

# **EEN VERGELIJKENDE STUDIE VAN ALTERNATIEVE RETOURPERIODES VOOR HYDRAULISCH ONTWERP**

Jorn Van de Velde

Studentennummer: 01302409

Promotoren: Prof. dr. Bernard De Baets, Prof. dr. ir. Niko Verhoest

Masterproef voorgelegd voor het behalen van de graad in: Master of Science in de  
Bio-ingenieurswetenschappen: Land- en Waterbeheer.

Academiejaar: 2017 - 2018

De auteur en promotoren geven de toelating deze scriptie voor consultatie beschikbaar te stellen en delen ervan te kopiëren voor persoonlijk gebruik. Elk ander gebruik valt onder de beperkingen van het auteursrecht, in het bijzonder met betrekking tot de verplichting uitdrukkelijk de bron te vermelden bij het aanhalen van resultaten uit deze scriptie.

The author and promoters give the permission to use this thesis for consultation and to copy parts of it for personal use. Every other use is subject to the copyright laws, more specifically the source must be extensively specified when using results from this thesis.

Ghent, 5 juni 2018

The promoters,

The author,

Prof. dr. Bernard De Baets, Prof. dr. ir. Niko Verhoest

Jorn Van de Velde

# **DANKWOORD**

Deze thesis was een heel werk, en zou niet tot stand zijn kunnen gekomen zonder de steun van verschillende mensen. De belangrijkste zijn ongetwijfeld mijn promotoren, Prof. dr. Bernard De Baets en Prof. dr. ir. Niko Verhoest, bij wie ik op elk moment, als hun deur tenminste open stond, kon binnen vallen en vragen kon stellen, soms over de kleinste details. Zonder hun adviezen, tips en opmerkingen zou deze thesis maar de helft zijn van wat hij geworden is, en ze hebben ongetwijfeld nog veel ideeën hoe we het nog beter hadden kunnen maken.

Naast het kantoor waar ik met vragen naar toe kon gaan, was er ook het lokaal waar ik tussen het werken door plezier kon hebben. Een thesis schrijf je niet alleen, en Dries, Eva, Nele, Kato, Tawhid en Klara hebben vele uren samen met me doorgebracht, waarin gethesist is, koffies gedronken zijn en heel af en toe eens iets minder serieus gedaan werd.

Iets verder van dat kantoor en het thesislokaal zijn er nog mensen te vinden die me gesteund hebben: mijn vriendin Marie en mijn ouders, zonder wie ik niet eens had kunnen studeren.

Als laatste wil ik ook nog de Universiteit Gent bedanken, die in mijn opleiding voorzien heeft, me veel bijgeleerd heeft, maar me vooral geleerd heeft dat 'durf denken' nogal holle woorden zijn zolang je ze zelf niet uitvoert. Durven en denken moeten ook ergens beginnen, en ik hoop dat ik in beide geslaagd ben.



# **INHOUDSOPGAVE**

|                                                                              |           |
|------------------------------------------------------------------------------|-----------|
| <b>Dankwoord</b>                                                             | <b>i</b>  |
| <b>Inhoudsopgave</b>                                                         | <b>iv</b> |
| <b>Nederlandstalige samenvatting</b>                                         | <b>v</b>  |
| <b>1 Inleiding</b>                                                           | <b>1</b>  |
| 1.1 Waterbeheer en hydrologie . . . . .                                      | 1         |
| 1.2 Doel en uitwerking van deze thesis . . . . .                             | 3         |
| <b>2 Probabilistische eigenschappen van extremen</b>                         | <b>5</b>  |
| 2.1 Inleiding . . . . .                                                      | 5         |
| 2.2 Gebruik van probabilistische verdelingen . . . . .                       | 6         |
| 2.3 Univariate retourperiodes . . . . .                                      | 7         |
| 2.4 Bivariate retourperiodes . . . . .                                       | 8         |
| 2.4.1 Belang van bivariate retourperiodes . . . . .                          | 8         |
| 2.4.2 Copula's . . . . .                                                     | 9         |
| 2.4.3 Retourperiodes aan de hand van copula's . . . . .                      | 15        |
| 2.4.4 Design met bivariate retourperiodes . . . . .                          | 19        |
| 2.5 Opmerkingen met betrekking tot het gebruik van een retourperiode . . . . | 20        |
| <b>3 Methodologie</b>                                                        | <b>25</b> |
| 3.1 Ukkeltijdreeks & datavoorbereiding . . . . .                             | 25        |
| 3.2 Debiet & variabelen . . . . .                                            | 26        |
| 3.2.1 Probability Distributed Model (PDM) . . . . .                          | 26        |
| 3.2.2 Hoogte . . . . .                                                       | 29        |
| 3.2.3 Selectie van de overstromingen . . . . .                               | 30        |
| 3.3 Analyse van de data . . . . .                                            | 33        |
| 3.4 Toepassing van copula's op de data . . . . .                             | 35        |

|          |                                                         |           |
|----------|---------------------------------------------------------|-----------|
| <b>4</b> | <b>Vergelijking van de verschillende retourperiodes</b> | <b>39</b> |
| 4.1      | Overzicht van de bekomen resultaten . . . . .           | 39        |
| 4.2      | Discussie . . . . .                                     | 57        |
| <b>5</b> | <b>Conclusie</b>                                        | <b>63</b> |
|          | <b>Bibliografie</b>                                     | <b>65</b> |

# **SAMENVATTING**

Om aan de stijgende nood voor overstromingsmitigatie te kunnen voldoen, is het essentieel om te weten welke extremen zich kunnen voordoen en hoe deze gedefinieerd kunnen worden. Hiervoor wordt gebruik gemaakt van retourperiodes. Omdat overstromingen en andere natuurlijke fenomenen beschreven kunnen worden door middel van verschillende variabelen, volstaan standaard probabilistische technieken niet om de retourperiodes in meerdere dimensies te gebruiken. Om hier aan tegemoet te komen, worden copula's gebruikt, die de gezamenlijke probabilliteit kunnen beschrijven zonder de noodzaak de marginale verdeling van de variabelen exact te kennen. Door middel van deze copula's kunnen verschillende situaties voor de retourperiodes beschreven worden. In dit werk zijn de OR-, AND-, COND- en SEC-situaties onderzocht en vergeleken met data gebaseerd op waarnemingen op het KMI te Ukkel.

Door middel van het Probability Distributed Model werd een debietreeks gemaakt, waarmee op basis van enkele drempelwaarden overstromingsvariabelen berekend zijn. De Frank copula is op deze verkregen gegevens toegepast, waarmee de verschillende types van retourperiodes berekend konden worden. Uit onderzoek bleek dat de situaties gebruikt voor het definiëren van de retourperiodes zich sterk verschillend gedragen, waaruit blijkt dat het van groot belang is om bij de toepassing rekening te houden met welke situatie gebruikt kan worden.

Daarnaast wordt aanbevolen om de retourperiode als concept niet meer te gebruiken, omdat deze het gebruik van de situaties onduidelijker maakt. Interessanter is de kans op falen. Zowel bij het gebruik van bivariate retourperiodes als bij de bivariate kans op falen is een correcte keuze van ontwerpwaarden echter nog een werkpunt, zodat beide concepten in de praktijk moeilijk toe te passen zijn.





# HOOFDSTUK 1

## **INLEIDING**

### **1.1 Waterbeheer en hydrologie**

De hydrologische cyclus heeft een grote invloed op de aarde en op het leven. Water is noodzakelijk voor het leefmilieu, maar bijvoorbeeld ook voor onze voedingsproductie, de industrie en het transport. Door het waterverbruik bij al deze factoren stuurt de mens de hydrologische cyclus bij. Om hierin een duurzaam evenwicht te vinden is waterbeheer noodzakelijk. Enkel door het water goed te beheren is het mogelijk om het te blijven gebruiken en om tegelijkertijd beschermd te blijven tegen verschillende onwenselijke omstandigheden, zoals een slechte waterkwaliteit, overstromingen en droogtes.

Wanneer het over hydrologische extremen gaat, zijn vooral overstromingen en droogtes belangrijk. Vaak hebben deze een natuurlijke oorsprong, wat het onmogelijk maakt om ze tegen te houden. Het is wel mogelijk om het gevaar in te schatten en de mogelijkheid tot mitigatie zo groot mogelijk te maken.

Waterbeheer kan in een mitigatie voorzien door het bouwen van overstromingsreservoirs op rivieren om overtollig water te kunnen opvangen, of door het bouwen van spaarbekkens waarin water verzameld kan worden om in tijden van droogte aan de vraag te kunnen voldoen. Vanuit verschillende instanties wordt benadrukt dat het zelfs noodzakelijk wordt om het installeren van mitigerende maatregelen nog te intensifiëren. Zo wordt in het meest recente rapport van het IPCC (*Intergovernmental Panel on Climate Change*) aangebracht dat het onbetwistbaar is dat klimaatverandering sinds de jaren '50 zorgt voor stijgende temperaturen, waardoor de hydrologische cyclus beïnvloed wordt (IPCC, 2014). Aangezien dynamiek in rivieren een deel van deze cyclus is, wordt aangenomen dat deze ook beïnvloed zal worden, wat gevolgen heeft voor de overstromingen die door deze dynamiek ontstaan. De manier waarop dat gebeurt, is onderdeel van lopend onderzoek, maar een stijgend risico op overstromingen is voorlopig zeer aannemelijk (zie o.a. Trenberth et al. (2015), Hirabayashi et al. (2013), Schaller et al. (2016), Merz et al. (2014)). Om tegen deze toename aan overstromingen gewapend te zijn, is een goede inschatting van het gevaar, nl. de

grootte van extreme debieten, noodzakelijk. De structuren die door de waterbeheerder aangelegd moeten worden, dienen namelijk gedimensioneerd te worden zodat ze kunnen omgaan met deze extremen. Deze extremen kunnen variëren in grootte, frequentie of duur.

Daarnaast is er binnen de Europese Unie sinds 2007 de Overstromingsrichtlijn (*European Flood Directive, EFD*) die de lidstaten verplicht om het overstromingsgevaar zoveel mogelijk te beperken en het gevaar in te schatten (European Union, 2007). De inschatting van dit gevaar dient weergegeven te worden door middel van overstromingskaarten, waarop het gecombineerde effect van verschillende extremen getoond wordt.

Om aan deze verschillende eisen te voldoen, is de hydrologische wetenschap noodzakelijk. Door de theoretische studie van de verschillende elementen van de hydrologische cyclus, en het wiskundig proberen vatten van deze schijnbaar onvoorspelbare natuurlijke fenomenen en hun interacties, kan de waterbeheerder over correcte informatie beschikken en de juiste inspanningen leveren. Echter noodzaakt dat het bestuderen van extremen. Om bijvoorbeeld extreme debieten te identificeren, dienen tijdreeksen onderzocht worden. Hiervoor zijn in het ideale geval lange tijdreeksen nodig. Omdat deze vaak niet beschikbaar zijn, wordt er vaker gebruik gemaakt van stochastisch gemodelleerde tijdreeksen. Een recent voorbeeld hiervan zijn neerslag- en evoporatietijdreeksen die gekoppeld gemodelleerd worden, zoals in het werk van Pham (2016).

Een belangrijke vraag binnen de hydrologie is echter hoe deze extremen gekozen en gedefinieerd worden. De standaardmethode is de retourperiode, die beschouwd kan worden als de gemiddelde tijd tussen twee extremen van dezelfde grootte. Traditioneel wordt hierbij gebruik gemaakt van univariate variabelen (bvb. retourperiodes van stormen op basis van de hoeveelheid neerslag), maar sinds enkele jaren is het duidelijk geworden dat dergelijke aanpak in vele omstandigheden niet voldoet (Salvadori et al., 2007). Vaak is het immers van belang om rekening te houden met verschillende variabelen, die vaak gecorreleerd zijn, zoals de piekhoogte bij een overstroming, het overstromingsvolume en de overstromingsduur. Sinds de introductie van copula's in de hydrologie (zie o.a. De Michele and Salvadori (2003), Salvadori and De Michele (2004) en Genest and Favre (2007)), ontstonden veel nieuwe mogelijkheden, waardoor het onderzoek dat gebruik maakt van een multivariate aanpak een sterke boost kent. Enkele recente voorbeelden hiervan zijn Wahl et al. (2015); Vernieuwe et al. (2015); Szolgay et al. (2016); Dodangeh et al. (2017); Qian et al. (2018).

Deze copula's beschrijven de afhankelijkheid tussen meerdere variabelen, zonder de noodzaak de marginale verdeling exact te kennen. Hierdoor wordt het veel gemakkelijker om de gezamenlijke probabiliteitsverdeling van deze variabelen te beschrijven. Dit kan voor verschillende toepassingen nuttig zijn, zoals voor multivariate retourperiodes of voor het simuleren van gekoppelde variabelen.

Een probleem dat opduikt bij het gebruik van multivariate retourperiodes op basis van copula's is dat de definitie en de praktische uitvoering minder eenduidig is. Gräler et al. (2013) onderzochten hoe de verschillende retourperiodes zich ten opzichte van elkaar gedragen en stelden een praktische werkwijze om tot ontwerpvariabelen te komen voor. Serinaldi (2015) stelde dan weer voor dat multivariate, maar ook univariate retourperiodes, als concept voor te veel verwarring zorgen. Hij stelde voor om terug te grijpen naar de kans op falen van een hydraulische structuur. Het is zeker nog niet duidelijk hoe en waarmee nu concreet gewerkt moet worden. Als deze techniek dus toegepast wordt om extremen te definiëren, bijvoorbeeld in het kader van ontwerpstudies, is het van groot belang dat de interpretatie van de verschillende types retourperiodes goed gekend is en dat stil gestaan wordt bij alle vragen die hieromtrent nog bestaan.

### **1.2 Doel en uitwerking van deze thesis**

Zoals uit voorgaande discussie blijkt, is een essentieel onderdeel van het waterbouwkundig ontwerp het selecteren van de correcte waarden voor extreme gebeurtenissen. Hiervoor zullen in deze thesis retourperiodes met verschillende definities met elkaar vergeleken worden, om zo te onderzoeken of deze effectief hun nut kunnen hebben binnen het waterbeheer en indien dat het geval is, hoe deze gebruikt kunnen worden.

Om de analyses te kunnen uitvoeren, wordt gebruik gemaakt van de neerslagreeks van Ukkel, zoals beschreven door Demarée (2003), aangevuld met een tijdreeks van evapotranspiratie, zoals gebruikt door Pham (2016).

Door beide tijdreeksen te gebruiken als forcing-data voor een conceptueel neerslag-afvoermodel, met name het Probability-Distributed Model (PDM) (Moore, 2007), wordt een debietsreeks verkregen. Op basis van deze debietsreeks zullen verschillende hydraulische variabelen berekend worden die gebruikt worden binnen waterbouwkundig ontwerp, met name de piekhoogte van een overstroming en het overstromingsvolume. Deze variabelen zijn de basis voor het fitten van copula's. Door middel van deze copula's zullen ontwerpvariabelen op basis van alternatieve retourperiode-definities bepaald en met elkaar vergeleken worden voor verschillende types van eenvoudige

hydraulische structuren, zoals een keermuur, een overstromingsreservoir, en dergelijke.

Om deze aanpak duidelijker te maken, zal in Hoofdstuk 2 de probabilistische theorie van toepassing op extremen, waaronder het gebruik van retourperiodes en copula's, kort geschetst worden. In Hoofdstuk 3 wordt ingegaan op de data en hoe deze door middel van verschillende bewerkingen gebruikt worden om de copula's te bepalen. In Hoofdstuk 4 tenslotte wordt onderzocht hoe de verschillende alternatieve retourperiodes zich, met betrekking tot de hier gebruikte data, ten opzichte van elkaar gedragen.

## HOOFDSTUK 2

# **PROBABILISTISCHE** **EIGENSCHAPPEN VAN** **EXTREMEN**

### **2.1 Inleiding**

Hoewel extreme gebeurtenissen vaak willekeurig lijken voor te komen, kunnen de eigenschappen ervan wiskundig gekarakteriseerd worden. Dit kan door de weerfenomenen die aan de basis liggen van deze gebeurtenissen voor te stellen als stochastische processen waarvan de gebeurtenissen de resultaten zijn. Om dit te kunnen doen, moeten doorgaans bepaalde aannames in acht genomen worden: de belangrijkste is dat deze gebeurtenissen onafhankelijk en identiek verdeeld (*independently and identically distributed*, iid.) zijn. Als dat het geval is, kan van elke gebeurtenis gezegd worden dat deze een manifestatie is van de onderliggende verdeling. Wanneer dan gegevens van verschillende gebeurtenissen van hetzelfde fenomeen beschikbaar zijn, is het mogelijk om probabilistische verdelingen aan deze gebeurtenissen te fitten, zoals de Generalized Extreme Value (of GEV)-verdeling of de Pearson Type III-verdeling, naast nog vele anderen (Salvadori et al., 2007). Op deze manier kunnen de gebeurtenissen en hun voorkomen wiskundig bestudeerd worden. Een voorbeeld van de mogelijkheden die deze werkwijze biedt, is het inschatten van de retourperiodes die overeenkomen met elk van de gebeurtenissen. Hierbij wordt voor extremen van een bepaalde grootte de tijd berekend waarmee deze gemiddeld terugkeren. Dit stelt de waterbeheerder in staat om het gevaar gekoppeld aan overstromingen van bijvoorbeeld een bepaalde grootte duidelijker weer te geven.

Tabel 2.1: Voorbeelden van gebruikte verdelingen binnen de hydrologie en enkele toepassingen.

|          | Verdeling        | Gebruik binnen de hydrologie                                                                 |
|----------|------------------|----------------------------------------------------------------------------------------------|
| Discreet | Binomiaal        | Al dan niet overschrijden van een drempelwaarde                                              |
|          | Poisson          | Voorkomen van stormen of overstromingen                                                      |
| Continu  | Uniform          | Genereren van random variabelen op basis van complexere gebeurtenissen                       |
|          | Normaal          | Variabelen die een optelsom zijn van onafhankelijke variabelen, zoals de jaarlijkse neerslag |
|          | Lognormaal       | Intensiteit van regenval, verdeling van druppelgroottes, ...                                 |
|          | Exponentieel     | Tijd tussen neerslaggebeurtenissen die een Poissongebeurtenis volgen                         |
|          | Pearson Type III | Jaarlijkse piekhoogtes van overstromingen                                                    |
|          | GEV              | Extreme regenval, droogtes                                                                   |

## 2.2 Gebruik van probabilistische verdelingen

Hoewel de verschillende gebeurtenissen die voorkomen binnen een hydrologisch systeem divers lijken, vallen deze vaak te beschrijven met enkele eenvoudige probabilistische verdelingen. Dit is een belangrijk gegeven, want dit stelt de hydroloog in staat om de eigenschappen van deze gebeurtenissen snel af te leiden en op basis van een beperkte tijdreeks toch aannames te doen over het voorkomen van bepaalde gebeurtenissen. Tabel 2.1 (naar Chow et al. (1988)) geeft enkele gangbare verdelingen weer, samen met voorbeelden van hoe deze gebruikt kunnen worden. Wegens de verschillende toepassingen is een goede keuze van verdeling noodzakelijk.

Aangezien de toepassingen die hier bestudeerd worden altijd extremen zijn, is de GEV-verdeling het meest van toepassing is. Deze wordt gebruikt voor extreme waarden, aangezien de verdeling hiervan sterk afwijkt van de globale verdeling van het fenomeen waartoe ze behoren. Deze extreme gebeurtenissen bevinden zich namelijk in de staart van deze globale verdeling.

De cumulatieve verdeling van de GEV wordt gekenmerkt door de volgende vergelijking:

$$F_X(x) = \exp \left[ - \left( 1 - k \frac{x-u}{\alpha} \right)^{\frac{1}{k}} \right]. \quad (2.1)$$

Hierbij zijn  $k$ ,  $u$  en  $\alpha$  (waarbij  $\alpha > 0$ ) parameters die gefit dienen te worden. Deze bepalen respectievelijk de vorm, locatie en schaal van de verdeling. Afhankelijk van de waarde van  $k$  wordt gesproken van verschillende types extremewaardenverdelingen.

1. Als  $k = 0$  is er sprake van een verdeling van het EVI-type, of een Gumbel-verdeling. Deze wordt typisch gebruikt wanneer extreme regenval of andere hoge waarden bestudeerd worden.
2. Als  $k < 0$  is er sprake van een verdeling van het EVII-type, ook wel gekend als de Fréchet-verdeling. Het bereik van deze verdeling heeft als ondergrens  $\frac{u + \alpha}{k}$ .
3. Als  $k > 0$  is er sprake van een verdeling van het EVIII-type (of een Weibull-verdeling indien als variabele  $-x$  gebruikt wordt). Deze is vooral van toepassing op extreem lage waarden (zoals droogte) en het bereik heeft als bovengrens  $\frac{u + \alpha}{k}$ .

## 2.3 Univariante retourperiodes

De retourperiode is een belangrijk concept binnen het waterbeheer, omdat het een beeld geeft van het gevaar dat gekoppeld is aan een extreme waarde met een bepaalde grootte. Ze geeft de gemiddelde tijd  $\tau$  tussen twee gebeurtenissen waarvan de variabele  $X$  groter is dan een vooropgestelde drempelwaarde  $x_\tau$ .

Onder andere in Salvadori et al. (2007) wordt duidelijk getoond hoe de probabiliteit dat een gebeurtenis waarvan de variabele grootte  $X \geq x_\tau$  heeft, voorkomt (dit is  $p = P(X \geq x_\tau)$ ), gerelateerd wordt aan de retourperiode.

Indien aangenomen wordt dat alle gebeurtenissen onafhankelijk zijn, wordt een gebeurtenis  $E$  gekenmerkt door een variabele  $X$  waarvan de onderliggende verdeling onbekend is. Bij deze gebeurtenis is  $X \geq x_\tau$  of  $X < x_\tau$ . Indien  $T_i$  dan de tijd is tussen twee willekeurige gebeurtenissen  $E_i$  en  $E_{i+1}$  (met  $i = 1, 2, \dots$ ), kan ervan uitgegaan worden dat  $T_i > 0$  en dat er een  $\mu_{T_i} = \mathbf{E}(T_i)$  bestaat die eindig is. Stel dat we enkel de gebeurtenissen waarbij de variabele  $X$  grootte  $x$  aanneemt beschouwen. Als  $N_x$  het aantal gebeurtenissen is tussen 2 opeenvolgende realisaties van  $E_x$ , dan kan  $T_x$ , de tijd tussen deze 2 realisaties, beschouwd worden als:

$$T_x = \sum_{i=1}^{N_x} T_i. \quad (2.2)$$

Indien aangenomen wordt dat de tussentijden  $T_i$  tussen de gebeurtenissen i.i.d. zijn, dan is het duidelijk dat:

$$\tau_x = \mathbf{E}(T_x) = \mathbf{E}(N_x) \mu_T, \quad (2.3)$$

waarbij  $\mu_T$  het gemiddelde van  $\mu_{T_i}$  is.  $N_x$  is geometrisch verdeeld, met probabilliteit  $p_x$ :

$$p_x = P(E_x) = P(X \geq x), \quad (2.4)$$

Dus  $E(N_x) = \frac{1}{p_x}$ . Hiermee valt te stellen dat

$$\tau_x = \frac{\mu_T}{p_x}, \quad (2.5)$$

met  $p_x$  de kans op overschrijden van een bepaalde grootte  $x_\tau$  bij het voorkomen van een gebeurtenis,  $\tau_x$  de eraan gekoppelde retourperiode en  $\mu_T$  de gemiddelde tijd tussen twee gebeurtenissen. Hieruit volgt dat in het univariate geval de retourperiode van een extreme waarde met grootte  $x_\tau$  eenvoudig te bepalen is met behulp van de probabilliteit dat een gebeurtenis met variabelen van deze grootte voorkomt. Deze probabilliteit kan dan weer bepaald worden wanneer de onderliggende verdeling (bijvoorbeeld een GEV-verdeling) bekend is.

