelijkingen samen voor de herhalingstijd T=100 en locaties A en B

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Trekken uit trekkingen | | Trekken uit kansverdeling | |
|  | Locatie A | Locatie B | Locatie A | Locatie B |
| Gumbel | 47 cm | 44 cm | 34 cm | 35 cm |
| Weibull | 497 cm | 45 cm | 515 cm | 45 cm |
| Gen.Pareto | 43 cm | 58 cm | 72 cm | 101 cm |
| Exponentieel | 62 cm | 43 cm | 65 cm | 54 cm |
| Lognormaal | -45 cm | 99 cm | 184 cm | 154 cm |

De uitkomsten zijn zeer wisselend. De ene keer resulteert bootstrapping uit de trekkingen tot een smallere betrouwbaarheidsband, dan weer tot een bredere.

Bij de Weibull-verdeling voor locatie A zien we dat het bootstrappen problemen heeft met de discontinuïteit: dit geldt voor beide vormen van trekken. Ook bij de lognormaal-verdeling zien we een artefact die het gevolg is van de discontinuïteit.

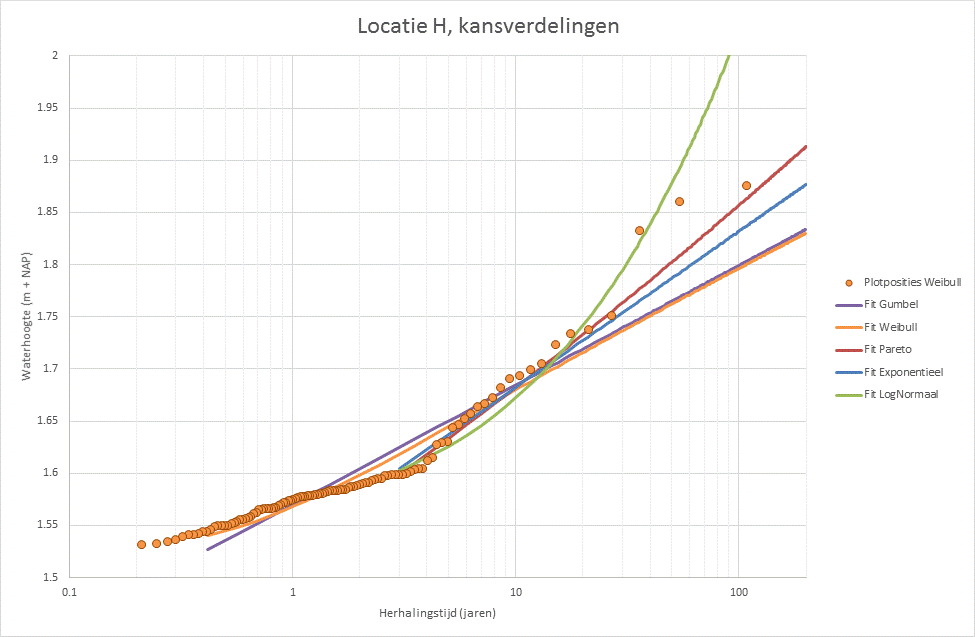
In het algemeen kunnen we wel stellen dat de bandbreedtes dusdanig groot zijn dat de praktische bruikbaarheid van het bootstrappen voor hoogwateranalyses vrijwel nihil is.

# Analyse beheergebied

In dit hoofdstuk analyseren we elk van de representatieve onderzoekslocaties. Hier komen de uitkomsten per statistisch onderdeel (zoals in de voorgaande hoofdstukken behandeld) samen.

## Locatie H

In Figuur 17 tonen we de uitkomst van het fitten van alle vijf de kansverdelingen. Voor de POT-verdelingen hebben we hier gekozen om de hoogste 40 maxima mee te nemen.



Figuur

Wanneer we het resultaat vergelijken met de plotposities, zien we dat drie kansverdelingen slecht presteren. Gumbel en Weibull worden scheefgetrokken als gevolg van de knik rond T=4 en de LogNormaal-verdeling ontspoort bij de grote herhalingstijden. Wanneer we dit resultaat bekijken vanuit de fysica zouden we bij grotere herhalingstijden eerder een afvlakking verwachten dan een verdere stijging.

Gen.Pareto en de Exponentiële verdeling geven beide een geloofwaardig resultaat, ook vanuit de fysica geredeneerd.

# Conclusies

Bij het kiezen en fitten van kansverdelingen aan maxima vormen discontinuïteiten in de werking van het watersysteem een cruciaal onderdeel. We zien een groot aantal situaties waarbij winter- en zomerstreefpeil langdurig gehandhaafd kan worden, wat zich uit in een horizontaal liggend gedeelte van de plotposities. Om goed te kunnen fitten aan een kansverdeling is het belangrijk om deze discontinuïteiten uit te sluiten. In de statistiek staat dit bekend als *left censoring*.

Bij het gebruik van de Gumbel of Weibull-verdelingen kan dit alleen door de kansverdeling te fitten aan het relevante deel van de plotposities. Dit is echter niet ideaal: het introduceert een nieuwe onzekerheid, namelijk welke plotposities te kiezen.