De retourperiode kan gemakkelijk gebruikt worden om een ontwerpwaarde te verkrijgen (Chow et al., 1988). Eerst moet de kans op falen  $p_M$  in de levensduur bepaald worden:

$$p_M = 1 - (F(x))^M = 1 - \left(1 - \frac{1}{\tau_x}\right)^M. \quad (2.6)$$

In deze vergelijking is  $M$  de levensduur van de structuur die ontworpen moet worden. De kans op falen  $p_M$  die hieruit verkregen wordt, kan vervolgens gebruikt worden voor een ontwerpwaarde  $x_d$ :

$$x_d = F^{-1}\left((1 - p_M)^{1/M}\right), \quad (2.7)$$

die dan toegepast kan worden om een structuur te dimensioneren.

## 2.4 Bivariate retourperiodes

### 2.4.1 Belang van bivariate retourperiodes

Hoewel univariate retourperiodes toelaten om het gedrag van extreme gebeurtenissen heel eenvoudig te beschrijven, zijn deze voor een groot deel van deze extreme



condities minder interessant. Het komt namelijk vaak voor dat verschillende variabelen samen het gedrag vaak veel beter kunnen beschrijven, zoals de piekhoogte en het overstromingsvolume bij overstromingen. In het bivariate geval is een retourperiode op zich even eenvoudig gedefinieerd als in het univariate geval:

$$\tau = \frac{\mu_T}{p}. \quad (2.8)$$

Hierbij zijn  $\tau$  en  $p$  echter niet enkel bepaald door een enkele variabele  $x_\tau$ , maar ook door een tweede variabele  $y_\tau$ . De combinatie van deze twee variabelen maakt het moeilijk om de klassieke retourperiodes te blijven gebruiken. Want hoewel deze  $p$  met behulp van gezamenlijke distributiefuncties zoals de bivariate exponentiële functie bepaald kan worden, zijn hier vele aannames en restricties aan verbonden. Zo dienen de twee verdelingen onafhankelijk te zijn, en moeten ze beide meestal van hetzelfde type zijn als de gezamenlijke verdelingen (Vandenberghé (2012); Genest and Favre (2007)).

Dit probleem werd opgelost door de introductie van copula's in de hydrologie (zie bijvoorbeeld De Michele and Salvadori (2003)). Bij copula's wordt de gezamenlijke distributie losgekoppeld van de marginale distributies. Dit maakt het gebruik van copula's zeer handig om de probabiliteit van verschillende natuurlijke fenomenen te berekenen, en de daaraan gekoppelde retourperiodes.

In dit onderdeel worden bondig de belangrijkste principes van copula's, evenals hoe hierdoor retourperiodes berekend kunnen worden, beschreven. Dit is gebaseerd op Nelsen (2006) en Salvadori et al. (2007), waarin een uitgebreidere uitleg terug te vinden is.

## 2.4.2 Copula's

### Copulatheorie

Wiskundig kunnen copula's gedefinieerd worden als volgt (Salvadori et al., 2007).

**Definitie 1. (2-copula).** *Gegeven  $\mathbb{I} = [0, 1]$ . Een 2-copula is een bivariate functie  $\mathbf{C} : \mathbb{I} \times \mathbb{I} \rightarrow \mathbb{I}$  zodanig dat:*

1. (Uniforme marginalen) voor alle  $u, v \in \mathbb{I}$ ,

$$\mathbf{C}(u, 0) = 0, \quad \mathbf{C}(u, 1) = u, \quad \mathbf{C}(0, v) = 0, \quad \mathbf{C}(1, v) = v; \quad (2.9)$$

2. (2-stijgend) voor alle  $u_1, u_2, v_1, v_2 \in \mathbb{I}$  zodanig dat  $u_1 \leq u_2$  en  $v_1 \leq v_2$ ,

$$\mathbf{C}(u_2, v_2) - \mathbf{C}(u_2, v_1) - \mathbf{C}(u_1, v_2) + \mathbf{C}(u_1, v_1) \geq 0. \quad (2.10)$$

Hierbij zijn  $u$  en  $v$  de waarden van de continue random variabelen  $U$  en  $V$ , waarbij  $U$  en  $V$  gegeven worden door:

$$\begin{cases} U = F_X(X) \\ V = F_Y(Y) \end{cases} \iff \begin{cases} X = F_X^{[-1]}(U) \\ Y = F_Y^{[-1]}(V) \end{cases}. \quad (2.11)$$

$U$  en  $V$  zijn uniform over  $\mathbb{I}$ .  $F_X$  en  $F_Y$  zijn cumulatieve distributiefuncties.  $F_X^{[-1]}$  en  $F_Y^{[-1]}$  zijn de quasi-inversen. Deze quasi-inversen zijn als volgt gedefinieerd.

**Definitie 2.** *zij  $F$  een univariate distributiefunctie. Een quasi-inverse van  $F$  is elke functie  $F^{[-1]} : \mathbb{I} \rightarrow \overline{\mathbb{R}}$  (met  $\overline{\mathbb{R}}$  de uitgebreide reële rechte, d.w.z. dat  $+\infty$  en  $-\infty$  inbegrepen zijn) zodanig dat:*

1. als  $t \in \text{Range}(F)$ , dan is  $F^{[-1]}$  elk getal  $x \in \overline{\mathbb{R}}$  zodanig dat  $F(x) = t$ , ofwel, voor alle  $t \in \text{Range}(F)$  :

$$F(F^{[-1]}(t)) = t; \quad (2.12)$$

2. als  $t \notin \text{Range}(F)$ , dan geldt

$$F(F^{[-1]}(t)) = \inf \{x \mid F(x) \geq t\} = \sup \{x \mid F(x) \leq t\}. \quad (2.13)$$

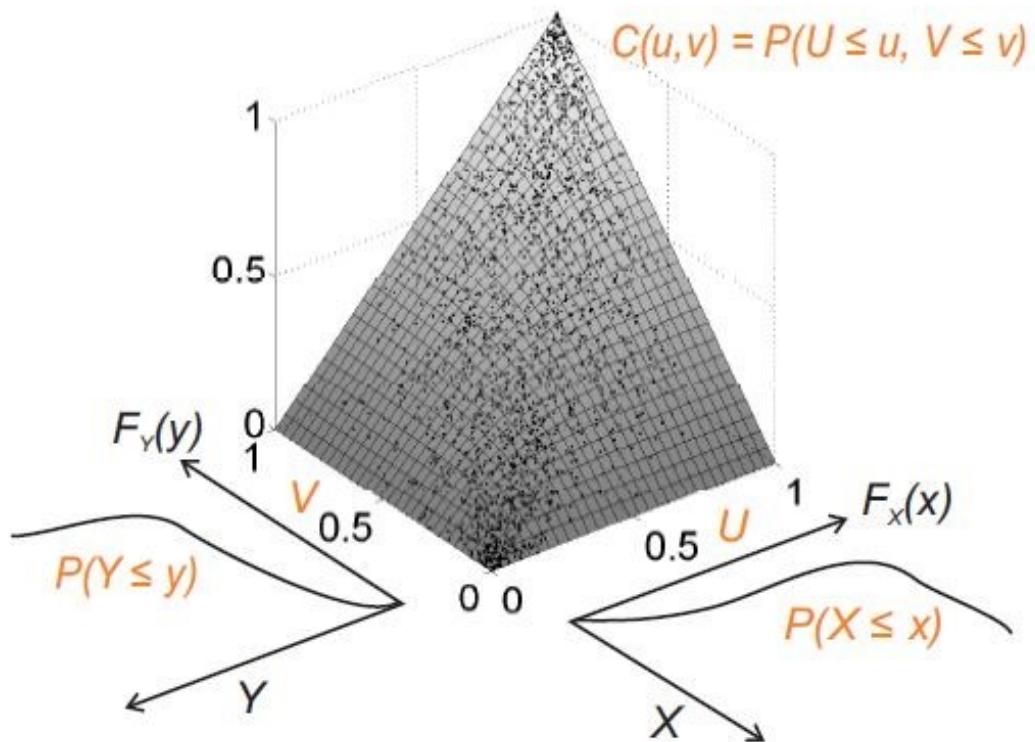
Hieruit volgt dat wanneer de distributiefunctie strikt stijgend is, de quasi-inverse gelijk is aan de gewone inverse  $F^{-1}$ .

Voor de eenvoud zal in wat volgt copula('s) geschreven worden in plaats van 2-copula('s), aangezien deze thesis enkel van toepassing is op 2-copula's. Enkel in bepaalde definities zal 2-copula nog gebruikt worden, zodat duidelijk is dat deze enkel daarop van toepassing zijn.

Het doel van de copula's in deze thesis is om gezamenlijke probabiliteit weer te geven. Dat de copula's hiervoor gebruikt kunnen worden, heeft Sklar in 1959 bewezen.

**Stelling 1.** *Er bestaat een copula  $\mathbf{C}$  zodat  $H(x, y) = \mathbf{C}[F_X(x), F_Y(y)]$ , met  $H(x, y) = P(X \leq x, Y \leq y)$  de gezamenlijke cumulatieve distributiefunctie en marginale cumulatieve distributiefuncties  $F_X$  en  $F_Y$ . Als  $F_X$  en  $F_Y$  continu zijn, is deze copula uniek.*

Het is deze stelling die de kracht van copula's aantoonst. Waar  $X$  en  $Y$  random variabelen zijn die het resultaat zijn van een stochastisch proces dat de werkelijkheid



Figuur 2.1: Voorstelling van een copula  $\mathbf{C}(u, v)$  die de marginalen  $U$  en  $V$  koppelt (Vandenberghe, 2012)

wiskundig voorstelt, zijn  $U$  en  $V$  de variabelen die uniform verdeeld zijn op basis van de marginale cumulatieve distributiefuncties van  $X$  en  $Y$ . Door hun uniforme verdeling zijn deze  $U$  en  $V$  gemakkelijk te gebruiken en gezamenlijk weer te geven. Hierin zit de kracht van copula's: ze koppelen de marginale distributies los van de gezamenlijke, doordat de uniforme verdelingen genoeg zijn om de probabiliteit te modelleren en men de onderliggende distributies niet dient te kennen. Hierdoor zijn copula's een *mapping tool* van  $\mathbb{I}^2(F_X, F_Y)$  naar  $\mathbb{I}(H)$  met  $[H(x, y), F_X(x), F_Y(y)]$  een punt in  $\mathbb{I}^3$  (Nelsen, 2003).

Naast de copula kan ook de zogenaamde survival-copula bepaald worden. Deze geeft  $P(U > u, V > v)$  en kan als volgt bepaald worden.

**Stelling 2. (Survival copula).** Als  $\mathbf{C}$  een 2-copula is en als de bivariate functie  $\bar{\mathbf{C}} : \mathbb{I}^2 \rightarrow \mathbb{I}$  gegeven wordt door

$$\bar{\mathbf{C}}(u, v) = u + v - 1 + \mathbf{C}(1 - u, 1 - v), \quad (2.14)$$

dan is  $\bar{\mathbf{C}}$  een 2-copula die de survival 2-copula van  $U$  en  $V$  genoemd wordt.

De kracht van de copula's in deze thesis ligt in het feit dat deze gezien kunnen worden al een probabiliteitsmaat. Deze probabiliteitsmaat wordt een C-maat genoemd en is gedefinieerd als.

**Definitie 3. (C-maat).** Een 2-copula  $\mathbf{C}$  induceert een probabiliteitsmaat  $\mu_C$  op  $\mathbb{I}^2$ , die de C-maat genoemd wordt en gegeven wordt door:

$$\mu_C \{[0, u] \times [0, v]\} = \mathbf{C}(u, v). \quad (2.15)$$

Meer concreet kan dus gezegd worden dat de C-maat  $\mu_C$  van een subset  $A$  van  $\mathbb{I}^2$  de probabiliteit dat twee random variabelen  $(U, V) \in \mathbb{I}^2$  met een gezamenlijke probabiliteit  $\mathbf{C}$  waarden hebben in  $A$ .

### Verschillende vormen van copula's

Om te zien hoe copula's binnen de context van de hydrologie gebruikt kunnen worden, is het belangrijk om te begrijpen dat ze in diverse vormen kunnen voorkomen. Dit is interessant voor de toepassing, aangezien er naar een zo goed mogelijke fit gezocht kan worden, maar zorgt er wel voor dat enig inzicht in de onderscheidende kenmerken van de verschillende vormen van copula's belangrijk is. Een eerste onderscheid is dat van de symmetrische versus de asymmetrische copula's. Voor symmetrische copula's geldt voor alle  $u, v \in \mathbb{I}$ :

$$\mathbf{C}(u, v) = \mathbf{C}(v, u). \quad (2.16)$$

Dit wil dus zeggen dat de random variabelen  $U$  en  $V$  uitwisselbaar zijn, of dat de copula symmetrisch is ten opzichte van zijn diagonaal. Hoewel asymmetrische copula's ook hun verdienstes hebben (Liebscher, 2008), worden deze in de hydrologie voorlopig eerder zelden gebruikt, niettegenstaande niet alle fenomenen met de symmetrische copula's correct weergegeven kunnen worden (Sadegh et al., 2017).

Copula's worden verder opgedeeld in klassen, waarvan de Archimedische één van de meest gebruikte in de hydrologie is. Dit omdat de copula's binnen deze klasse gemakkelijk te implementeren zijn, breed toegepast kunnen worden en een sterke associatie tussen de variabelen induceren (Genest et al., 1998). Archimedische copula's zijn symmetrisch en kunnen altijd als volgt weergegeven worden:

$$\mathbf{C}(u, v) = \gamma^{[-1]}(\gamma(u) + \gamma(v)), \quad (2.17)$$

waarbij  $\gamma$  een (additieve) generator is en convex dient te zijn.  $\gamma^{[-1]}$  is de pseudo-inverse van  $\gamma$ . Deze is gedefinieerd als:

**Definitie 4.** Als  $\gamma : \mathbb{I} \rightarrow [0, \infty]$  zodanig dat  $\gamma$  continu en strikt dalend is, met  $\gamma(1) = 0$  en als  $\gamma^{-1}$  de inverse is, dan is de pseudo-inverse van  $\gamma$  de functie  $\gamma^{[-1]} : [0, \infty] \rightarrow \mathbb{I}$  die gegeven wordt door

$$\gamma^{[-1]} = \begin{cases} \gamma^{-1}(t) & , \text{ als } 0 \leq t \leq \gamma(0) \\ 0 & , \text{ als } \gamma(0) \leq t \leq \infty \end{cases} \quad (2.18)$$

Hierbij geldt dat  $\gamma^{[-1]}$  gelijk is aan de gewone inverse wanneer deze strikt dalend is. Dankzij deze generator zijn Archimedische copula's handig in gebruik: deze kan meestal berekend worden door middel van de copulaparameter  $\theta$ . Bijgevolg kan de copula gemakkelijk bepaald worden indien de copulaparameter en daaruit de generator bepaald kunnen worden. Vaak is het bepalen van deze copulaparameter mogelijk met behulp van Kendall's tau,  $\tau_K$  (Kendall, 1938; Kruskal, 1958), een maat voor de associatie van twee reeksen, gebaseerd op de rangen.

Een tweede reden voor de populariteit van Archimedische copula's is dat de *level curves* expliciet uitgeschreven kunnen worden. Deze curves worden als volgt gedefinieerd.

**Definitie 5.** Gegeven  $0 < t \leq 1$ , worden de curves

$$L_t = \{(u, v) \in \mathbb{I}^2 \mid \mathbf{C}(u, v) = t\} \quad (2.19)$$

de *level curves* van  $\mathbf{C}$  genoemd.

Deze curves zijn convex bij de Archimedische copula's en laten het toe om de copula's in de ruimte  $\mathbb{I}^2$  te plotten, wat de interpretatie sterk vereenvoudigt.

Hoewel deze veel gebruikt worden, zijn Archimedische copula's zeker niet de enige die toegepast worden in de hydrologie. Ook meta-elliptische copula's ((Fang et al., 2002; Genest et al., 2007)), Gaussiaanse copula's (Wist et al., 2004; Renard, 2007), extreme waardencopula's (Salvadori and De Michele, 2007), etc. kennen hun toepassingen.

Copula's zijn onderverdeeld in verschillende families die in één of meerdere van deze klassen onderverdeeld kunnen worden. In Tabel 2.2 staan 3 regelmatig toegepaste families, die allen behoren tot de Archimedische copulaklasse. Zo wordt de Frank copula toegepast door De Michele and Salvadori (2003) om neerslagintensiteit en -duur te koppelen, is de Gumbel-Hougaard copula al gebruikt geweest om het falen van een dam door gekoppelde piekhoogte en volume in te schatten (De Michele et al.,

Tabel 2.2: Enkele copulafamilies, met hun uitdrukking, het domein van de parameter en de uitdrukking van de generator.

| Copulafamilie   | $\mathbf{C}_\theta(u, v) =$                                                                   | $\theta \in$    | $\gamma(t) =$                                         |
|-----------------|-----------------------------------------------------------------------------------------------|-----------------|-------------------------------------------------------|
| Frank           | $\frac{1}{\ln \theta} \ln \left[ 1 + \frac{(\theta^u - 1)(\theta^v - 1)}{\theta - 1} \right]$ | $[0, +\infty]$  | $-\ln \left( \frac{\theta^t - 1}{\theta - 1} \right)$ |
| Gumbel-Hougaard | $e^{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}}$                                         | $[1, +\infty]$  | $(-\ln t)^\theta$                                     |
| Clayton         | $(\max\{(u^{-\theta} + v^{-\theta} - 1), 0\})^{-1/\theta}$                                    | $[-1, +\infty]$ | $\frac{1}{\theta}(t^{-\theta} - 1)$                   |

2005) en werd de Clayton copula al aangewend om de duur en ernst van droogtes op de Gele Rivier te koppelen (Shiau et al., 2007). Andere voorbeelden zijn te vinden in Zhang and Singh (2007) (Frank, Gumbel-Hougaard), Genest and Favre (2007) (Clayton), Singh and Zhang (2007) (Frank), Shiau and Modarres (2009) (Clayton).

Als laatste dient er gezegd te worden dat er naast deze theoretische voorstelling van copula's ook gebruik gemaakt kan worden van empirische copula's (Deheuvels, 1979). De empirische copula kan gebruikt worden om het gezamenlijk gedrag van variabelen weer te geven en telt in feite het aantal paren van variabelen dat voldoet aan gegeven voorwaarde. Op die manier geven ze een benadering van de copula die het het paar variabelen  $(U, V)$  koppelt. Empirische copula's zijn gedefinieerd als.

**Definitie 6.** Wanneer  $(R_k, S_k)$  de rangen geassocieerd met een random sample  $\{(X_k, Y_k)\}$ ,  $k = 1, \dots, n$  zijn, dan is de daarmee overeenkomende empirische copula  $\mathbf{C}_n$ :

$$\mathbf{C}_n(u, v) = \frac{1}{n} \sum_{k=1}^n \mathbf{1} \left( \frac{R_k}{n+1} \leq u, \frac{S_k}{n+1} \leq v \right). \quad (2.20)$$

Hierbij zijn  $u, v \in \mathbb{I}$  en is  $\mathbf{1}$  de indicatorfunctie.

### Fitten van copula's

In de meeste toepassingen is het van belang om uit de vele copulafamilies die bestaan een goede keuze te maken, zodat de gezamenlijke verdeling zo goed mogelijk weergegeven wordt. Er zijn verschillende technieken die gebruikt worden om de *goodness of fit* zo goed mogelijk te bepalen, maar het onderzoek hiernaar blijft verder gaan. Enkele klassieke methodes en momenteel vaak gebruikte methodes worden gegeven in Genest and Favre (2007) en Salvadori et al. (2007). Deze zijn:

1. technieken gebaseerd op de grafische weergave (Genest and Rivest, 1993);
2. technieken gebaseerd op meer formele tests en de empirische copula (Genest and Rémillard, 2005; Genest et al., 2006);

3. technieken gebaseerd op de probabiliteitsintegraal-transformatie van Rosenblatt (1952) (Breyman et al., 2003);
4. technieken gebaseerd op *kernel smoothing* (Fermanian, 2005; Panchenko, 2005; Scaillet, 2007).

Waar nodig, zullen deze technieken uitgebreider besproken worden. Wat opmerkelijk is aan de opsomming van goodness-of-fitmethodes, is dat er nog steeds geen eenduidige methode is om deze te bepalen. Sadegh et al. (2017) hebben een poging gedaan om hier aan tegemoet te komen, door een toolbox te ontwikkelen. Deze toolbox berekent voor verschillende copula's de *goodness-of-fit*, de copulaparameters en hun onzekerheid. Dit doen ze ook voor enkele copula's die voorlopig maar zelden gebruikt worden in de hydrologie. Op deze manier, hopen ze meer structuur in de fit en parameterbepaling van copula's te krijgen.

### 2.4.3 Retourperiodes aan de hand van copula's

Eén van de grootste verschillen met het univariate geval, en daardoor meteen ook een sterk voordeel van het multivariate, is dat er zich verschillende situaties kunnen voordoen. Dit is afhankelijk van de eisen die gesteld worden aan de twee variabelen. In wat volgt, zullen vier van dergelijke situaties beschreven worden: de OR-, AND-, SEC- en COND-case. Er zijn meer mogelijkheden, zoals ook Serinaldi (2015) aangeeft. In Tabel 2.3 staat een korte omschrijving van de hier gebruikte situaties. Elk van deze geeft een in de praktijk sterk verschillende situatie aan, wat gevolgen heeft voor het hydraulisch ontwerp.

Een tweede belangrijke verschil is dat in tegenstelling tot het univariate geval er niet één waarde is die voldoet aan een bepaalde retourperiode. Er zijn namelijk verschillende waarden voor  $u$  en  $v$  (en dus verschillende waarden voor  $X$  en  $Y$ ) die tot dezelfde retourperiode kunnen leiden. Hierdoor komen in  $\mathbb{I}^2$  (of  $\mathbb{R}^2$  wanneer de oorspronkelijke random variabelen bekeken worden) isolijnen met dezelfde retourperiodes voor. Dit maakt de interpretatie moeilijker ten opzichte van het univariate geval. In plaats van één probabiliteit  $p$  moet er namelijk op zoek gegaan worden naar een isolijn van probabiliteiten, de level curve, waarmee de isolijn van een retourperiode berekend kan worden.

Voor de OR-situatie kan de probabiliteit omschreven worden als de kans dat één van beide variabelen overschreden wordt. Hier gaat het dus over de piekhoogte of het overstromingsvolume die een bepaalde grenswaarde overschrijden. Aangezien door de copula de gezamenlijke probabiliteit van de twee variabelen berekend kan worden,

Tabel 2.3: De verschillende bestudeerde situaties die kunnen voorkomen bij het gebruik van copula's.

| Situatie | Beschrijving                                                                       |
|----------|------------------------------------------------------------------------------------|
| OR       | Eén van de twee variabelen overschrijdt een bepaalde waarde.                       |
| AND      | Beide variabelen overschrijden een bepaalde waarde.                                |
| SEC      | Een situatie met variabelen groter dan beide variabelen vindt plaats.              |
| COND     | Een variabele wordt overschreden, gegeven dat de andere variabele overschreden is. |

is de OR-probabiliteit als volgt (aangenomen dat er met een Archimedische copula gewerkt wordt):

$$p_{u,v}^v = P(U > u \vee V > v) = 1 - \mathbf{C}(u, v). \quad (2.21)$$

Hieruit volgt dat de retourperiode uitgedrukt kan worden als:

$$\tau_{u,v}^v = \frac{\mu_T}{p_{u,v}^v} = \frac{\mu_T}{1 - \mathbf{C}(u, v)}, \quad (2.22)$$

waarbij  $\tau_{u,v}^v : \mathbb{I} \rightarrow [\mu_T, \infty]$ .

Bij AND is de probabiliteit de kans dat de twee variabelen tegelijkertijd overschreden worden: de piekhoogte en het overstromingsvolume overschrijden allebei een bepaalde grenswaarde. Deze kan met behulp van de survival copula (vergelijking (2.14)) berekend worden, of:

$$\begin{aligned} p_{u,v}^\wedge &= P(U > u \wedge V > v) = \bar{\mathbf{C}}(1-u, 1-v) \\ &= 1 - u - v + \mathbf{C}(u, v). \end{aligned} \quad (2.23)$$

Dit leidt er toe dat de retourperiode berekend kan worden als:

$$\tau_{u,v}^\wedge = \frac{\mu_T}{p_{u,v}^\wedge} = \frac{\mu_T}{1 - u - v + \mathbf{C}(u, v)}. \quad (2.24)$$

De SEC-situatie is minder eenduidig dan de twee voorgaande. De SEC, wat staat voor secundaire retourperiode (tegenover de OR-situatie als 'primaire' retourperiode), verschilt van de andere retourperiodes omdat ze niet de probabiliteit weergeeft dat een bepaalde gebeurtenis voorkomt, maar dat een gebeurtenis voorkomt die gevaarlijker is dan een gebeurtenis met een bepaalde  $x_T$  en  $y_T$ , of de equivalenten daarvan op de level curve. Als een dam een overstroming met een bepaalde piekhoogte en een



bepaald volume aankan, zal deze situatie de retourperiode geven van de mogelijkheid dat er een gebeurtenis optreedt waarbij de dam niet meer adequaat is. Dat kan in theorie een gebeurtenis zijn met een lagere piekhoogte, maar een groter overstroomingsvolume.

Om deze retourperiode te kunnen bepalen, moet er eerst een verzameling  $B_C(t)$  bepaald worden. Deze verzameling is het gebied in  $\mathbb{I}^2$  dat afgelijnd wordt door de level curve  $L_t$  en wordt gegeven door:

$$B_C(t) = \{(u, v) \in \mathbb{I} \mid \mathbf{C}(u, v) \leq t\}, \forall t : 0 < t \leq 1. \quad (2.25)$$

Nelsen (2006) heeft bepaald dat er voor deze verzameling als volgt een C-maat bepaald kan worden (zie Definitie 3).

**Stelling 3.** *Als  $0 < t \leq 1$ , dan is de functie  $K_C(t)$  de C-maat van de verzamelings  $B_C(t)$  en wordt deze gegeven door*

$$K_C(t) = t - \frac{\gamma(t)}{\gamma'(t^+)}. \quad (2.26)$$

Hierbij is  $\gamma$  de generator van de Archimedische copula en  $\gamma'$  zijn eerste afgeleide, waarbij  $t$  positief benaderd dient te worden.

Deze  $K_C$  kan als volgt geïnterpreteerd worden (Nelsen, 2006).

**Stelling 4.**  *$U$  en  $V$  zijn uniforme random variabelen waarvan de gezamenlijke distributie gegeven wordt door de Archimedische copula  $\mathbf{C}$  die gegenereerd wordt met  $\gamma$ .  $K_C$  is dan de distributiefunctie van de random variabele  $W = \mathbf{C}(U, V)$ , ofwel:*

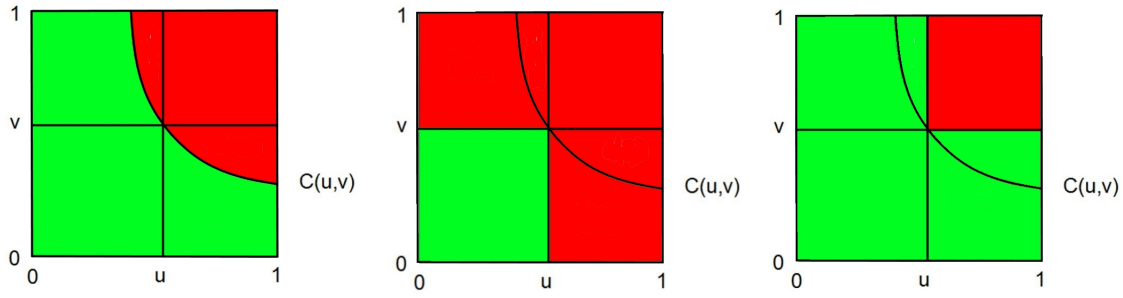
$$F_W(w) = P(W \leq w) = K_C(w), \forall w : 0 < w < 1, \quad (2.27)$$

met  $F_W(w) = 0$  voor  $w \leq 0$  en  $F_W(w) = 1$  voor  $w \geq 1$ .

Uit de twee voorgaande stellingen volgt:

$$(u, v) \in B_C(t) \iff \tau_{u,v} \leq \theta(t) \iff w = \mathbf{C}(u, v) \leq t, \forall t : 0 < t \leq 1. \quad (2.28)$$

Hierbij is  $\theta(t)$  een grenswaarde die  $\mathbb{I}^2$  opdeelt in twee ruimtes en die geassocieerd kan worden met een bepaalde probabiteit  $F_W(t)$ , die in functie van de maat  $K_C$  uitgedrukt kan worden. Hieruit volgt dat er een functie is die aangeeft dat een gebeurtenis behoort tot het complement van  $B_C(t)$ . Dit complement is de ruimte waarin de copula



Figuur 2.2: Voorstelling van 3 retourperiodedefinities (SEC, OR, AND) in  $\mathbb{I}^2$  (naar Gräler et al. (2013)). De groene oppervlaktes worden als 'veilig' beschouwd, de rode oppervlaktes zijn de kritische zones die beschouwd worden bij de berekening van de probabilliteit.

een waarde aanneemt die groter is dan de waarde van de level curve waarmee  $B_C(t)$  bepaald werd.

$$\bar{F}_W(t) = 1 - F_W(t) = \bar{K}_C(t). \quad (2.29)$$

Dit laat toe dat er een kritisch niveau  $t_*$  gedefinieerd kan worden en er een probabilliteit bepaald kan worden dat er een extremere gebeurtenis voorkomt, door middel van  $\bar{K}_C(t)$ . Door middel van deze probabilliteit kan de bijhorende retourperiode bepaald worden. Deze secundaire retourperiode is als volgt gedefinieerd:

**Definitie 7.** Als een  $t_*$  geselecteerd is, met  $0 < t_* \leq 1$ , dan is

$$\tau_* = \frac{\mu_T}{\bar{K}_C(t_*)} = \frac{\mu_T}{1 - F_W(t_*)} \quad (2.30)$$

de secundaire retourperiode geassocieerd met het kritisch niveau  $t_*$ .

Deze 3 voorgaande retourperiodes kunnen ook voorgesteld worden in  $\mathbb{I}^2$ , zoals door Gräler et al. (2013) gedemonstreerd (zie Figuur 2.2). Hierbij zijn de groene oppervlaktes deze die 'veilig' zijn volgens de respectievelijke definities en de rode de kritische zones die beschouwd worden wanneer de probabilliteit berekend wordt.

Een laatste situatie wordt beschreven door de conditionele retourperiode (COND), waarbij de probabilliteit dat  $V > v$  gegeven  $U > u$  gebruikt wordt (dit kan ook omgekeerd gedefinieerd worden, zonder verschillen, aangezien de copula symmetrisch is). Dit kan bijvoorbeeld een situatie zijn waarbij de piekhoogte groter is dan een drempelwaarde gegeven dat het overstromingsvolume al een ontwerpvolume overschreden heeft.

Om deze retourperiode te bepalen, moet eerst de retourperiode voor  $U > u$  bepaald worden, die vrij eenvoudig weergegeven wordt als volgt:

$$\tau_{U>u} = \frac{\mu_T}{1-u}. \quad (2.31)$$

Deze geeft de gemiddelde tijd tussen twee gebeurtenissen met grootte  $u$  en kan dus gebruikt worden als tussentijd voor de volgende stap, waarbij er rekening gehouden moet worden met de probabiliteit  $P(U > u \wedge V > v)$ . De conditionele retourperiode  $\tau_v|u$  kan dus als volgt berekend worden:

$$\tau_v|u = \frac{\tau_{U>u}}{P(U > u \wedge V > v)} = \frac{\mu_T}{(1-u)(\bar{\mathbf{C}}(1-u, 1-v))}. \quad (2.32)$$

Hierbij moet de noemer op basis van vergelijking (2.23) als volgt vereenvoudigd worden:

$$(1-u)(\bar{\mathbf{C}}(1-u, 1-v)) = (1-u)(1-u-v + \mathbf{C}(u, v)) \quad (2.33)$$

waardoor ook deze retourperiode berekend kan worden.

Interessant aan deze verschillende retourperiodes is dat deze gebaseerd zijn op dezelfde concepten, en daardoor ook met elkaar gelinkt zijn. Hierdoor is het mogelijk de probabiliteiten waarop deze gebaseerd zijn volgens grootte te ordenen. Yue and Rasmussen (2002) hebben dit bijvoorbeeld gedaan, waarvoor voor één punt  $(U, V)$  de verschillende probabiliteiten vergeleken werden. Dit gaf het volgende resultaat:

$$p_{\text{OR}} \geq \max(p_U, p_V) \geq \max(p_U, p_V) \geq p_{\text{AND}} \geq p_{\text{COND}}. \quad (2.34)$$

Dit heeft als gevolg dat de retourperiodes zich als volgt gedragen (in eenzelfde punt):

$$T_{\text{OR}} \leq \min(T_U, T_V) \leq \max(T_U, T_V) \leq T_{\text{AND}} \leq T_{\text{COND}}. \quad (2.35)$$

#### 2.4.4 Design met bivariate retourperiodes

Net als bij univariate retourperiodes is het uiteindelijke doel bij bivariate retourperiodes om bepaalde ontwerpwaardes te vinden waarmee een hydraulische structuur ontworpen kan worden. In tegenstelling tot het univariate geval zijn deze retourperiodes immers van toepassing op verschillende combinaties van variabelen  $U$  en  $V$ , wat het ondubbelzinnig bepalen moeilijk maakt. Er zijn al enkele pogingen gedaan om een degelijke werkwijze te vinden om deze bivariate retourperiodes te gebruiken,

zoals door Gräler et al. (2013). Zij stellen voor om door middel van ensembles te werken, maar ze hebben geen methode gevonden die breed voor alle mogelijkheden van toepassing is. Dit is dus nog onderdeel van lopend onderzoek.

## 2.5 Opmerkingen met betrekking tot het gebruik van een retourperiode

Hoewel de retourperiode als concept sterk ingeburgerd is en vaak gebruikt wordt, vallen er wel enkele kanttekeningen bij te maken. Zo haalt Serinaldi (2015) aan dat het concept nog steeds slecht toegepast en naar het publiek vertaald wordt, zelfs door mensen die het concept goed zouden moeten kennen. Uitspraken zoals 'deze storm met een retourperiode van 50 jaar kan nu niet voorkomen, want hij is enkele jaren geleden nog maar geweest' geven een slecht beeld van de probabiliteit die achter de retourperiode zit (in tegenstelling tot 'deze storm zal elk jaar met een kans van  $\frac{1}{50}$  kunnen voorkomen'). Op deze manier wordt het gevaar vaak slecht gecommuniceerd, wat het begrip bij het publiek niet verbetert.

Deze communicatiefout is al het onderwerp van discussie bij univariate retourperiodes, maar bij de bivariate/multivariate retourperiodes is dit een nog groter probleem doordat er verschillende definities zijn. Hierdoor ontstaat het gevaar dat door de bomen het bos niet meer te zien is. Wat is immers in welke situatie van toepassing? Yue and Rasmussen (2002) hebben hier al duidelijkheid in proberen scheppen, maar volgens Serinaldi (2015) kunnen hun conclusies voor meer verwarring zorgen. Zij stelden immers dat een ontwerp op basis van aparte univariate verdelingen voor een meerkost ten opzichte van een gezamenlijke verdeling kan zorgen, en de univariate frequentie-analyse dus geen goed beeld geeft van de multivariate gebeurtenis. Een univariate en multivariate toepassing gaan echter over intrinsiek verschillende situaties, wat uit deze uitleg niet duidelijk blijkt.

Er zijn wel enkele antwoorden in deze richting op de vele mogelijkheden die bivariate retourperiodes geven. Zo stelde Shiau al in 2003 dat er gekeken moet worden naar wat er gebeurt:  $\tau_{OR}$  is vooral interessant wanneer beide variabelen apart gevaar kunnen veroorzaken,  $\tau_{AND}$  dan weer wanneer beide nodig zijn om een gevaar te veroorzaken. Een ander voorstel is om de secundaire retourperiode vaker te gebruiken, zoals gedaan door Salvadori and De Michele (2010, 2013), Durante and Salvadori (2010), Vandenberghe et al. (2011) en anderen. Dit deden ze omdat één gezamenlijke probabiliteit gepaard gaat met verschillende (in feite oneindig veel) mogelijkheden voor OR- en AND-situaties, in tegenstelling tot de duidelijke situatie die door de secundaire retourperiode gegeven wordt. Dit is aangegeven in Figuur 2.3, waarbij in het do-

mein van de secundaire retourperiodes naast  $u$  en  $v$  ook  $v'$ ,  $u'$ ,  $v''$  en  $u''$  aangeduid staan. Voor elk van deze  $u$ 's en  $v$ 's kunnen er OR- en AND-retourperiodes gedefinieerd worden, die alle verschillend zijn, ook al liggen ze op dezelfde level curve. In de secundaire retourperiode is er voor deze ene level curve één retourperiode, wat een duidelijkere situatie is.

Hierbij is het probleem echter dat ook dit type retourperiode een andere situatie beschrijft dan de OR- en AND-situaties, waardoor het niet zomaar als vervanging gebruikt kan worden. Deze onvervangbaarheid tussen de verschillende definities van retourperiodes blijft bestaan, ondanks de pogingen om ze mathematisch te vergelijken (zie Vergelijkingen (2.34) en (2.35)). Deze vergelijkingen zijn meestal gebaseerd op puur mathematische regels en daarom niet altijd veelzeggend. Dit is verduidelijkt in figuur 2.3, waarin getoond wordt hoe de verschillende situaties gedefinieerd zijn over verschillende domeinen en voldoen aan verschillende voorwaarden. Serinaldi (2015) stelde daarop voor om de retourperiode als concept minder naar voor te schuiven, en terug te grijpen naar de kans dat een extreme gebeurtenis plaats vindt, als onderliggende grondgedachte. Hoewel de retourperiode door de weergave in jaren als gebruiksvriendelijk overkomt, is hier teveel verwarring aan verbonden, in tegenstelling tot probabiliteitsweergaven. Wat bijvoorbeeld in elke omstandigheid duidelijk gebruikt kan worden, is de kans op falen van een structuur tijdens een levensduur  $M$ . Deze wordt gegeven door (Chow et al., 1988):

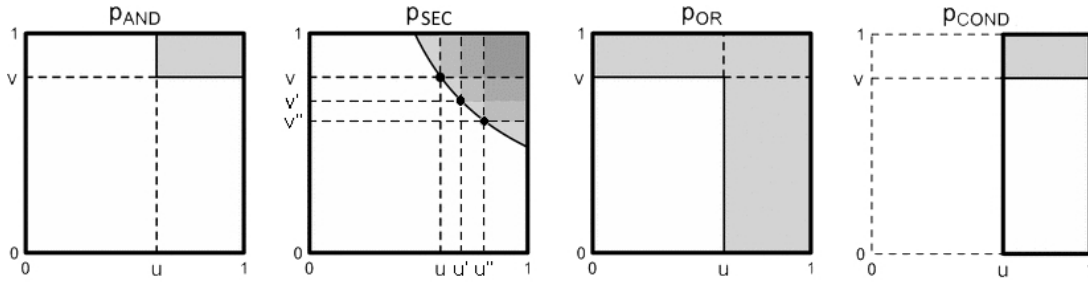
$$p_M = 1 - (F(x_d))^M. \quad (2.36)$$

Hierbij is  $x_d$  de ontwerpwaarde voor de variabele waarmee rekening gehouden moet worden. Deze waarde wordt bij het gebruik van retourperiodes op een veel te omslachtige manier verkregen, vindt Serinaldi (2015), terwijl deze vergelijking duidelijk is. In het multivariate geval kan deze vergelijking als volgt geschreven worden:

$$p_M = 1 - H_M(E_1, E_2, \dots, E_M), \quad (2.37)$$

waarbij  $E_j$ ,  $j = 1, \dots, M$ , een gebeurtenis voorstelt die bepaalt of er al dan niet een falen optreedt. Dit kan elk van de gebeurtenissen zoals beschreven bij de retourperiode zijn:  $\{E_j \mid X_j \leq x_d \wedge Y_j \leq y_d\}$ ,  $\{E_j \mid X_j \leq x_d \vee Y_j \leq y_d\}$  of elke andere situatie die bij het ontwerp voor kan komen.  $H_M$  is de gezamenlijke kans op voorkomen van deze gebeurtenissen doorheen de levensduur van de waterbouwkundige structuur.

Ondanks dit interessante voorstel blijven er wel enkele vragen. Een eerste belangrijke vraag is hoe deze multivariate kans op falen concreet uitgewerkt kan worden.



Figuur 2.3: Voorstelling van het domein waarover de AND, SEC, OR en COND gedefinieerd zijn (dikke zwarte lijnen), en de kritische regio's (grijs) die voldoen aan de voorwaarde voor de betreffende situaties, naar Serinaldi (2015). In de secundaire situaties zijn ook  $u'$ ,  $v'$ ,  $u''$  en  $v''$  aangegeven, die andere mogelijke combinaties voor de OR- en AND-retourperiode aangeven ten opzichte van  $u$  en  $v$ .

Tabel 2.4: Overzicht van de formules voor de bivariate kans op falen, met  $M$  de levensduur van een hydraulische structuur.

| Situatie | Formule                                                 |
|----------|---------------------------------------------------------|
| OR       | $p_M^V = 1 - (\mathbf{C}(F_X(X), F_Y(Y)))^M$            |
| AND      | $p_M^A = 1 - (\overline{\mathbf{C}}(F_X(X), F_Y(Y)))^M$ |
| SEC      | $p_M^{\text{SEC}} = 1 - K_C(t)^M$                       |

Het onderzoek hiernaar is beperkt, maar een aanzet is bijvoorbeeld al gegeven door Salvadori et al. (2016). Zij beschrijven hoe deze ideeën, samen met de door hen geïntroduceerde Hazard Scenario's gebruikt kunnen worden om de basisideeën van bivariate of multivariate retourperiodes beter toe te passen en stellen voor de verschillende situaties een formule voor, die opnieuw op de copula's gebaseerd is. De formules voor de kans op falen  $p_M$  gedurende de levensduur  $M$  van een hydraulische structuur zijn gegeven in Tabel 2.4.

Daarnaast moet er ook stil staan worden bij het juiste gebruik van een situatie in de bepaalde omstandigheden. Hier is al enig onderzoek naar gebeurd, maar door de vele mogelijkheden, is dit verre van afgerond. Salvadori et al. (2016) hebben enkele voorbeelden samengevat. Een voorbeeld waar de relevantie van de OR-situatie al heel duidelijk aangetoond is, is het samenvloeien van twee rivieren. Hierbij kan een overstroming namelijk veroorzaakt worden door de ene rivier, de andere, of beide rivieren samen (Favre et al., 2004; Wang et al., 2009; Bender et al., 2013). Een ander voorbeeld is het falen van de uitloop van een dam, dat zowel door volume als debiet apart veroorzaakt kan worden (De Michele et al., 2005; Salvadori et al., 2011). Een voorbeeld voor de AND-situatie dat gegeven wordt, is dat van Dung et al. (2015). Zij bestudeerden de Mekong Delta, waar men voorbereid is op overstromingen. Overstromingen zijn daar pas een gevaar wanneer de verschillende variabelen (overstromingsvolume en piekdebiet) samen zeer groot worden. Een ander voorbeeld is te vinden in Shiau (2006), waarbij de ernst en duur van droogtes gekoppeld zijn. Uit

Corbella and Stretch (2012) blijkt dat de secundaire retourperiode goed gebruikt kan worden om kusterosie te analyseren. In het geval van de conditionele situatie, zijn minder eenduidige conclusies beschikbaar.

Het gebruik van deze multivariate kans op falen en de toepassingen van de verschillende situaties zijn dus nog volop in onderzoek, maar de kritiek op de retourperiode blijft een belangrijk punt. Dat dient dan ook verder doorheen deze thesis in gedachten gehouden te worden.

## 2.5. OPMERKINGEN MET BETREKKING TOT HET GEBRUIK VAN EEN RETOURPERIODE



## HOOFDSTUK 3

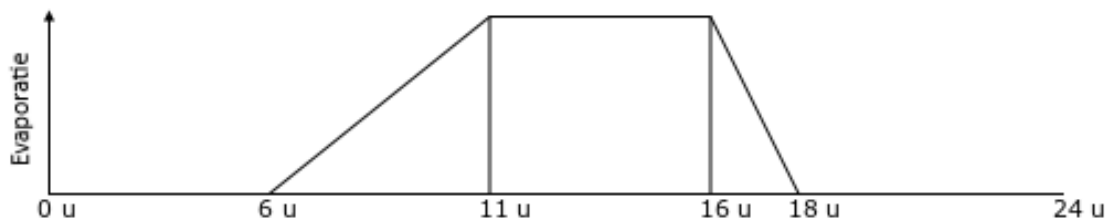
# **METHODOLOGIE**

### **3.1 Ukkeltijdreeks & datavoorbereiding**

De data die voor deze thesis gebruikt worden, zijn allemaal afkomstig of gebaseerd op metingen bij het KMI te Ukkel. De neerslagtijdreeks, met een resolutie van 10 minuten, is gemeten van 01/01/1898 tot 31/12/2002 met een Hellmann-Fuess pluviograaf (Demarée (2003)). Gezien de lengte en resolutie is dit een vrij unieke dataset binnen de hydrologie, die al gebruikt is voor verschillende toepassingen (zoals in Verstraeten et al. (2006), De Jongh et al. (2006) en Willems (2013)). Daarnaast is, zoals Ntegeka and Willems (2008) aangeven, deze tijdreeks gedurende de hele periode met dezelfde meetinstrumenten gemeten en verwerkt, wat tot een grote consistentie in de data leidt.

Deze data werden gebruikt in een hydrologisch model om er debieten mee te simuleren. In dit hydrologisch model zijn er twee types van forcingdata nodig, nl. neerslag en potentiële evaporatie. Neerslag is rechtstreeks beschikbaar via de Ukkeltijdreeks, maar voor evaporatie werd een tijdreeks, bepaald door middel van de Penman-vergelijking (Penman, 1948), gebruikt die eveneens door Pham (2016) werd toegepast. Hierin zijn de dagelijkse evapotranspiratiewaarden [mm/dag] van 1931 tot 2002 gegeven. Door deze dagelijkse waarden te verwerken naar uurlijkse data, kunnen ze gecombineerd worden met dat deel van de neerslagreeks tussen 1931 en 2002.