In onze optiek is het daarom beter om over te stappen naar een van de kansverdeling die speciaal is ontworpen voor het werken met drempeloverschrijdingen: de Gen.Pareto-, Exponentiële of Lognormaal-verdeling. Bijkomend voordeel is dat ook daadwerkelijk alle pieken worden meegenomen; niet alleen de jaarmaxima.

# Literatuur

Lit. 1 Meteobase.nl; toetsingsreeksen voor het huidige klimaat;www.meteobase.nl

Lit. 2 Mathwave (2018); *toelichting bij het statistische programma EasyFit*; http://www.mathwave.com/articles/extreme-value-distributions.html



# BIJLAGEN

w

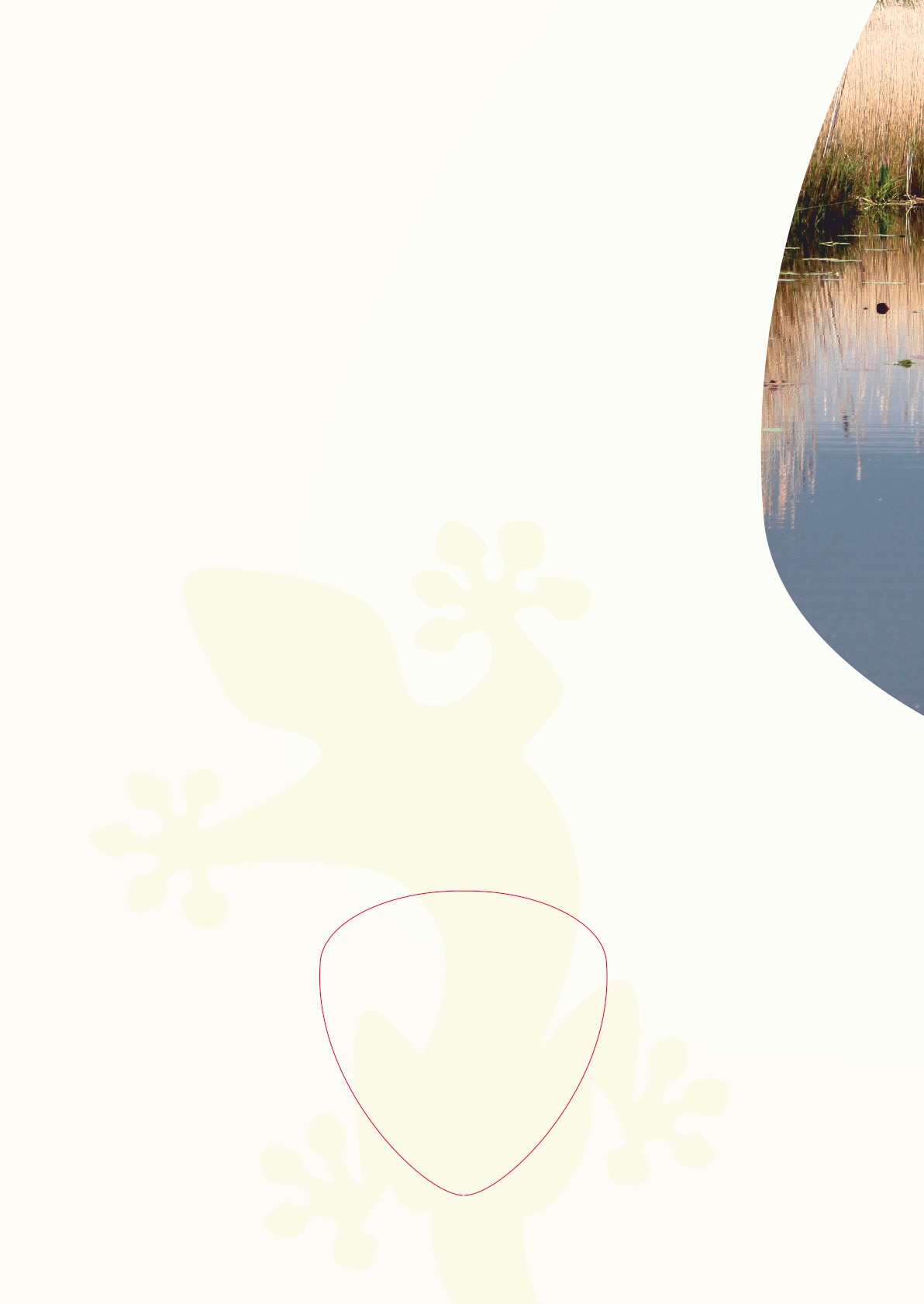
BIJLAGE A

Titel

## Inleiding

.

## Elementen



**hydroconsult©**

**Lulofstraat 55, unit 47**

**2521 AL Den Haag**

**The Netherlands**

**+31 (0)6 17 682 689**

[**hydroconsult@siebebosch.nl**](mailto:hydroconsult@siebebosch.nl)

**www.hydroconsult.info**