Deze verwerking naar uurlijkse data gebeurt door de dag op te delen in uren met en zonder zonlicht (of dus dag en nacht). In de nacht vindt er geen instraling plaats en is er bijgevolg geen evapotranspiratie. Deze uren worden voor de eenvoud het jaar rond gedefinieerd als de uren tussen 18 u 's avonds en 6 u 's ochtends. Gedurende de rest van de dag wordt de evapotranspiratie echter ook niet als evenwaardig beschouwd: ze wordt zodanig verdeeld dat de evapotranspiratie stijgt van 6 u tot en met 10 u en afneemt vanaf 16 u. Vanaf 11 u tot 16 u is de evapotranspiratie dus maximaal. Deze procedure is weergegeven in Figuur 3.1. Deze werkwijze garandeert dat de evaporatie zich op een vrij natuurlijke wijze gedraagt. De winst ten opzichte van een



Figuur 3.1: Verdeling van de evaporatie over de tijd.

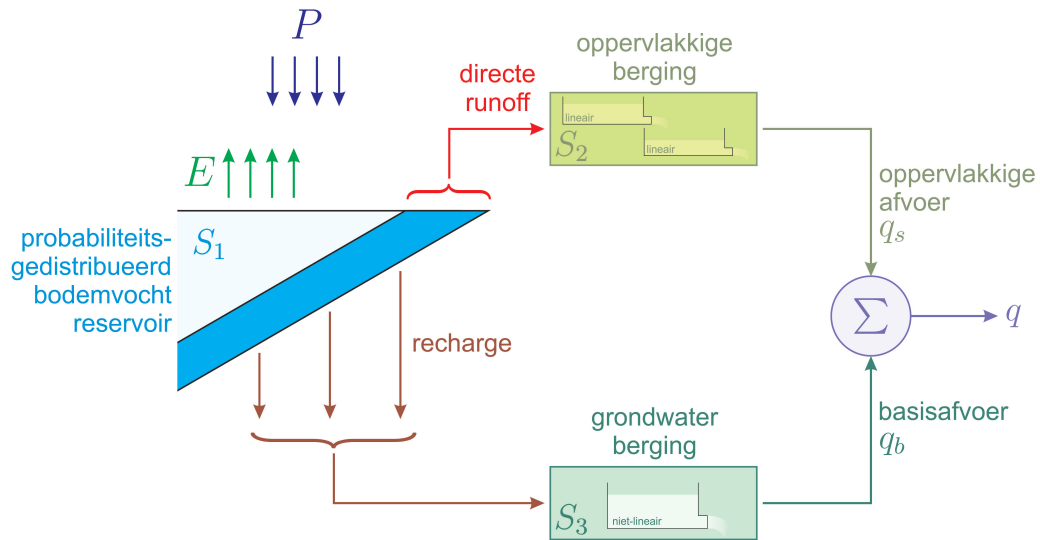
lineaire verdeling van de evaporatie over een dag, is echter zeer beperkt gebleken. Hierdoor valt de vraag te stellen of dit nuttig is.

## 3.2 Debiet & variabelen

### 3.2.1 Probability Distributed Model (PDM)

Het model om debietreeksen te bekomen is het PDM (Moore, 2007; Cabus, 2008), zoals gebruikt door Pham (2016). Dit is een geaggregeerd model dat debieten modelleert op basis van de veronderstelling dat het stroomgebied conceptueel voorgesteld kan worden als een collectie van drie verschillende reservoirs, waarvan in Figuur 3.2 een overzicht gegeven wordt. Een meer gedetailleerde uitleg is terug te vinden in Moore (2007).

Het eerste reservoir ( $S_1$ ) is een probabilistisch verdeeld bodemvochtreservoir. Hierbij wordt gebruik gemaakt van een Pareto verdeling om weer te geven welke fractie van het stroomgebied een bepaalde bergingscapaciteit heeft. Het tweede reservoir ( $S_2$ ) stelt het oppervlaktereservoir voor, dat de directe runoff omzet in oppervlakte-afvoer. Als laatste is er een reservoir dat het grondwatertransport voorstelt ( $S_3$ ), om zo de recharge om te zetten in basisafvoer. De input voor  $S_1$  is de netto neerslag ( $P - E$ ), met  $P$  de neerslag en  $E$  de evapotranspiratie.  $S_1$  kan daarnaast water verliezen door directe runoff en recharge. Zoals hierboven staat, vormt het reservoir  $S_2$  de directe runoff om naar oppervlakte-afvoer  $q_s$  door middel van twee lineaire reservoirs met kleine tijdsconstanten  $k_1$  en  $k_2$ . De aanvulling van het grondwater wordt gecontroleerd door de tijdsconstante  $k_g$  en de omzetting naar basisafvoer  $q_b$  door  $S_3$  gebeurt met een traag reagerend niet-lineair reservoir. Dit heeft een grote tijdsconstante  $k_b$ . De som van  $q_s$  en  $q_b$  bepaalt uiteindelijk de totale afvoer  $q$ . Eventueel kan een con-

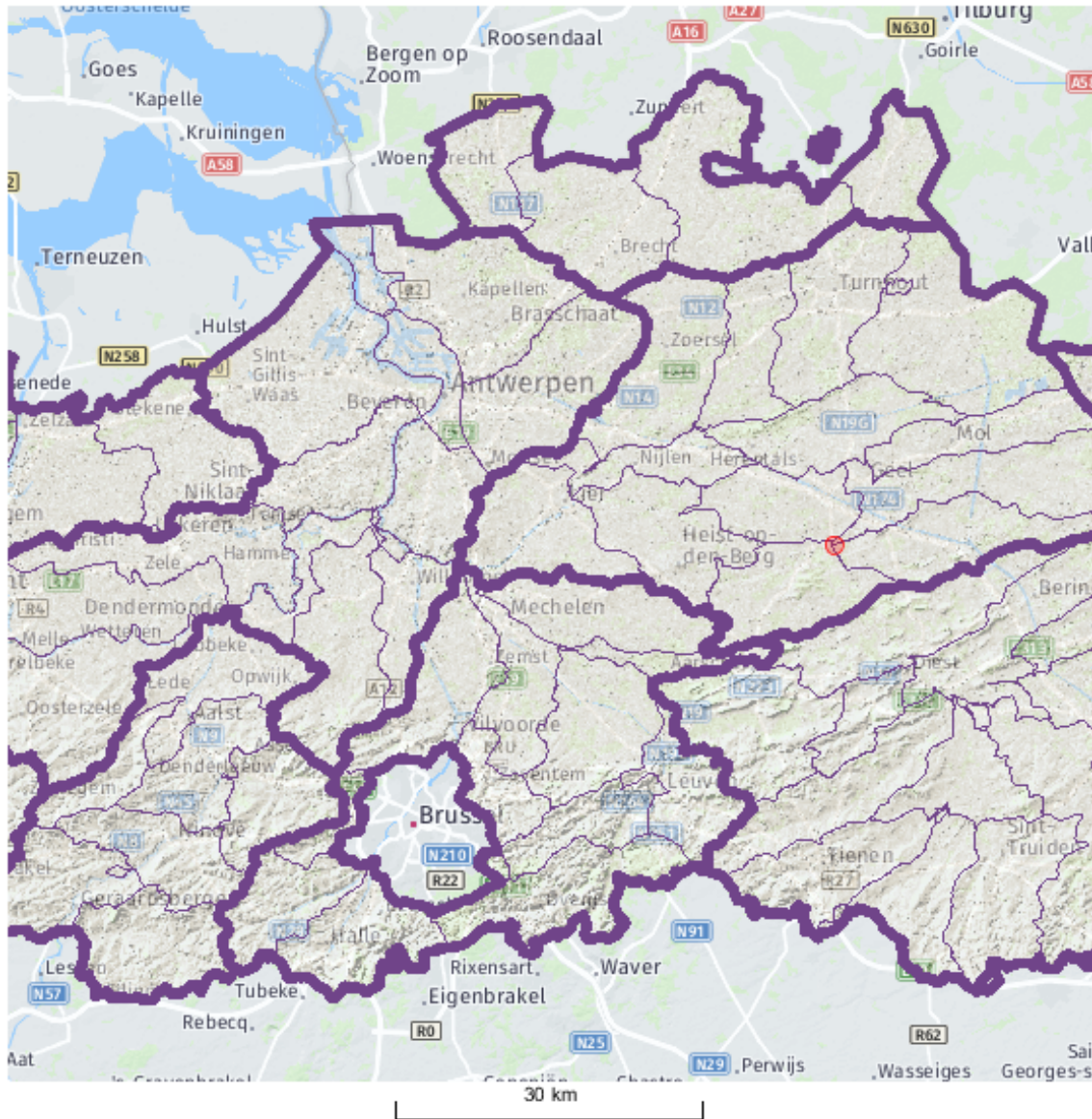


Figuur 3.2: Het PDM neerslag-afvoermodel

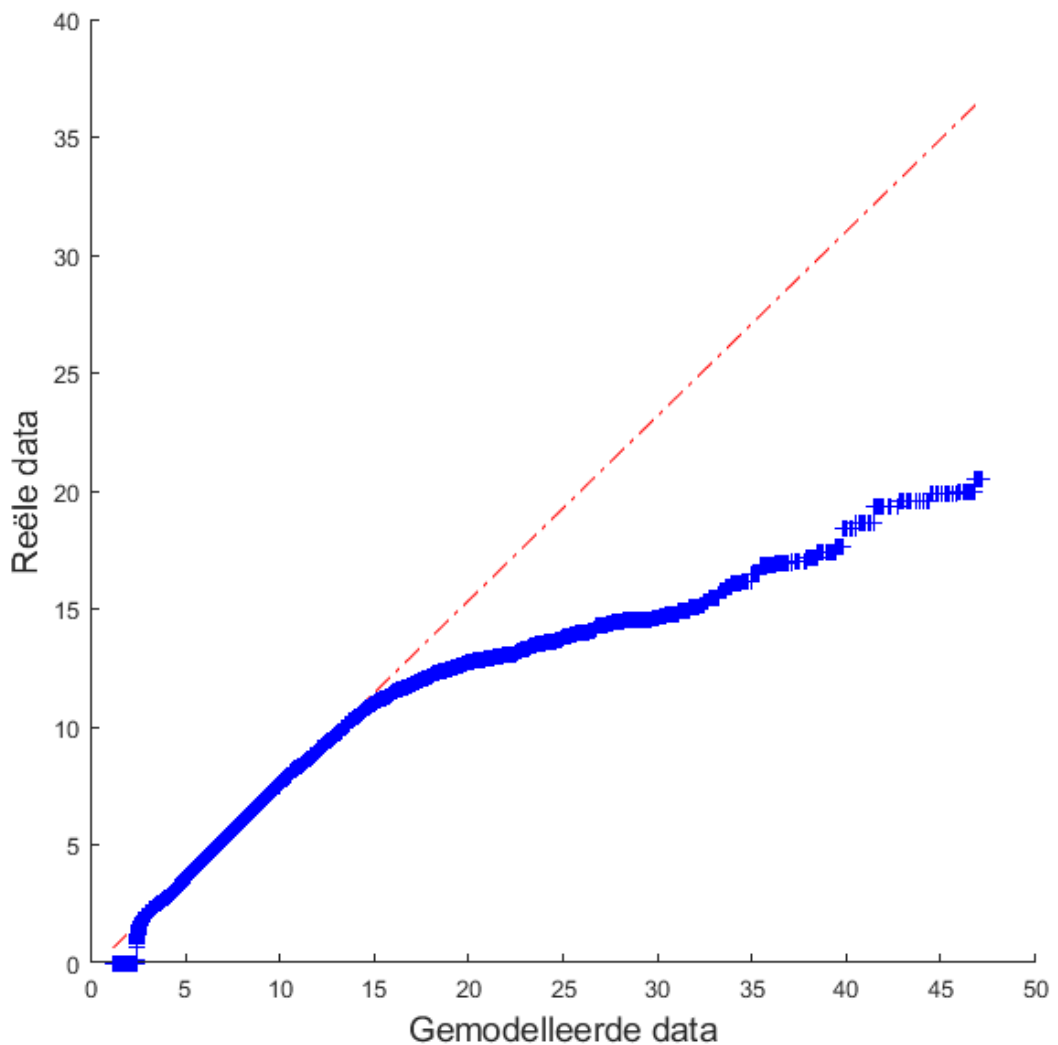
stante afvoerparameter  $q_{const}$  toegevoegd worden, om het onttrekken of toevoegen van extra water in het stroomgebied in rekening te brengen.

Het PDM dat hier gebruikt is, werd door Cabus (2008) gekalibreerd op basis van het Grote Netebekken door middel van het Particle Swarm Optimization algoritme (PSO) (Eberhart and Kennedy, 1995). Om doorheen de rest van de verwerking consistentie te houden, is telkens wanneer gebruik gemaakt wordt van hydrologische data een meetpunt genomen dat in dit bekken ligt. Meer specifiek is gekozen voor het meetpunt dat een bekkenoppervlakte heeft die gelijkaardig is aan de bekkenoppervlakte die gebruik is in het model, met name  $385 \text{ km}^2$ . Het meetpunt dat hieraan voldoet bevindt zich in Geel-Zammel op de Grote Nete. De data hiervan zijn beschikbaar op [waterinfo.be](http://waterinfo.be) (VMM, 2018). Op basis van de ingegeven tijdreeks van 72 jaar is hieruit een debietstijdreeks van 631128 waarden bekomen, die als input dient voor de volgende stappen.

Ter controle is deze gemodelleerde debietreeks vergeleken met deze die voor het meetpunt beschikbaar is. Deze gaat van 31 december 1984 tot 14 maart 2018 (laatste datum van controle), met enkele onderbrekingen. In de gemodelleerde debietreeks zitten enkele piekdebieten die niet voorkomen in de werkelijke debietreeks en vice versa, maar er zijn meer overeenkomstige pieken dan pieken die maar in één van beide tijdreeksen voorkomen. De verschillen kunnen verklaard worden doordat de originele neerslag en evaporatie voor Ukkel gelden, terwijl de debietreeks op de Grote Nete gemeten is. De buien die op beide locaties gepasseerd zijn, zijn niet per se dezelfde, waardoor de verschillen ontstaan.



Figuur 3.3: Locatie van het meetpunt Geel-Zammel (rood punt) (kaart verkregen via [www.geopunt.be](http://www.geopunt.be) op 14/05/2018).



Figuur 3.4: QQ-plot van de gemodelleerde debieten t.o.v. de reële debieten.

Deze twee reeksen zijn statistisch vergeleken door middel van een QQ-plot (Figuur 3.4). Deze figuur bevestigt de conclusie dat de reeksen niet geheel overeenkomen: bij de hogere waarden van de data zijn er grote verschillen tussen beide tijdreeksen. Vooral in de extremen zitten er verschillen, wat ook in de QQ-plot duidelijk is.

### 3.2.2 Hoogte

Om overstromingen te bepalen, dient de hoogte van het water op elk beschikbaar moment berekend te worden. Dit kan door een zogenaamd Q-h-verband op te stellen en hiermee de debieten om te zetten in hoogtes. Dit verband is uniek voor elk punt in de rivier. Bijgevolg dient gebruik gemaakt te worden van de data van Geel-Zammel, het eerder vermelde meetpunt. De ruim 30 jaar aan beschikbare data van dit meetpunt is hiervoor volledig gebruikt, om een zo groot mogelijke verscheidenheid aan water-

hoogtes en debieten beschikbaar te hebben. Er zijn twee belangrijke opmerkingen bij deze tijdreeksen. Ten eerste bestaan ze uit dagelijkse data, in tegenstelling tot de gemodelleerde tijdreeksen. Aangezien deze niet rechtstreeks met elkaar gecombineerd gaan worden, is dit echter niet van belang. Daarnaast is de waterhoogte waarmee dit verband is opgesteld, de waterhoogte t.o.v. de bedding. Na het toepassen van de bekende vergelijking op de gemodelleerde debieten, wordt de hoogte van de bedding t.o.v. van de zeespiegel (13.48 m TAW) er bijgeteld. Vanaf dan wordt de waterhoogte in m TAW gebruikt voor verdere berekeningen, zoals het fitten van de copula. In feite is dit geen noodzaak, maar dit geeft een duidelijker beeld van wat er gaande is.

Om het verband op te stellen is gekozen voor een logaritmisch verband. Deze keuze is gebaseerd op de afzwakking van de invloed van het debiet: het logaritme zorgt ervoor dat hogere debieten voor een kleinere stijging in de waterhoogtes zorgen. Om de parameters te bepalen, wat hier gebeurde op basis van initiële waarden, was trial & error nodig. Op deze manier is het volgende verband gekomen:

$$h(m) = \log(Q) + 0.3615 \quad (3.1)$$

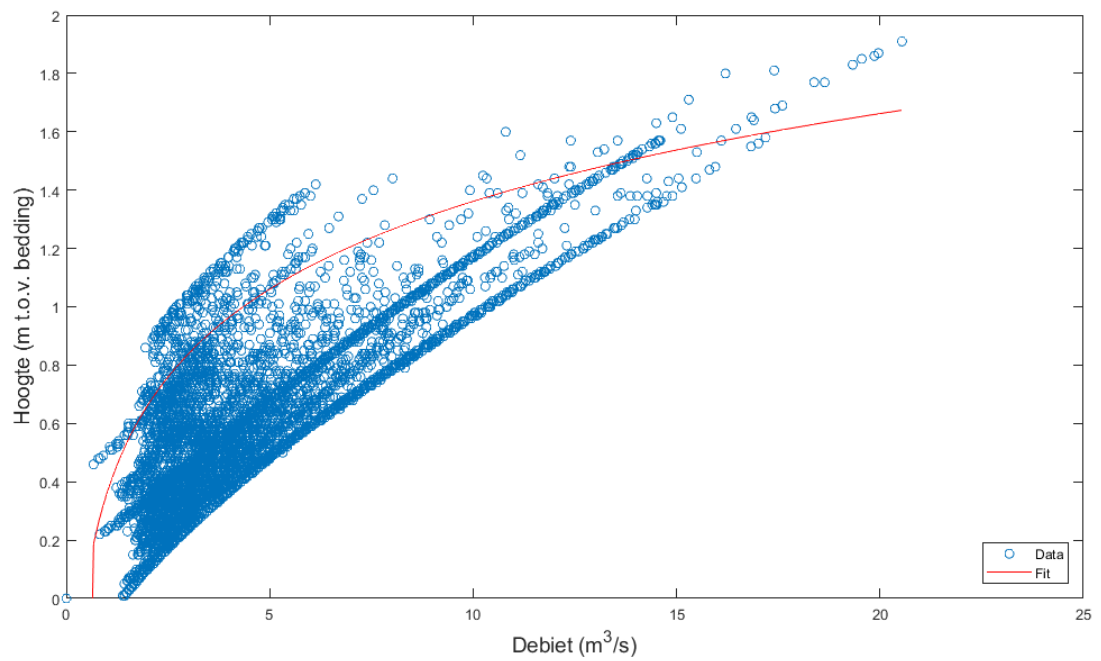
Zoals te zien is op Figuur 3.5 is er veel spreiding op de data. Dit zou veroorzaakt kunnen zijn door veranderingen in de rivier doorheen de tijd, waardoor het Q-h-verband beïnvloed wordt. Het verband geeft nog steeds werkbare resultaten, dus dit is geen groot probleem, hoewel een lichte onderschatting van de hoogtes bij hoge debieten kan plaats vinden en een mogelijke overschatting bij de lagere debieten.

Dit verband geeft in combinatie met de gemodelleerde debieten een hoogtetijdreeks die in Figuur 3.6 weergegeven wordt. Uit deze figuur en de bijhorende data blijkt dat er synthetische hoogtes tussen 13.77 m TAW en 15.68 m TAW voorkomen, wat dus neerkomt op een bereik van 1.91 m.

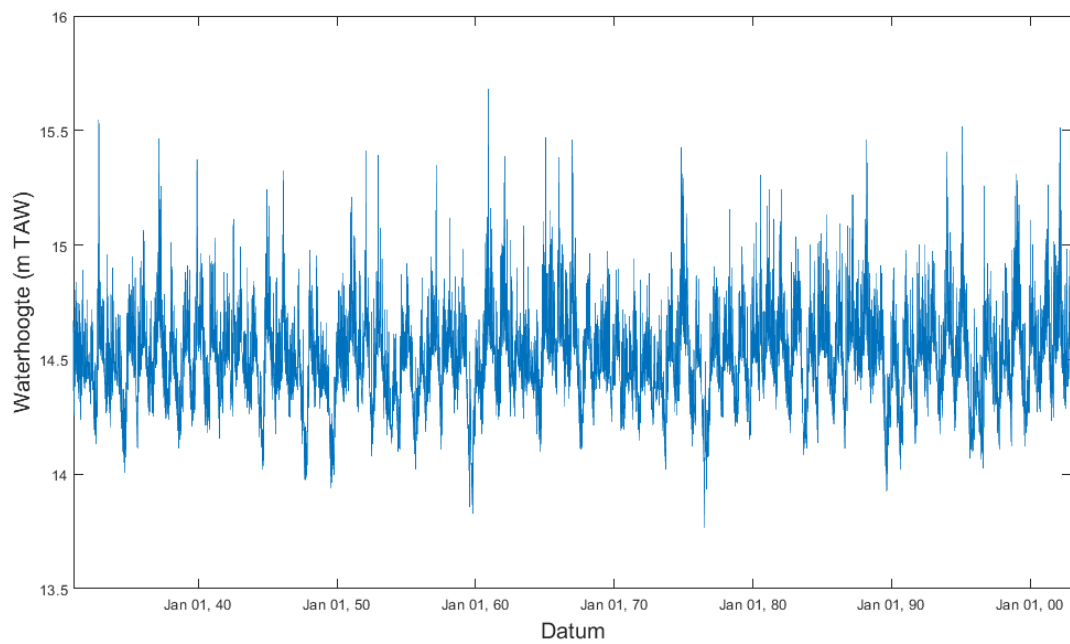
### 3.2.3 Selectie van de overstromingen

Op basis van de synthetische overstromingstijdreeks kunnen de verschillende overstromingen bepaald worden. Om dit te kunnen doen, moeten echter een aantal drempelwaarden vastgelegd worden: vanaf wanneer er over een overstroming gesproken wordt en welke tijd noodzakelijk tussen twee overstromingen moet liggen zodat deze onafhankelijk zijn.

In principe kan een overstroming gedefinieerd worden als een te hoog waterniveau waarbij de oevers van een rivier overschreden worden. In dit geval, zijn er echter geen gegevens over de oeverhoogte. Om enige algemeenheid te behouden is er voor



Figuur 3.5: Verband tussen het debiet en de waterhoogte



Figuur 3.6: Tijdreeks van de hoogte

Tabel 3.1: Geteste grenswaarden en het aantal overstromingsgebeurtenissen dat bekomen is. De hoeveelheid voor een grenswaarde van 75% ligt hoger dan deze voor 50% door het onafhankelijk houden van de overstromingen.

| Grenswaarde                | 50% | 75% | 90% | 99% |
|----------------------------|-----|-----|-----|-----|
| Hoeveelheid overstromingen | 930 | 970 | 620 | 94  |

gekozen om voor de overstromingshoogte een drempelwaarde te selecteren. Deze drempelwaarde is een theoretische hoogte waarboven er sprake is van overstromingen en is hier berekend door middel van het procentueel aantal hoogtes waarbij er sprake is van overstromingen. Verschillende waardes werden getest (Tabel 3.1), maar uiteindelijk bleek dat met een drempelwaarde van 90% (ofwel: enkel de 10% hoogste waterhoogtes geven aanleiding tot een overstroming) er een goed evenwicht was tussen het aantal overstromingsgebeurtenissen (dat hoog genoeg dient te zijn om relevante berekeningen op te doen) en de relevantie van een overstroming. Dit kwantiel komt overeen met een hoogte van 14.75 m TAW.

De tijd tussen twee overstromingen speelt ook een rol. Dit is gebaseerd op het idee dat wanneer gebeurtenissen als deze bestudeerd worden, ze beschouwd moeten worden als statistisch onafhankelijk. Voor stormen is dit al uitvoerig bestudeerd, waarbij Restrepo-Posada and Eagleson (1982) een formule opstelden om de tijd te berekenen die zich tussen twee stormen moet bevinden om ze als onafhankelijk te kunnen beschouwen. Verhoest et al. (1997) testten deze formule voor België en bekwamen een naar boven afgeronde tussentijd van 24 u. Deze tijd is echter niet als dus van toepassing op overstromingen. Wanneer neerslag een verhoogd debiet veroorzaakt, zit hier namelijk een vertragend effect op. Wanneer aangenomen wordt dat een storm over een volledig bekken van toepassing is, is er een bepaalde tijd nodig vooraleer de neerslag die het verst van de uitlaat van een bekken valt, dit punt bereikt. Dit is de zogenaamde concentratietijd. Hierdoor zit er een verschil op het neerslaghydrogram en dat van het debiet (Chow et al., 1988). Hier is gekozen voor een tussentijd van 36 u om onafhankelijke overstromingen te bekomen. Dit was een subjectieve keuze, en eventueel verder onderzoek kan deze keuze verduidelijken.

Deze twee drempelwaarden, de tussentijd en de hoogte van overstromingen, worden gebruikt om de overstromingen en overeenkomstige variabelen mee te bepalen. Dit gebeurt door de waterhoogtetijdreeks te overlopen en telkens wanneer de 90%-hoogte overschreden wordt een nieuwe overstromingsgebeurtenis aan te maken. Van deze gebeurtenis worden de begintijd, eindtijd (wanneer de drempelwaarde opnieuw overschreden wordt), de maximale waterhoogte, het maximale debiet en het volume bij gehouden. Hierbij wordt het overstromingsvolume berekend door het debiet af te toppen. Vervolgens wordt per tijdstap de hoeveelheid die afgetopt werd verme-



nigvuldigd met de gebruikte tijdstap. Uiteindelijk worden al deze volumes vanaf de begintijd tot de eindtijd van de overstroming opgeteld.

De nodige tijd tussen twee overstromingen wordt hierbij in rekening gebracht door bij het overlopen van de hoogtetijdreeks te controleren hoeveel tijd er tussen twee opeenvolgende overstromingen is. Als deze kleiner is dan de tussentijd, wordt geen aparte gebeurtenis aangemaakt voor deze overstroming, maar wordt de piekhoogte berekend over de twee oorspronkelijke overstromingen heen, wordt het volume van beide overstromingen opgeteld, geldt als begintijd van de nieuwe overstroming de begintijd van de eerste overstroming en als eindtijd de eindtijd van de tweede overstroming. Op deze manier wordt gegarandeerd dat de gegevens van de overstromingen zo onafhankelijk mogelijk zijn.

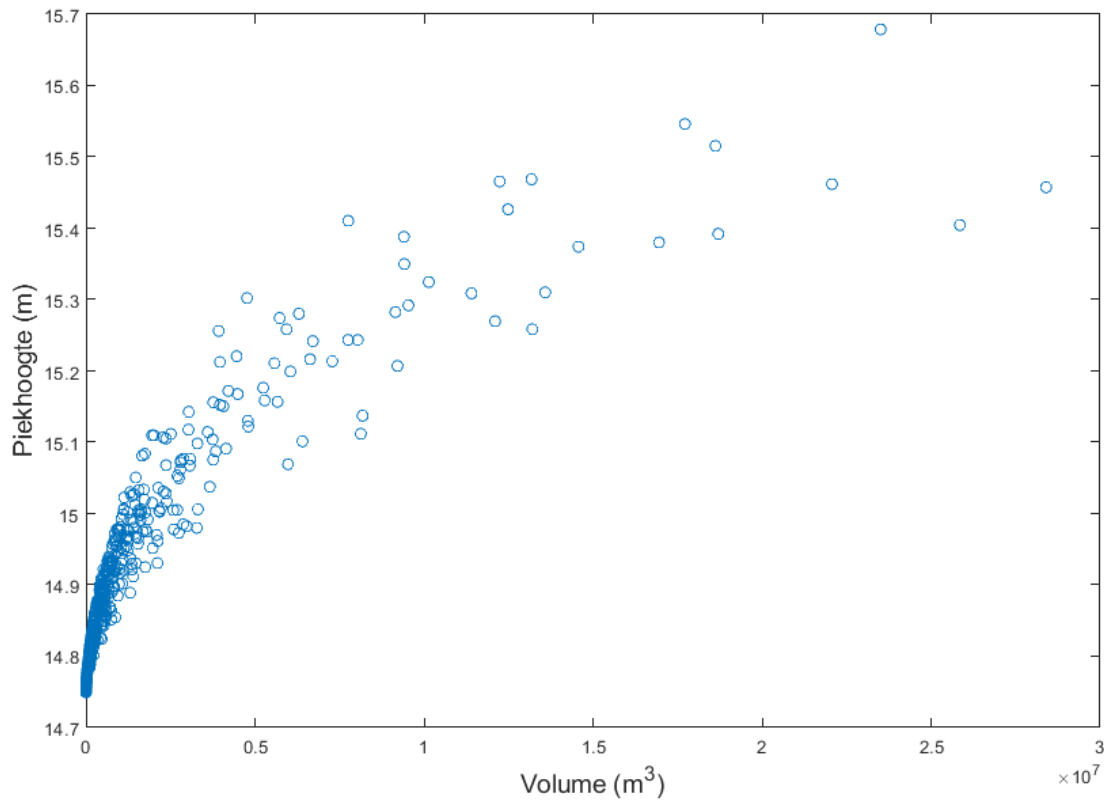
De combinatie van al de voorgaande stappen zorgt er voor dat op basis van de gemodelleerde tijdreeks van 631128 waarden een reeks bekomen is met 620 overstromingen, die verder geanalyseerd zullen worden en waarop een copula gefit zal worden.

### **3.3 Analyse van de data**

Alvorens de copula te fitten op de data, is het noodzakelijk om meer inzicht te krijgen in de geselecteerde overstromingen. Enkele algemene gegevens, met betrekking tot de variabelen piekhoogte, overstromingsvolume, tijdsduur en piekdebiet van de overstromingen, staan in Tabel 3.2. Hierbij vallen enkele zaken op. Een eerste is de sterke spreiding van enkele variabelen. Bij de piekhoogte is de spreiding beperkt (ongeveer 0.9 meter), terwijl deze bij het volume en de overstromingsduur relatief gezien veel groter is. Een relatief kleine stijging in piekhoogte kan dus al voor grote veranderingen in het volume zorgen, of omgekeerd: grote veranderingen in het volume, het piekdebiet of de duur van een overstroming, gaan vaak gepaard met een kleine verandering in de piekhoogte.

Daarnaast dient opgemerkt te worden dat deze gegevens niet noodzakelijk tot dezelfde overstroming behoren: de overstroming met het maximale volume ( $3.07 \cdot 10^7 \text{ m}^3$ ) is een andere dan deze met de maximale overstromingsduur (998 u).

Al laatste dient opgemerkt te worden dat, hoewel dit hoge waarden lijken, dit niet per se het geval is. Om op een oppervlak van  $385 \text{ km}^2$  een volume van  $3.07 \cdot 10^7 \text{ m}^3$  te bekomen, moet er ongeveer 80 mm aan neerslag vallen. Uit klimatologische data blijkt dat er in de periode die hier gemodelleerd is 155.5 mm neerslag viel, waardoor er, rekening houdende met evaporatie, nog voldoende overblijft om 80 mm te voorzien.



Figuur 3.7: Piekhoogte t.o.v. overstromingsvolume van de geselecteerde overstromingen

Tabel 3.2: Overzicht van enkele variabelen.

|            | Piekhoogte (m) | Volume (m <sup>3</sup> ) | Overstromingsduur (u) | Piekdebiet (m <sup>3</sup> ) |
|------------|----------------|--------------------------|-----------------------|------------------------------|
| Minimum    | 14.75          | 21.84                    | 2                     | 8.09                         |
| Gemiddelde | 14.89          | $1.36 \cdot 10^6$        | 101.75                | 12.18                        |
| Maximum    | 15.68          | $3.07 \cdot 10^7$        | 998                   | 68.76                        |

Naast deze variabelen werden ook de gemiddelde tijd tussen events,  $\mu_T$  en de correlatie uitgedrukt door Spearmans  $\rho_S$  bepaald. Deze bedragen respectievelijk 0.1163 jaren en 0.9879. De hoge waarde voor Spearmans  $\rho_S$  wijst op een sterke correlatie, zoals ook te zien is in Figuur 3.7.

Ondanks het belang van copula's blijft het interessant om ook het marginale gedrag van de variabelen te kennen. Om dit te bestuderen, is de GEV-verdeling op het overstromingsvolume en de piekhoogte gefit. Dit is schijnbaar de meest toepasselijke verdeling, aangezien het hier allemaal extreme events betreft. In Tabel 3.3 staan de parameters die hiervoor bekomen zijn. Wat hierbij opvalt, is het verschil tussen beide waarden voor  $k$ . Waar er van uit gegaan kon worden dat de verdeling hier een Gumbel-verdeling zou benaderen (aangezien het hier om afgeleiden van extreme regenval gaat), is deze in geen van beide gevallen ongeveer 0. In het geval van de piekhoogte kan nog geargumenteed worden dat de waarde laag genoeg is (0.51) om

Tabel 3.3: Parameters voor de GEV-verdeling van de twee variabelen

| Reeks               | k    | $\alpha$  | u         |
|---------------------|------|-----------|-----------|
| Piekhoogte          | 0.51 | 0.07      | 14.81     |
| Overstromingsvolume | 1.66 | 195082.95 | 106547.95 |

een Gumbel-verdeling te zijn, in het geval van het overstromingsvolume (1.66) is dit twijfelachtig en kan eerder voor een Weibull-verdeling gekozen worden.

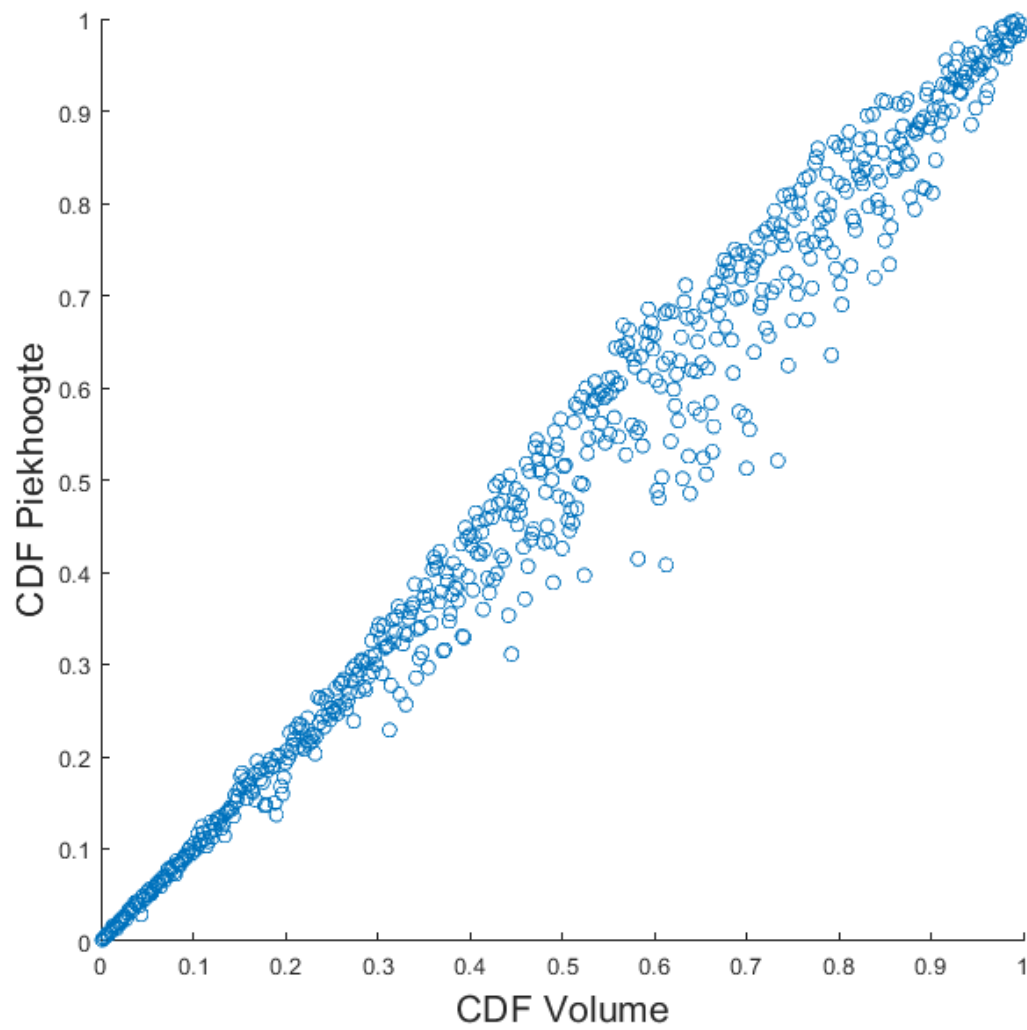
Op basis van deze werkelijke data kunnen ook de empirische CDF's opgesteld worden, wanneer gebruik gemaakt wordt van de rangen van de data. Dit zorgt ervoor dat het fitten aan en gebruik van een marginale verdeling voor deze data overbodig wordt. De Kendalls  $\tau_K$  is 0.9165, wat net als Spearmans  $\rho_S$  wijst op de hoge correlatie die deze twee datareeksen hebben. Door de CDF's van de twee variabelen ten opzichte van elkaar te plotten, wordt Figuur 3.8 bekomen. In deze figuur komt de hoge correlatie die Kendalls  $\tau_K$  al aantoonde weer naar voor. Deze sterke correlatie is belangrijk, want dit betekent dat het fitten van een copula aan de data niet strikt noodzakelijk is: de afhankelijkheid tussen de twee datareeksen kan evengoed door middel van eenvoudigere technieken, zoals een regressie beschreven worden. In dit geval is de copula echter nodig om de bivariate retourperiodes te bepalen. Omdat de correlatie sterk is, is het fitten van een specifieke copula echter minder van belang.

### 3.4 Toepassing van copula's op de data

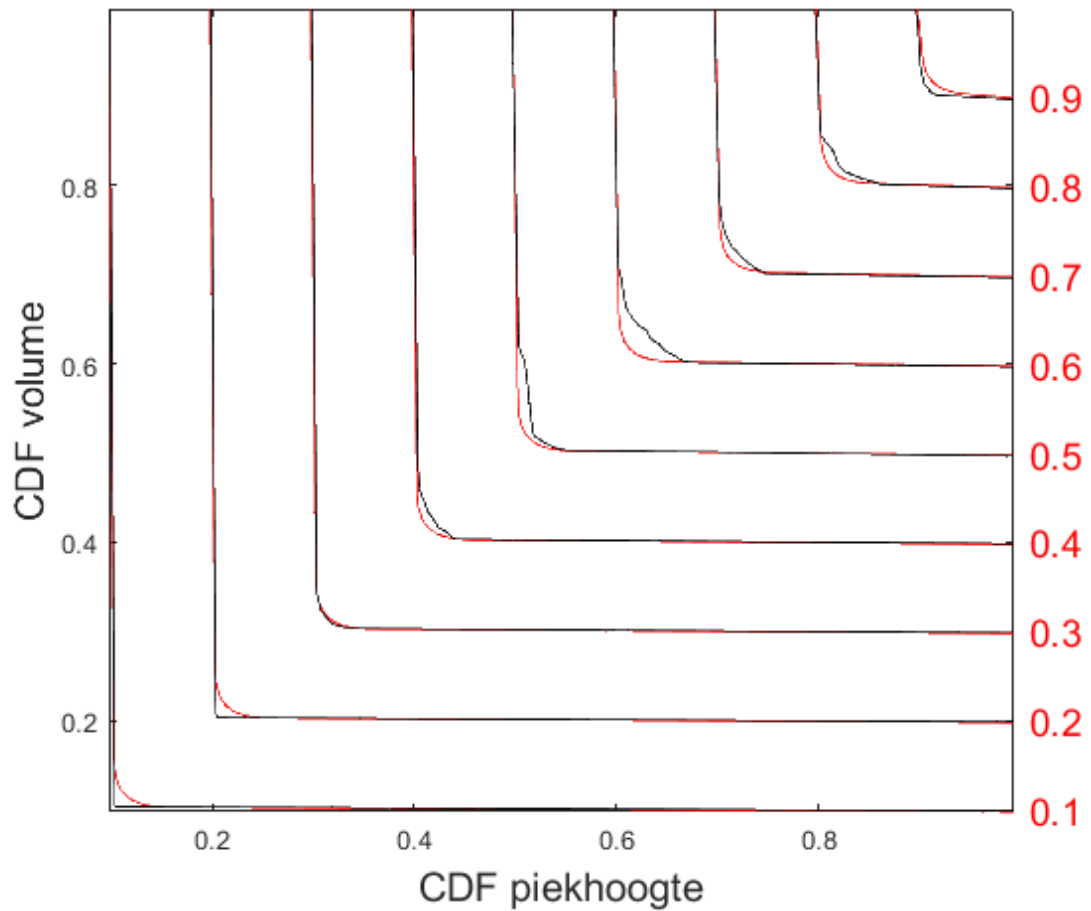
Doordat de correlatie tussen de twee datareeksen zo groot is, is het zoeken naar een beste fit van de copula minder noodzakelijk: de meeste die in aanmerking komen zullen gelijkaardige resultaten geven. De goodness-of-fitmethodes gegeven in Hoofdstuk 2 zullen dus niet gebruikt worden. Een copula is wel nog nodig om de retourperiodes te berekenen. Er werd gekozen om de Frank copula te gebruiken. Zoals in Hoofdstuk 2 beschreven, is dit een Archimedische copula waarvoor de generator bepaald kan worden door middel van de copulaparameter  $\theta$ . Deze parameter kan dan weer bepaald worden met behulp van Kendalls  $\tau_K$ , door middel van (Salvadori et al., 2007):

$$\tau_K(\theta) = 1 - 4 \frac{D_1(-\ln\theta) - 1}{\ln\theta}. \quad (3.2)$$

Hierbij is  $D_1$  de Debye functie van orde 1 (Luke, 1969). De hierop gebaseerde functie om via Kendalls  $\tau_K$  de copulaparameter te bepalen, is geïmplementeerd in Matlab. Dit leverde voor  $\theta$  de waarde 46.2199 op.



Figuur 3.8: Empirische CDF's van de piekhoogte t.o.v. overstromingsvolume



Figuur 3.9: De gefitte Frank copula (rood) versus de empirische copula (zwart)

Vervolgens kon met deze parameter de copula bepaald worden. Het resultaat hiervan is te zien in Figuur 3.9, waarbij de gefitte copula samen met de empirische copula geplot is. In de figuur is duidelijk te zien dat de empirische copula de gefitte goed benadert, wat betekent dat de keuze voor de Frank copula te ondersteunen valt. Deze copula kan dan gebruikt worden om de retourperiodes te bepalen, met de formules zoals gegeven in Hoofdstuk 2.



## HOOFDSTUK 4

# VERGELIJKING VAN DE VERSCHILLENDE RETOURPERIODES

### 4.1 Overzicht van de bekomen resultaten

Door middel van de Frank copula zijn retourperiodes voor de vier verschillende situaties (OR, AND, COND en SEC) bekomen. Aangezien deze retourperiodes gelden voor een, in principe, oneindige lijst van punten, kunnen ze weergegeven worden met behulp van isolijnen. Deze isolijnen zijn voor de verschillende situaties gevisualiseerd in Figuren 4.1, 4.2, 4.3, 4.4. In deze figuren vallen alvast een aantal zaken op. Zo is te zien dat er door de berekeningen retourperiodes kunnen voorkomen die groter zijn dan de lengte van de gebruikte tijdreeks. Enkel in het geval van OR zijn de retourperiodes kleiner dan 72 jaar, terwijl in de andere gevallen retourperiodes tot 200 of zelfs 500 jaar kunnen voorkomen.

De grote verschillen tussen de verschillende situaties vallen te verklaren door het proces dat ze weergeven. Bij de OR moet maar één van beide waarden de drempelwaarde overschrijden om effect te hebben. Dit betekent dat zelfs de extremen die een hoge gezamenlijke copula-waarde hebben, relatief gemakkelijk voorkomen. Hieruit volgt dat de meeste gebeurtenissen voorkomen met een zeer lage retourperiode, en enkel de gevaarlijkste een retourperiode hebben die meer dan enkele tientallen jaren omvat. Bij de AND is het een ander proces dat speelt: beide extremen moeten bereikt worden. De mogelijkheid dat dit gebeurt, ligt lager, wat verklaart dat voor dezelfde waarden  $U$  en  $V$  een hogere retourperiode bekomen wordt dan in de OR-situatie. Door de berekeningswijze met de survival copula (zie Vergelijkingen (2.23) en (2.24)) en de sterke correlatie kan deze retourperiode zelfs buiten het oorspronkelijke interval (72 jaar) liggen. Bij de grotere waarden van  $U$  en  $V$  is de waarde van de

gezamenlijke verdeling amper groter dan deze van  $U$  en  $V$ , of zelfs kleiner, waardoor  $1 - u - v + \mathbf{C}(u, v)$  zeer klein wordt en de retourperiode bijgevolg zeer groot.

In het conditionele geval speelt ongeveer hetzelfde effect: de survival copula wordt gebruikt, wat er voor zorgt dat dezelfde vergelijking als bij de AND-situatie in de noemer staat (zie Vergelijkingen (2.32) en (2.33)). Hierbij staat in de noemer ook nog  $1 - u$ , wat ervoor zorgt dat de waarde bekomen via  $1 - u - v + \mathbf{C}(u, v)$  nog kleiner wordt. Dit is zichtbaar in Figuur 4.3, waar de hogere retourperiodes bij lagere waarden voor  $U$  en  $V$  voorkomen dan in Figuur 4.2.

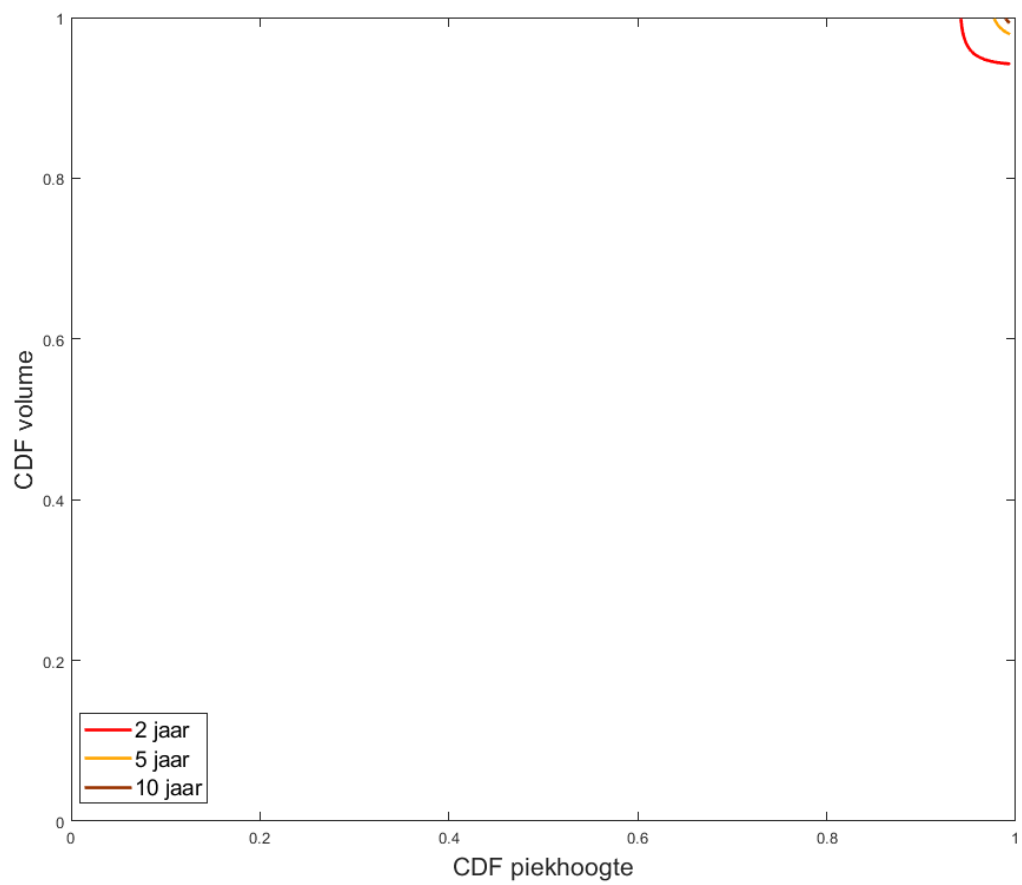
Bij de secundaire retourperiode wordt de kans bepaald dat een gebeurtenis voorkomt uit het oppervlak dat begrensd wordt door de level curve met een gegeven probabieliteit. Het geeft dus de gemiddelde kans op voorkomen van een van deze gebeurtenissen, waardoor zelfs voor lage extremen (die deel zijn van een level curve met lage probabieliteit) deze retourperiode al hoog ligt. Hierdoor induceren de hogere extreme waarden zeer grote retourperiodes, tot 500 jaar.

Wat nog opvalt, zijn de sterke verschillen tussen de isolijnen en hun vorm, die ontstaat door de berekeningswijze. Bij de OR is de berekening voor de retourperiodes het meest eenvoudig op basis van de copula, hierdoor benadert de vorm van de isolijn van de retourperiodes zeer sterk die van de level curves van de copula. In het secundaire geval geldt ook dat de isolijn vergelijkbaar is met de level curve, daar deze laatste gebruikt worden om de isolijnen te berekenen. Bij de AND en de secundaire retourperiode is de formule ingewikkelder en gebaseerd op de survival copula.

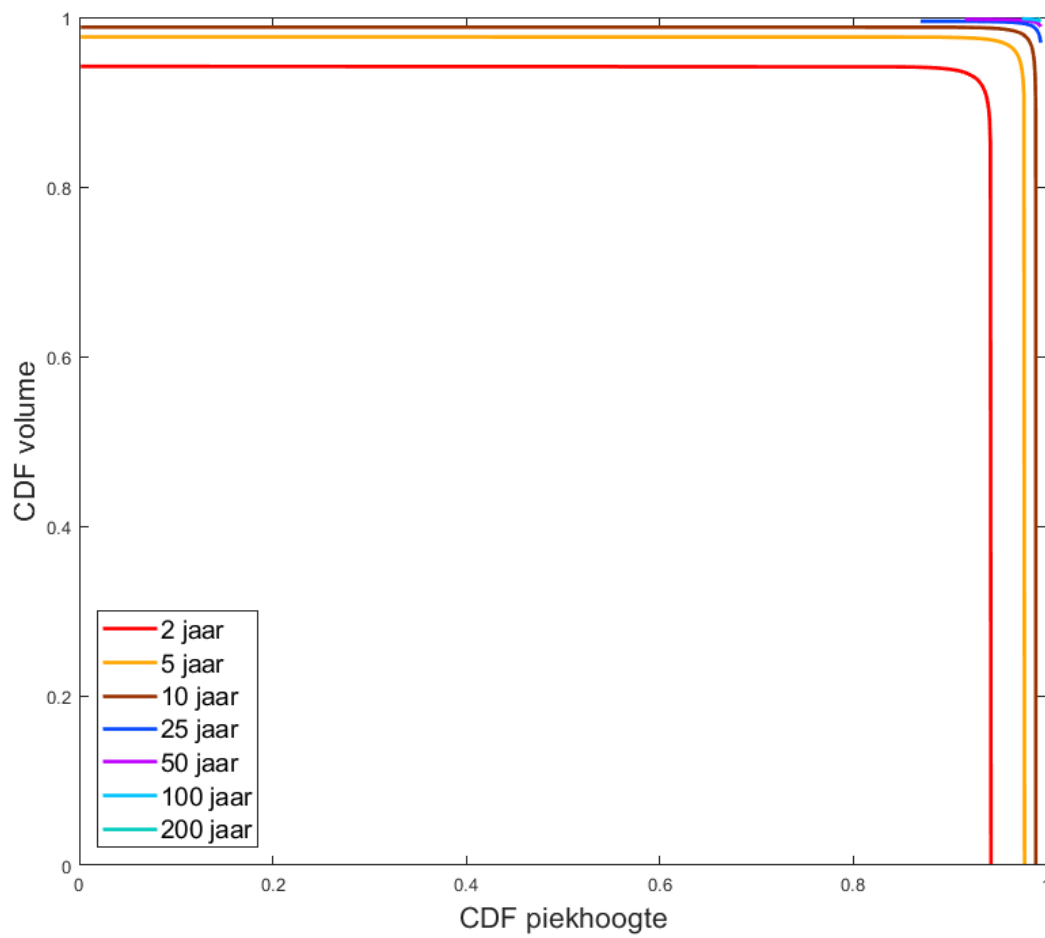
Deze isolijnen kunnen ook samen geplot worden. Het resultaat hiervan wordt voor de retourperiodes van 2 tot 10 jaar weergegeven in Figuur 4.5. Hierbij is telkens ook de CDF-waarde van het univariate geval van de piekhoogte geplot. In deze figuur is opnieuw duidelijk te zien hoe de verschillende retourperiodes zich ten opzichte van elkaar gedragen: de secundaire retourperiode lijkt gemiddeld de laagste waarde aan te nemen, dan de COND, vervolgens de AND; de OR heeft de hoogste waarde, met de univariate retourperiode daar tussenin. Dit is, hoewel hier niet vergeleken wordt voor één paar  $U$ - en  $V$ -waarden, gelijkaardig aan de mathematische vergelijking zoals in Hoofdstuk 2 gegeven. Hier moet wel bij opgemerkt worden dat er door de isolijnen geen uitspraken gedaan kunnen worden over mogelijke ontwerpwaarden. Om interpreteerbare waarden te verkrijgen, moeten er op de isolijnen enkele bewerkingen toegepast worden, zodat hier wel duidelijke data uit te halen vallen.

Om te zorgen dat er een interpreteerbare plot gemaakt kan worden, is er gekozen om per variabele het gemiddelde van de verschillende isolijnen te bepalen. Door hiernaast ook 25%- en 75%-kwantiel te nemen, kan de spreiding op de waarden gelinkt aan de verschillende retourperiodes geplot worden. Het is belangrijk om te begrijp-

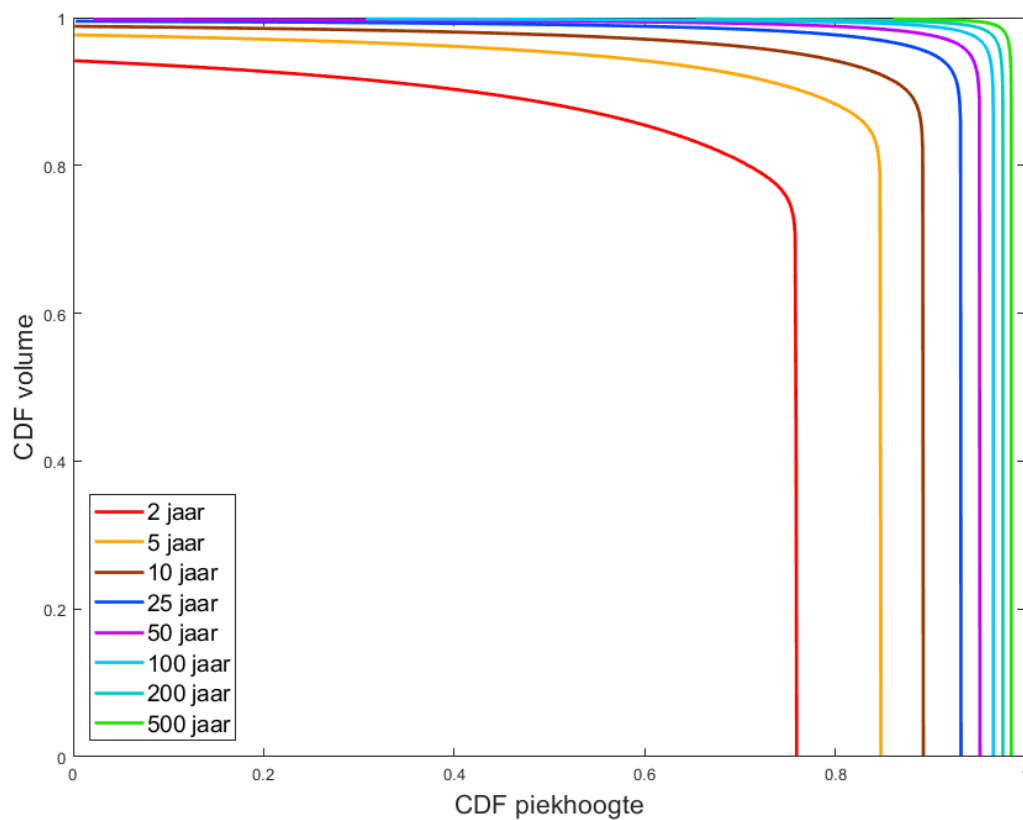




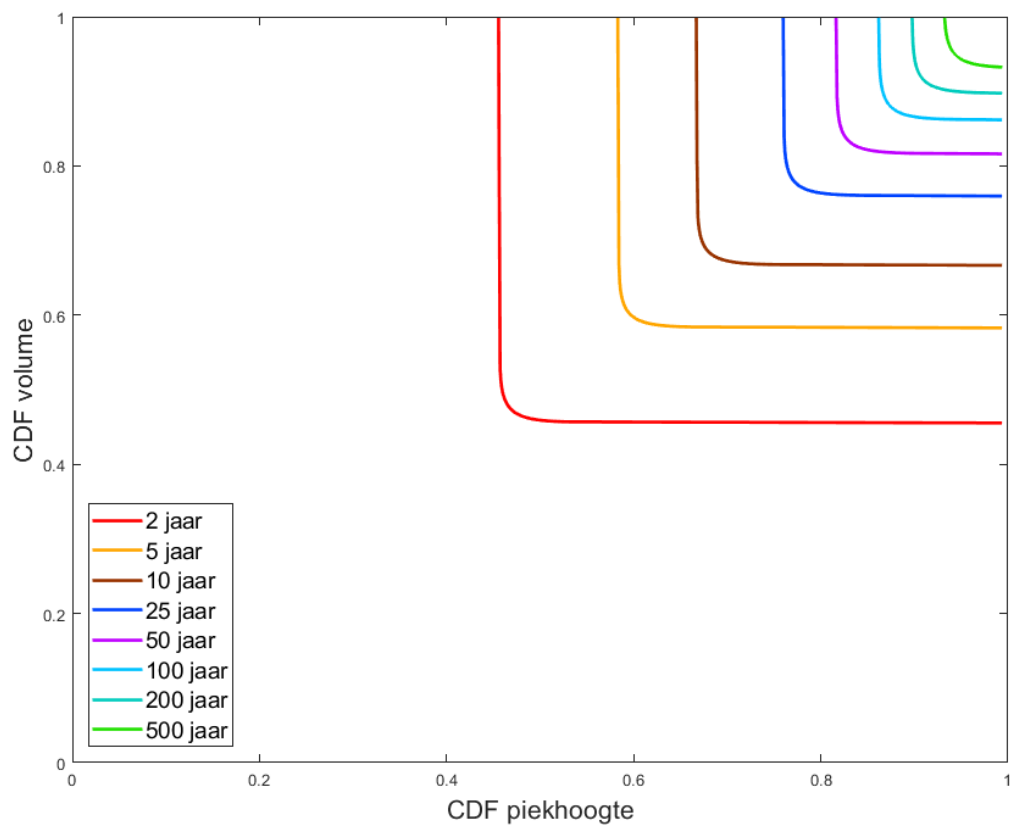
Figuur 4.1: Isolijnen OR-situatie



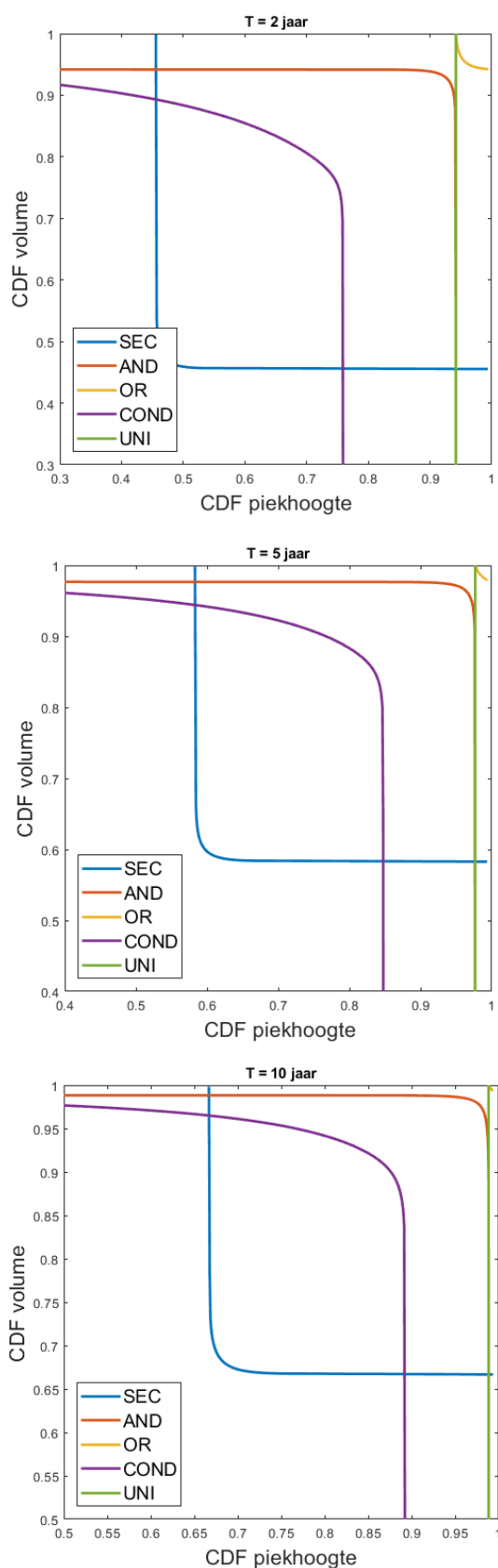
Figuur 4.2: Isolijnen AND-situatie



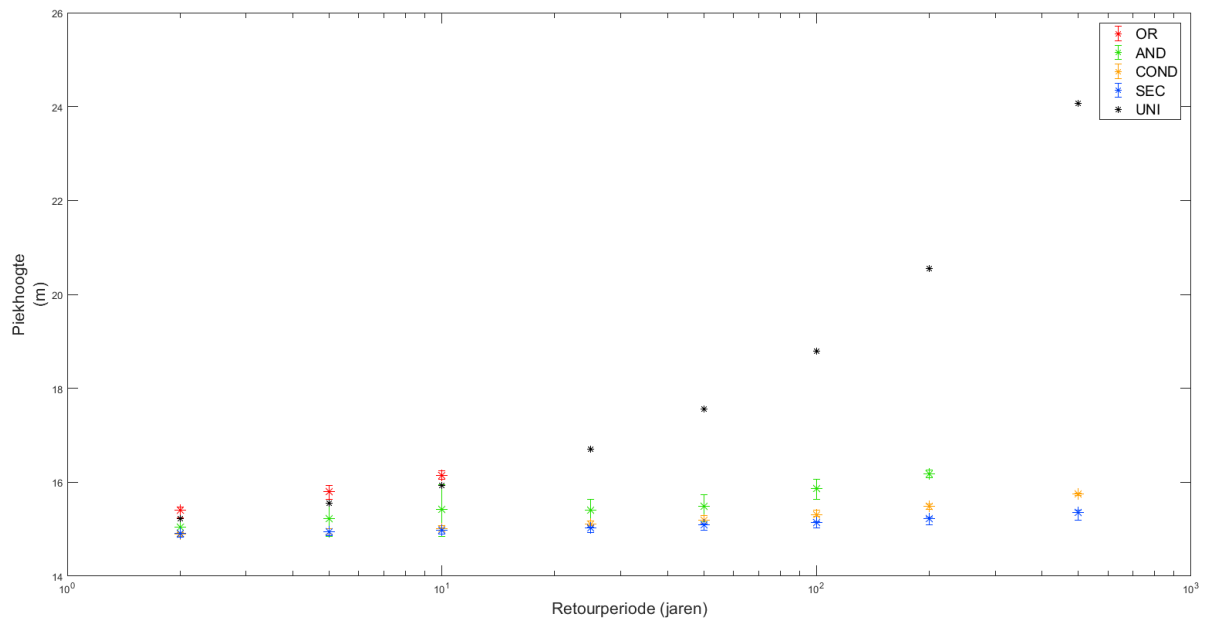
Figuur 4.3: Isolijnen COND-situatie



Figuur 4.4: Isolijnen SEC-situatie



Figuur 4.5: Isolijnen van de verschillende situaties samen geplot, voor de retourperiode van 2 tot 10 jaar.



Figuur 4.6: Gemiddelde reële waarde, 25%- en 75%-kwantiel van de piekhoogte voor elke situatie en enkele retourperiodes

pen wat deze retourperiodes in realiteit betekenen, dus met behulp van de geparametriseerde GEV-verdelingen uit Hoofdstuk 3 zijn de waarden voor de CDF's van de extremen omgerekend naar de waarden zoals ze in de realiteit zouden zijn. Voor elk type situatie zijn deze gemiddelde retourperiodes en hun kwantielen dan samen geplott, zodat de verschillen tussen de retourperiodes duidelijk worden. Het resultaat is voor de piekhoogte te zien in Figuur 4.6 en voor het volume in Figuur 4.7. Hierbij zijn telkens ook de waarden voor de univariate retourperiodes (in de figuren gegeven door 'UNI') van de variabele geplott, zodat het onderscheid daarmee ook gemaakt kan worden. De waarden van de gemiddeldes, de 25%- en 75%-kwantielen zijn samen met de exacte waarden voor de univariate retourperiode gegeven in Tabel 4.1 voor de piekhoogte en in Tabel 4.2 voor het overstromingsvolume.

In deze figuren valt opnieuw duidelijk te zien dat voor eenzelfde retourperiode de waarde van de variabelen sterk kan verschillen. De volgorde van de hoogte van de gemiddelde waarden is altijd als volgt (van hoog naar laag): OR, univariaat, AND en dan COND en SEC, waartussen het onderscheid meestal kleiner is. Dit blijkt ook uit de cijfers, waarbij de gemiddelde piekhoogte voor een retourperiode van 10 jaar voor OR 16.15 m is, voor AND 15.42 m, voor COND 15.02 m en voor SEC 14.98 m. Dit komt door de verschillen tussen de situaties, die in verband met de figuren van de isolijnen ook al besproken zijn. Wat wel nog meer aandacht verdient, is de univariate situatie, die nog niet eerder besproken is. Het is interessant om te zien dat de waarden die voor OR bekomen worden hoger liggen dan deze van de univariate situatie. Een voorbeeld hiervan is bij het overstromingsvolume, voor een retourperiode van 10 jaar, waarbij

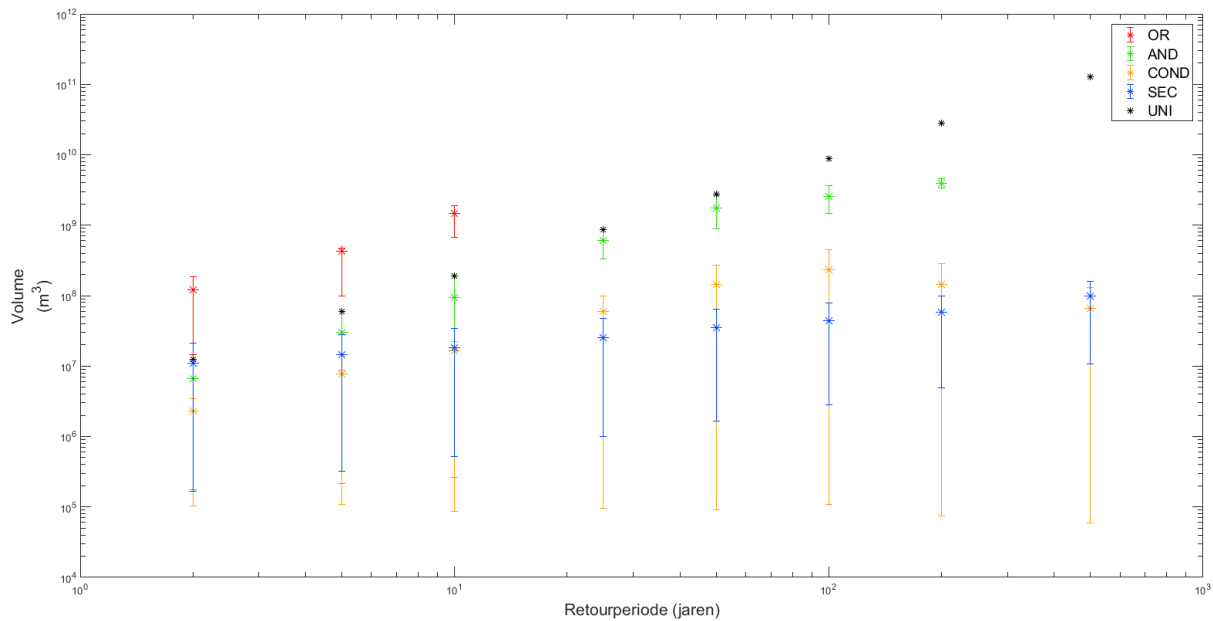
Tabel 4.1: Overzicht van de gemiddelde waardes, 25%-kwantielen, 75%-kwantielen en in het univariate geval de exacte waardes voor alle retourperiodes, voor de piekhoogte (in m). Voor de OR-situatie zijn er voor de retourperiodes vanaf 25 jaar niet genoeg data beschikbaar, voor de AND-situatie vanaf 500 jaar.

| Situatie | Waarde        | Retourperiode |        |         |         |         |          |          |          |
|----------|---------------|---------------|--------|---------|---------|---------|----------|----------|----------|
|          |               | 2 jaar        | 5 jaar | 10 jaar | 25 jaar | 50 jaar | 100 jaar | 200 jaar | 500 jaar |
| OR       | 25%-kwantiel  | 15.25         | 15.64  | 16.06   | n.v.t.  | n.v.t.  | n.v.t.   | n.v.t.   | n.v.t.   |
|          | Gemiddelde    | 15.40         | 15.81  | 16.15   | n.v.t.  | n.v.t.  | n.v.t.   | n.v.t.   | n.v.t.   |
|          | 75%-kwantiel  | 15.46         | 15.92  | 16.24   | n.v.t.  | n.v.t.  | n.v.t.   | n.v.t.   | n.v.t.   |
| AND      | 25%-kwantiel  | 14.83         | 14.84  | 14.85   | 15.08   | 15.15   | 15.64    | 16.10    | n.v.t.   |
|          | Gemiddelde    | 15.04         | 15.22  | 15.42   | 15.40   | 15.49   | 15.87    | 16.18    | n.v.t.   |
|          | 75%-kwantiel  | 15.23         | 15.56  | 15.94   | 15.63   | 15.73   | 16.06    | 16.26    | n.v.t.   |
| COND     | 25%-kwantiel  | 14.85         | 14.89  | 14.97   | 15.07   | 15.21   | 15.35    | 15.54    | 15.77    |
|          | Gemiddelde    | 14.89         | 14.95  | 15.02   | 15.11   | 15.20   | 15.31    | 15.49    | 15.75    |
|          | 75%-kwantiel  | 14.93         | 15.01  | 15.08   | 15.18   | 15.28   | 15.40    | 15.55    | 15.77    |
| SEC      | 25%-kwantiel  | 14.83         | 14.86  | 14.89   | 14.93   | 14.97   | 15.03    | 15.09    | 15.20    |
|          | Gemiddelde    | 14.91         | 14.95  | 14.98   | 15.03   | 15.09   | 15.15    | 15.22    | 15.35    |
|          | 75%-kwantiel  | 14.91         | 14.95  | 14.98   | 15.04   | 15.10   | 15.17    | 15.25    | 15.40    |
| UNI      | Exacte waarde | 15.23         | 15.56  | 15.94   | 16.70   | 17.56   | 18.79    | 20.55    | 24.07    |

Tabel 4.2: Overzicht van de gemiddelde waarden, 25%-kwantilen, 75%-kwantilen en in het univariate geval de exacte waarden voor alle retourperiodes, voor het overstromingsvolume (in m<sup>3</sup>). Voor de OR-situatie zijn er voor de retourperiodes vanaf 25 jaar niet genoeg data beschikbaar, voor de AND-situatie vanaf 500 jaar.

| Situatie | Waarde        | Retourperiode |            |             |              |
|----------|---------------|---------------|------------|-------------|--------------|
|          |               | 2 jaar        | 5 jaar     | 10 jaar     | 25 jaar      |
| OR       | 25%-kwantiel  | 14528796      | 99822429   | 670350698   | n.v.t.       |
|          | Gemiddelde    | 119735897     | 428423289  | 1458362226  | n.v.t.       |
|          | 75%-kwantiel  | 52036130      | 387618347  | 1910302316  | n.v.t.       |
| AND      | 25%-kwantiel  | 175982        | 212828     | 264760      | 327940842    |
|          | Gemiddelde    | 6624539       | 30059158   | 94333582    | 605538628    |
|          | 75%-kwantiel  | 12578911      | 59373972   | 191191888   | 864788515    |
| COND     | 25%-kwantiel  | 104099        | 109683     | 87254       | 94613        |
|          | Gemiddelde    | 2317006       | 7687958    | 16610515    | 59440147     |
|          | 75%-kwantiel  | 3605320762    | 8850333    | 10979499    | 21166486     |
| SEC      | 25%-kwantiel  | 164835        | 317620     | 517923      | 998461       |
|          | Gemiddelde    | 11105573      | 14548703   | 17983377    | 25401181     |
|          | 75%-kwantiel  | 786508        | 1317340    | 1940917     | 3595299      |
| UNI      | Exacte waarde | 12595884      | 59505226   | 190062252   | 875896321    |
|          |               | 50 jaar       | 100 jaar   | 200 jaar    | 500 jaar     |
| OR       | 25%-kwantiel  | n.v.t.        | n.v.t.     | n.v.t.      | n.v.t.       |
|          | Gemiddelde    | n.v.t.        | n.v.t.     | n.v.t.      | n.v.t.       |
|          | 75%-kwantiel  | n.v.t.        | n.v.t.     | n.v.t.      | n.v.t.       |
| AND      | 25%-kwantiel  | 896404164     | 1484975836 | 3393336466  | n.v.t.       |
|          | Gemiddelde    | 1736994669    | 2574729673 | 3956683518  | n.v.t.       |
|          | 75%-kwantiel  | 2657623946    | 3605320762 | 3605320762  | n.v.t.       |
| COND     | 25%-kwantiel  | 90055         | 108545     | 73739       | 59392        |
|          | Gemiddelde    | 143730491     | 234082241  | 145206614   | 65237560     |
|          | 75%-kwantiel  | 19588509      | 18254265   | 2880292     | 1038191      |
| SEC      | 25%-kwantiel  | 1661465       | 2816427    | 4907290     | 10743943     |
|          | Gemiddelde    | 25401181      | 25401181   | 25401181    | 98365989     |
|          | 75%-kwantiel  | 5991450       | 9603037    | 16763037    | 37927647     |
| UNI      | Exacte waarde | 2775412883    | 8785747231 | 27798338531 | 127393836821 |

#### 4.1. OVERZICHT VAN DE BEKOMEN RESULTATEN



Figuur 4.7: Gemiddelde reële waarde, 25%- en 75%-kwantiel van het overstromingsvolume voor elke situatie en enkele retourperiodes

de univariate waarde  $1.90 \cdot 10^8 \text{ m}^3$  bedraagt en de waarde voor OR  $14.58 \cdot 10^8 \text{ m}^3$ . Een verklaring hiervoor ligt opnieuw bij de situatie die zich voordoet: van OR is sprake wanneer één van de beide variabelen overschreden wordt. Hierdoor kan het in meer gevallen voorkomen dan de univariate situatie, die aangeeft hoe vaak een enkele variabele overschreden wordt. Voor dezelfde waarde van de variabele zijn er in de OR-situatie meer mogelijkheden, wat voor een lagere retourperiode zorgt. Of omgekeerd: bij dezelfde retourperiode zal de waarde voor de variabele bij de OR-situatie hoger liggen.

Wat uit deze figuren en tabellen eveneens blijkt, is dat door het gebruik van de onderliggende GEV (waarvan de parameters gegeven zijn in Tabel 3.3) om de waarden om te rekenen deze hoger kunnen liggen dan de waarden die in de data voorkomen. Deze omrekening kan ook gebeuren met behulp van de kwantielen om zo binnen het oorspronkelijke bereik van de data te blijven, maar dit levert een mogelijke onderschatting van het gevaar op. In de oorspronkelijke data was de maximale piekhoogte 15.68 m en het maximale overstromingsvolume  $3.07 \cdot 10^7 \text{ m}^3$ , terwijl de gemiddelde waarde van de piekhoogte voor de OR-situatie bij een retourperiode van 5 jaar al  $15.81 \text{ m}$  bedraagt en deze voor het overstromingsvolume in dezelfde situatie  $4.28 \cdot 10^8 \text{ m}^3$ .

Wat daarnaast ook te zien valt in deze figuren zijn de verschillen in spreiding en in de mate waarin de waardes toenemen per retourperiode. Voor de univariate retourperiode en de bivariate OR- en AND-situaties neemt de waarde per retourperiode veel sterker toe dan in de COND- en SEC-situaties. Zo neemt de gemiddelde waarde van



OR voor de piekhoogte van 5 naar 10 jaar toe van 15.81 m naar 16.15 m (dus met 0.34 m), de gemiddelde waarde voor AND van 15.22 m naar 15.42 m (dus met 0.20 m). Voor COND en SEC is dit in hetzelfde geval respectievelijk van 14.95 m naar 15.02 m (dus met 0.07 m) en van 14.95 m naar 14.98 m (dus met 0.03 m). Vooral de univariate waarden stijgen sterk in vergelijking met de rest, voor zowel piekhoogte als overstromingsvolume. Bij de piekhoogte is het verschil tussen 50 en 100 jaar 1.23 m (van 17.56 m naar 18.79 m). Het valt te verwachten dat dit ook voor de OR-situatie het geval is, maar wegens de beperkte data valt dit niet meer af te leiden voor retourperiodes hoger dan 10 jaar. De minder sterke stijging van COND en SEC (en in zekere mate ook AND) valt grotendeels te verklaren doordat deze, in tegenstelling tot het univariate geval, ook gemodelleerd zijn door middel van de oorspronkelijke data. Wat echter ook een verschil uitmaakt, is dat zeker bij de COND- en SEC-situaties de retourperiodes allemaal gekoppeld zijn aan lagere waarden van  $U$  en  $V$  dan bij de AND-situatie. Bij de omrekening door middel van de GEV-verdeling, zullen hier dus ook lagere waarden uit komen, wat het verschil met de AND- en, nog duidelijker, met de OR-situatie verklaart. Dit effect is wel sterker bij het overstromingsvolume dan bij de piekhoogte, wat verklaart wordt doordat de isolijnen van de AND- en de OR-situatie voor de hogere retourperiodes niet helemaal geplot kunnen worden wegens een gebrek aan data. Vooral bij de hogere waardes voor de CDF van het volume zijn er genoeg data beschikbaar, in tegenstelling tot deze van de piekhoogte. Dit zorgt voor een groter bereik voor de piekhoogte dan op deze van het volume. Hierdoor is er een verschuiving van de gemiddelde waarde van de isolijn naar hogere waarden voor het volume. Bij het overstromingsvolume is in de SEC-situatie de toename zelfs helemaal afwezig: van 25 tot 200 jaar blijft een stijging van de gemiddelde waarde afwezig. Dit kan verklaard worden doordat hoewel het bereik van de isolijn steeds meer naar hogere waarden voor het volume verschuift, er meer datapunten zijn met een lage waarden, waardoor het gemiddelde hetzelfde blijft.

Het belangrijkste gevolg van de variërende toename per situatie is dat de verschillen tussen de situaties steeds groter wordt naargelang de retourperiode hoger wordt. Hoewel bij 2 en 5 jaar de situatie niet veel verschil lijkt uit te maken, is dat bij de hogere retourperiodes al veel meer het geval. Zo is de gemiddelde waarde van de piekhoogte bij 2 jaar voor OR 15.40 m, voor AND 15.04 m, voor COND 14.89 m, voor SEC 14.91 m en voor univariaat 15.23 m, of dus een maximaal verschil van 0.51 m (tussen OR en COND). Bij 10 jaar is dit maximale verschil tussen de gemiddelde waardes al 1.17 m (met de gemiddelde hoogte voor OR 16.15 m, voor AND 15.42 m, voor COND 15.02 m, voor SEC 14.98 m en voor univariaat 15.94 m). Dit effect is veel sterker bij de piekhoogte, waar het oorspronkelijke bereik kleiner was dan bij het overstromingsvolume en de waarden bij lage retourperiodes dus dicht bij elkaar lijken te

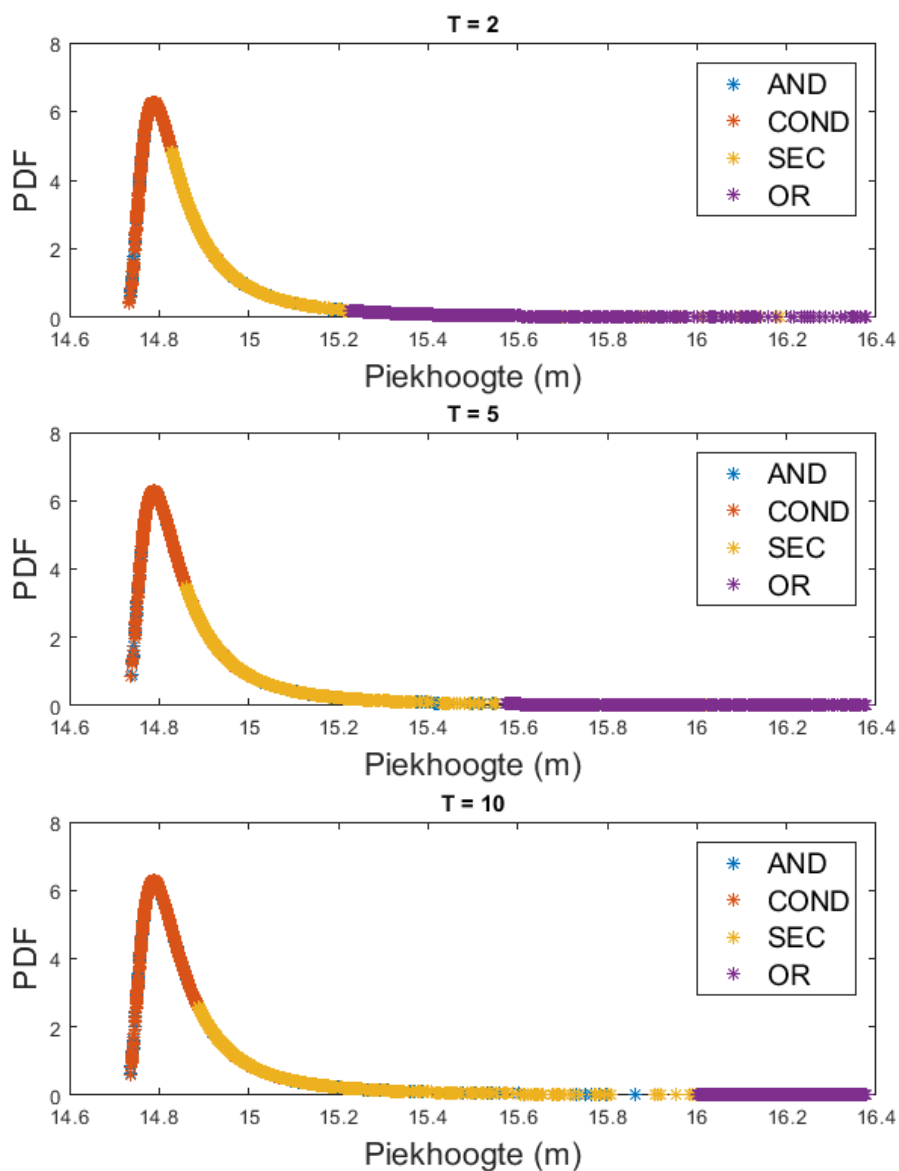
liggen. In vergelijking daarmee verschillen de situaties bij het overstromingsvolume vanaf een retourperiode van 2 jaar al sterk.

De spreiding die op sommige waarden zit, volgt uit de combinatie van de berekening van deze spreiding met het bereik van de originele datareeksen. Doordat per variabele het gemiddelde, het 25%- en 75% kwantiel genomen is van alle waarden van de isolijn van die variabele, is de spreiding afhankelijk van de vorm van de isolijn. Deze vorm hangt sterk af van de situatie: door hun opbouw hebben de COND- en AND-situatie in theorie een groter bereik naarmate de retourperiode toeneemt, terwijl de OR- en SEC-situatie een kleiner bereik hebben naarmate de retourperiode groter wordt. Dit is te zien op Figuren 4.1 tot 4.4. Bij de AND en OR-situatie wordt dit wel gecompenseerd doordat, zoals al eerder vermeld, de isolijnen wegens een gebrek aan data niet volledig geplot kunnen worden. Door deze verschillen in bereik daalt het 25%-kwantiel van de COND-situatie zelfs naarmate de retourperiode stijgt en is het 75%-kwantiel van de COND- en SEC-situaties soms lager dan het gemiddelde. De GEV-verdeling waarmee de  $U$ - en  $V$ -waardes van de isolijn omgerekend worden naar de reële waarden is gebaseerd op de oorspronkelijke data en heeft een bereik dat daardoor beïnvloed wordt, maar wel groter is dan het oorspronkelijke bereik van de data. Zoals in Tabel 3.2 te zien valt, is het bereik op overstromingsvolume relatief gezien veel groter dan dat op de piekhoogte.

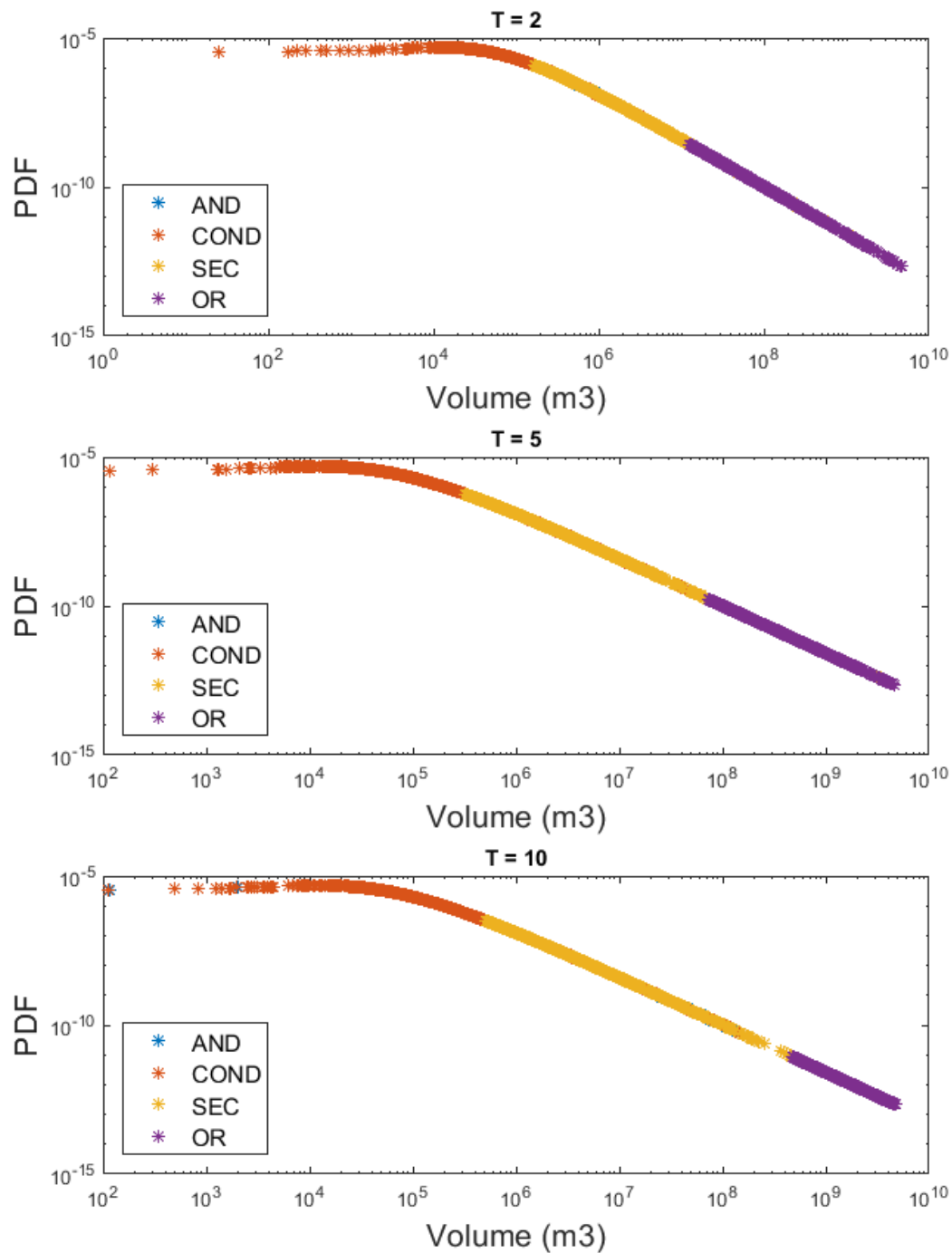
De spreiding van de resultaten zorgt er soms voor dat het onderscheid tussen de resultaten minder duidelijk is dan uit de gemiddeldes op het eerste zicht kan lijken. Bij het overstromingsvolume zorgt de spreiding er bijvoorbeeld voor dat de COND- en SEC-situaties amper van elkaar te onderscheiden zijn. Daarnaast verandert de situatie met de hoogste gemiddelde waarde afhankelijk van de retourperiodes. Tot een retourperiode van 10 jaar heeft SEC de hoogste gemiddelde waarde, daarna is deze van COND groter, behalve bij 500 jaar. Over alle retourperiodes heen is er een sterke overlap tussen deze twee situaties. Zo is zijn de 25%- en 75%-kwantielen bij 100 jaar voor SEC respectievelijk  $2.82 \cdot 10^6 \text{ m}^3$  en  $9.60 \cdot 10^6 \text{ m}^3$ , terwijl deze voor COND  $1.09 \cdot 10^5 \text{ m}^3$  en  $1.83 \cdot 10^7 \text{ m}^3$  zijn. Van  $2.82 \cdot 10^6 \text{ m}^3$  tot  $9.60 \cdot 10^6 \text{ m}^3$  overlappen deze retourperiodes dus, waarbij er dan nog rekening gehouden mee moet worden dat de gemiddeldes bij beide hoger kunnen liggen dan het 75%-kwantiel, door de sterke uitschieters bij de data. Zowel bij het overstromingsvolume als bij de piekhoogte is bij de lagere retourperiodes, wanneer de stijging van het univariate geval beperkt is, ook te zien hoe deze en de AND-situatie, door de spreiding, moeilijk van elkaar te onderscheiden zijn. Bij de piekhoogte, voor 10 jaar, is de univariate waarde 15.94 m, terwijl het 75%-kwantiel van AND ook 15.94 m is. Bij het overstromingsvolume zijn deze, in hetzelfde geval,  $1.90 \cdot 10^8 \text{ m}^3$  en  $1.91 \cdot 10^8 \text{ m}^3$ . Dit impliceert dat ook de gemiddelde waardes geen ideale indicator zijn van het gedrag van de situaties.

Een volgende weergave is op basis van de marginale GEV-verdelingen van de hoogte en het volume. Door voor de retourperiodes van 2 tot 10 jaar uit de isolijnen van de verschillende situaties te samplen, kunnen de PDF's van deze marginale GEV-verdelingen geplot worden. Op deze manier kan vergeleken worden hoe deze samples overeenkomen met elkaar en met de originele verdeling van de data. Het resultaat hiervan staat in Figuren 4.8 en 4.9. Hoewel de samples van de verschillende situaties telkens met dezelfde GEV-parameters omgezet zijn naar de PDF, valt duidelijk op hoe deze nooit het volledige bereik van de verdeling bevatten. De OR-situatie heeft altijd hogere waarden, AND en COND altijd lage waarden en de secundaire retourperiode waarden daartussen. Dit is gelijklopend met de volgorde van de retourperiodes in eerdere figuren. Naarmate de retourperiode vergroot, komen de situaties meer uiteen te liggen, wat gelijkaardig is aan de verschillen in stijging, afhankelijk van de retourperiode.

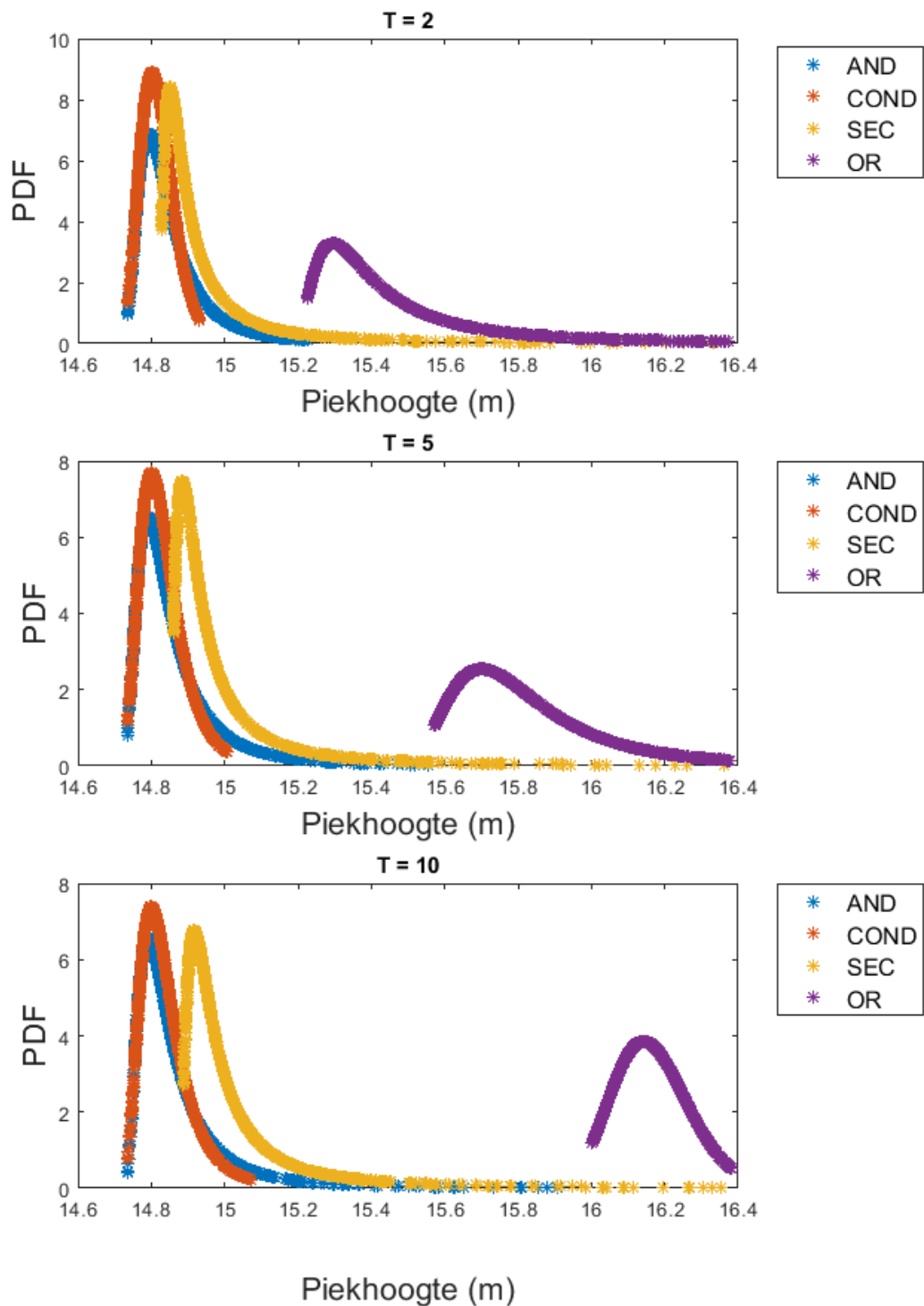
In plaats van de marginale GEV-verdeling te gebruiken, kan de GEV ook aan elke situatie apart gefit worden. Dit is voor de piekhoogte en het overstromingsvolume berekend door te samplen uit de isolijn op de respectievelijke assen. Op deze samples is dan telkens per situatie en per retourperiode een GEV gefit. Doordat deze situaties elk hun eigen bereik en waarden hebben, zal dit GEV's opleveren die verschillend zijn van de oorspronkelijke waarden. Dit levert een extra mogelijkheid op om te vergelijken hoe deze situaties zich gedragen. Dit gedrag blijkt duidelijk verschillend te zijn voor de verschillende situaties (Figuren 4.10 en 4.11, Tabellen 4.3 en 4.4). De AND- en COND-situatie gedragen zich nog enigszins gelijk, met een hoge piek die een lage reële waarde heeft. Voor een retourperiode van 2 jaar bij de piekhoogte zijn deze waardes respectievelijk 14.80 m en 14.81 m en bij het overstromingsvolume beide  $9.2 \cdot 10^4$ . De secundaire retourperiode heeft daarentegen een al sterker verschillende  $\alpha$  (bijvoorbeeld voor de piekhoogte, met retourperiode 5 jaar, bedraagt deze 0.0624 ten opzichte van 0.0479 voor COND) en hogere reële waarde voor de piek (voor hetzelfde geval: 14.91 m ten opzichte van 14.80 m). Dit verschilt bij de OR-situatie al sterk. In dit laatste geval is de reële waarde voor de piek bij een retourperiode van 5 jaar voor de piekhoogte 15.73 m, en de waarde voor  $\alpha$  is 0.1478. Hiermee wordt benadrukt dat de verschillende situaties zich wel degelijk sterk verschillend gedragen. Wat hierbij ook opvalt, is dat de gefitte waarde voor  $k$  in het geval van de piekhoogte sterker afwijkt dan in het geval van de originele datareeks, terwijl deze van het overstromingsvolume 0 is. Dit wijst op een groot verschil ten opzichte van de originele GEV-verdelingen.



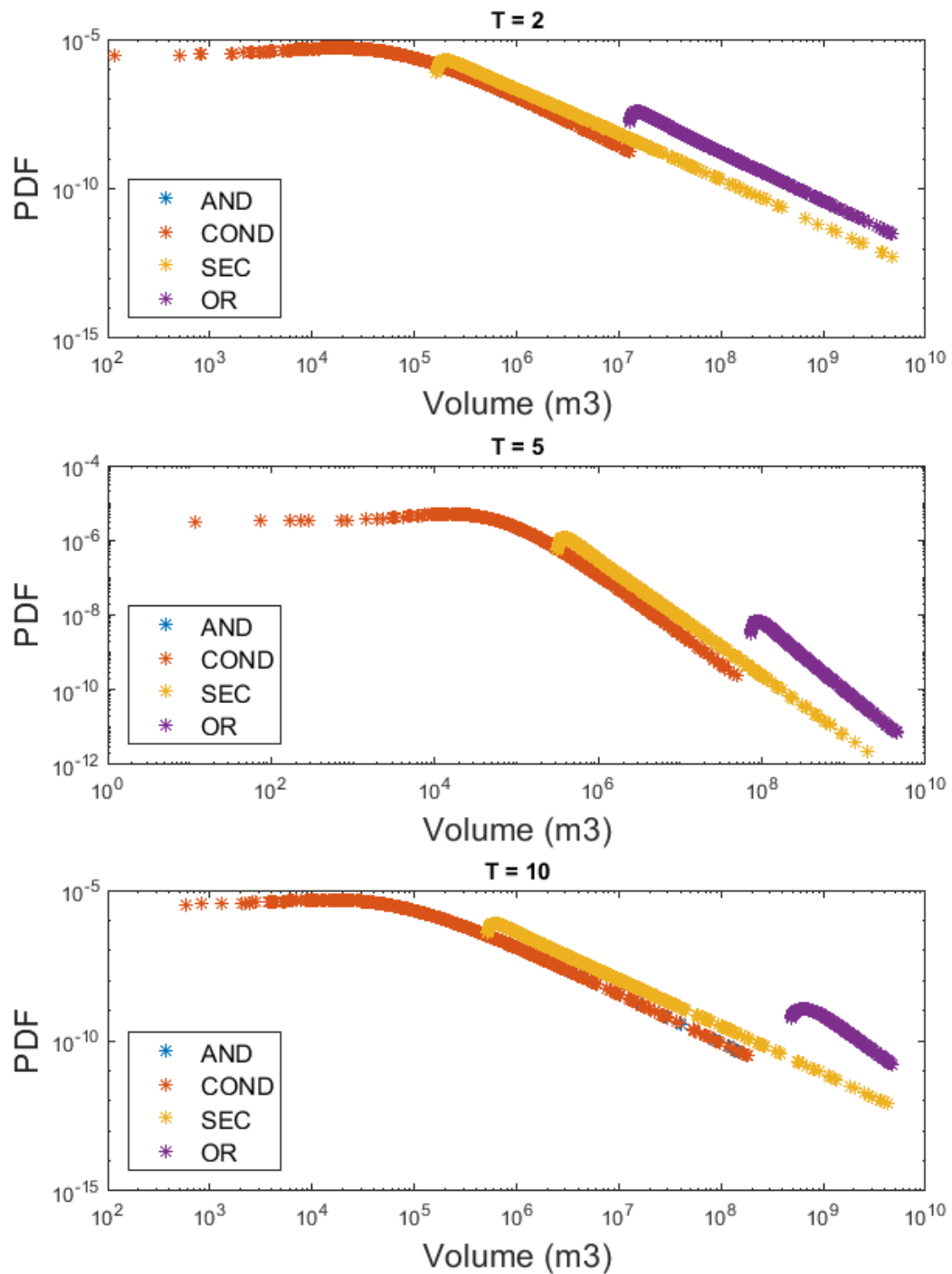
Figuur 4.8: PDF van de marginale GEV-verdeling van de hoogte. Deze PDF is gemaakt met waarden die gesampled zijn uit de isolijn en omgerekend zijn volgens de GEV-verdeling van de hoogte. Dit is berekend voor  $T = 2, 5$  en  $10$  jaar.



Figuur 4.9: PDF van de marginale GEV-verdeling van het volume. Deze PDF is gemaakt met waarden die gesampled zijn uit de isolijn en omgerekend zijn volgens de GEV-verdeling van het volume. Dit is berekend voor  $T = 2, 5$  en  $10$  jaar.



Figuur 4.10: GEV-verdeling gefit op de piekhoogte voor elke situatie, voor  $T= 2, 5$  en 10 jaar.



Figuur 4.11: GEV-verdeling gefit op het overstromingsvolume voor elke situatie, voor  $T = 2, 5$  en  $10$ .

Tabel 4.3: Overzicht van de verschillende parameters voor de gefitte GEV-verdelingen in Figuur 4.10, per situatie en retourperiode

| Situatie |      | $k$     | $\alpha$ | $u$   |
|----------|------|---------|----------|-------|
| 2 jaar   | AND  | 0.3075  | 0.0558   | 14.81 |
|          | COND | -0.0813 | 0.0414   | 14.80 |
|          | SEC  | 0.7941  | 0.0565   | 14.88 |
|          | OR   | 0.5646  | 0.1281   | 15.34 |
| 5 jaar   | AND  | 0.3957  | 0.0604   | 14.81 |
|          | COND | 0.0605  | 0.0479   | 14.80 |
|          | SEC  | 0.7640  | 0.0624   | 14.91 |
|          | OR   | 0.2104  | 0.1478   | 15.73 |
| 10 jaar  | AND  | 0.4313  | 0.0607   | 14.81 |
|          | COND | 0.1524  | 0.0504   | 14.81 |
|          | SEC  | 0.7375  | 0.0679   | 14.95 |
|          | OR   | -0.1838 | 0.0972   | 16.12 |

Tabel 4.4: Overzicht van de verschillende parameters voor de gefitte GEV-verdelingen in Figuur 4.11, per situatie en retourperiode

| Situatie |      | $k$    | $\alpha$          | $u$               |
|----------|------|--------|-------------------|-------------------|
| 2 jaar   | AND  | 0.0000 | $1.47 \cdot 10^5$ | $9.2 \cdot 10^4$  |
|          | COND | 0.0000 | $1.47 \cdot 10^5$ | $9.2 \cdot 10^4$  |
|          | SEC  | 0.0000 | $5.31 \cdot 10^5$ | $4.51 \cdot 10^5$ |
|          | OR   | 0.0000 | $2.16 \cdot 10^7$ | $2.56 \cdot 10^7$ |
| 5 jaar   | AND  | 0.0000 | $1.6 \cdot 10^5$  | $1.0 \cdot 10^5$  |
|          | COND | 0.0000 | $1.6 \cdot 10^5$  | $1.0 \cdot 10^5$  |
|          | SEC  | 0.0000 | $8.0 \cdot 10^5$  | $7.6 \cdot 10^5$  |
|          | OR   | 0.0000 | $9.51 \cdot 10^7$ | $1.37 \cdot 10^8$ |
| 10 jaar  | AND  | 0.0000 | $1.8 \cdot 10^5$  | $1.0 \cdot 10^5$  |
|          | COND | 0.0000 | $1.8 \cdot 10^5$  | $1.0 \cdot 10^5$  |
|          | SEC  | 0.0000 | $1.23 \cdot 10^6$ | $1.20 \cdot 10^6$ |
|          | OR   | 0.0000 | $3.97 \cdot 10^8$ | $8.19 \cdot 10^8$ |



## 4.2 Discussie

Een belangrijke vraag die uit de resultaten volgt, is hoe hiermee omgegaan moet worden. Om dit te bestuderen, is het interessant om deze resultaten eens te vergelijken met gangbare, eenvoudige hydraulische structuren: een keermuur en een overstromingsreservoir. Stel dat we deze structuren voor een levensduur van zeker 50 jaar willen ontwerpen, en rekening willen houden met overstromingen met een retourperiode van 25 jaar (aangezien hiervoor data van de meeste situaties beschikbaar zijn). Via Vergelijking (2.6) kan dan berekend worden dat de kans op falen gelijk is aan  $1 - \left(1 - \frac{1}{25}\right)^{50} = 0.8701$ , wat een onaanvaardbaar hoge kans is. Er moet dus gewerkt worden met retourperiodes die hoger liggen. Uit de resultaten blijkt echter dat hoe hoger de retourperiode, hoe verder de gemiddelde waardes voor de verschillende situaties uit elkaar liggen. Dit betekent dat het nodig is om te bekijken welke situaties relevant zijn voor de hier gegeven toepassingen. Een OR-situatie betekent in dit geval dat ofwel het overstromingsvolume ofwel de piekhoogte hoog genoeg moeten zijn om een overstroming te veroorzaken. Bij een keermuur is schijnbaar enkel de piekhoogte belangrijk. Bij het overstromingsreservoir zijn de hoogte en het volume gerelateerd, al kan een OR-situatie wel gebruikt worden als het water door een inlaat het reservoir in kan stromen, maar het water ook over een dijk in het reservoir kan stromen. Op die manier kan het reservoir falen als het volume dat het moet bevatten te groot is, of als de waterhoogte in de rivier te hoog wordt en op die manier een te grote toevloed (zowel door de inlaat als over een dijk) in het reservoir veroorzaakt, waardoor deze toevloed niet meer gecontroleerd kan worden. Een andere optie is dat het overstromingsreservoir en de keermuur zich op dezelfde locatie bevinden en een falen van een van beide een overstroming kan veroorzaken, waarbij de twee overstromingen verschillend zijn.

Een AND-situatie lijkt voor beide toepassingen minder relevant: de keermuur wordt vooral bepaald door de piekhoogte, het overstromingsreservoir door het volume. Het lijkt voor beide toepassingen onwaarschijnlijk dat beide variabelen hoog genoeg moeten zijn om een overstroming te veroorzaken. Een situatie waarin dat zou kunnen voorkomen, is wanneer het overstromingsreservoir pas vanaf bepaalde hoogtes gevuld wordt: een combinatie van een te groot overstromingsvolume (dat door het reservoir niet meer geborgen kan worden) en een grote piekhoogte zal dan een overstroming induceren. Dit kan voorkomen wanneer dit volume op een korte tijd bereikt wordt, in tegenstelling tot een situatie waarin de hoogte gedurende lange tijd net over de drempelwaarde gaat: in de berekeningen zal dit gezien worden als een overstromingssituatie, maar het kan zijn dat dit volume het overstromingsreservoir nooit zal vullen.

Wat echter logischer is voor deze toepassing, is de conditionele situatie, aangezien het idee van een drempelwaarde hier standaard in zit. Deze situatie niet van toepassing op een keermuur. Wat in beide gevallen wel gebruikt kan worden, is de secundaire retourperiode, aangezien die de mogelijkheid geeft dat extremen groter dan een bepaalde grens voorkomen. Dat kan echter wel te algemeen zijn, aangezien sommige combinaties van de variabelen pas een probleem vormen als ze heel specifiek voorkomen.

De moeilijkheid van deze keuze vult in combinatie met de resultaten (zie bijvoorbeeld de sterk verschillende GEV-verdelingen) aan wat Serinaldi (2015) al aangebracht had: er zitten te sterke verschillen op de retourperiodes om ze zomaar op een hoopje te gooien. Want hoewel de verschillende situaties mathematisch vergeleken kunnen worden, zoals o.a. Yue and Rasmussen (2002) en Vandenberghe et al. (2011) gedaan hebben, heeft dit geen aanvullende betekenis: de formules stellen intrinsiek verschillende situaties voor. De figuren, cijfers en mogelijkheden tot toepassing die besproken zijn, verduidelijken dit alleen maar.

Het moet dan bekeken worden of het interessant blijft om deze retourperiodes te gebruiken. Het valt te beargumenteren van wel: doordat elk van de situaties een verschillende combinatie van extremen voorstelt, loont het om voor de juiste omstandigheden de juiste situatie te kiezen. Achter het falen van een hydraulische structuur gaat immers een zekere combinatie van variabelen schuil, die correct ingeschat dient te worden. De opdracht waar een studie bureau bij een ontwerp dus op zou moeten focussen, is het correct inschatten van de omstandigheden die zich voordoen en de daarmee overeenkomstige situatie gebruiken. Hiervoor is wel veel informatie nodig, meer dan in het univariate geval.

Vaak is deze informatie niet of minder beschikbaar, of dient deze via een omweg verkregen te worden, zoals in deze studie een piekhoogte en overstromingsvolume verkregen zijn door middel van een model. De vraag kan gesteld worden of het dan niet eenvoudiger en goedkoper is om voor een univariate inschatting te kiezen, waarbij de verschillende variabelen die een rol spelen apart bestudeerd worden: uit de voorgaande cijfers is immers gebleken dat een univariate waarde vaak een van de grootste waarden levert. Deze insteek gaat echter voorbij aan drie belangrijke aspecten. Het eerste belangrijke aspect is dat de bivariate retourperiodes een isolijn beslaan, waartoe verschillende waarden behoren. Door hier op te ontwerpen, wordt de informatie van de volledige isolijn meegenomen, wat echter ook betekent dat er een zekere spreiding op zit. Zoals de cijfers uit Tabellen 4.1 en 4.2 aantonen, kan deze spreiding ervoor zorgen dat het gevaar dat gepaard gaat bij waarden van ontwerpvariabelen bepaald via een bivariate retourperiode groter is dan wat verkregen wordt indien met de univariate verdeling wordt gewerkt.

Naast het mogelijk grotere gevaar ten gevolge van de isolijn van de bivariate retourperiodes is ook gebleken dat de exacte waarde van het univariate geval niet altijd het hoogste is, wanneer deze vergeleken wordt met de gemiddeldes van de isolijnen. De OR-situatie levert in principe altijd een hogere waarde, dus als blijkt dat meerdere variabelen een rol spelen bij het proces dat onderzocht wordt en apart van elkaar voor gevaar kunnen zorgen, is het sowieso nodig met deze OR-situatie rekening te houden.

Een laatste aspect is dat kan blijken dat de omstandigheden die aanleiding geven tot een falen, ook veel lager kunnen uitvallen, wanneer er bijvoorbeeld sprake is van een SEC- of COND-situatie. In dat geval kan het op deze manier naar de combinatie van variabelen kijken er voor zorgen dat er kosten kunnen bespaard wordt, omdat er niet voor beide variabelen van het univariaat grootste gevaar uitgegaan moet worden.

Het lijkt dus het meest logisch om, wanneer meerdere variabelen samen een invloed hebben, deze ook samen te beschouwen in het ontwerp. Zo wordt de situatie juist aangepakt, het gevaar juist ingeschat en bespaard of noodzakelijke kosten gedaan waar nodig. Elke andere aanpak zal leiden tot fouten, die uiteindelijk tot materiële schade of zelfs het verlies van mensenlevens kunnen leiden.

Dan nog is er de vraag hoe er dan exact ontworpen moet worden. Door de isolijnen kampen de bivariate retourperiodes met het fundamentele probleem dat verschillende waarden van de variabelen aanleiding kunnen geven tot eenzelfde retourperiode. Wanneer op basis van een benodigde levensduur, kostprijs of retourperiode een bepaalde combinatie aan ontwerpvariabelen gekozen wordt, dan betekent dat dit, in theorie, een oneindig aantal andere combinaties kan voorkomen. Een optie is de keuze per variabele dan te beperken tot het bereik dat in de oorspronkelijke datareeks voorkomt, maar dit gaat in tegen het basisidee van extremewaardenanalyse. Dat idee is namelijk dat er een onderliggende verdeling is die de extremen bepaalt, en die er voor kan zorgen dat er nog hogere extremen kunnen voorkomen, die op deze manier gevonden kunnen worden. Dit wordt ook benadrukt door de data, waaruit blijkt dat de meeste waarden hoger zijn dan wat er voorkomt in de data. Gräler et al. (2013) geven wel ideeën hoe er ontworpen kan worden, maar benadrukken eveneens dat er nog veel onderzoek nodig is. Een relevante vraag daarbij is de rol van de copulakeuze. De copula kan immers, afhankelijk van de correlatie tussen twee variabelen, de isolijnen die met de probabiliteiten samengaan sterk bepalen. Het belang van een goede fit en een ruime keuze aan copulafamilies om deze fit op toe te passen is dus groot. Ideeën zoals bij Sadegh et al. (2017) kunnen helpen om deze keuzes goed te maken.

Dan is er nog de vraag of, zoals Serinaldi (2015) ook aangeeft, het concept van retourperiode niet als geheel achterwege gelaten kan worden. Het lijkt immers logischer om

Tabel 4.5: De bivariate kans op falen voor een paar  $(U, V)$  dat aanzet geeft tot  $t = 0.90$  en  $M = 50$  jaar, voor de OR-, AND- en SEC-situaties.

| Situatie | $p_{50}$ |
|----------|----------|
| OR       | 0.9948   |
| AND      | 0.9948   |
| SEC      | 0.0316   |

bij de berekening te vertrekken vanuit de kans op falen  $p_M$  (zoals gegeven in Vergelijking (2.37)), in plaats van deze voor een bepaalde retourperiode te berekenen. Hier kunnen de verschillende situaties ook in verwerkt worden, net zoals in het gebruik van de retourperiodes, maar kan er tegelijkertijd eenduidiger afgebakend worden wat dit betekent voor het ontwerp. Aangezien deze nog steeds gebaseerd is op de copula, is hiermee het probleem van de grote variabiliteit van bivariate retourperiodes nog niet opgelost, al kan een duidelijke manier om het gevaar in te schatten het onderzoek naar een goede keuze van ontwerpvariabelen wel vooruit helpen. Dit is bijvoorbeeld door Salvadori et al. (2016) uitgewerkt, die de kans op falen gebruikt hebben om te tonen hoe deze in het ontwerp gebruikt kan worden. Hoewel ook zij nog steeds veel onzekerheid in hun resultaten hebben, is de basisgedachte die hierin zit duidelijker en correcter. Dat laat toe om de juiste situaties effectief toe passen, wat zoals eerder beschreven noodzakelijk is. In Tabel 2.4 staan enkele formules gegeven, waarmee de kans op falen berekend kan worden. In Tabel 4.5 staan de waarden wanneer ter illustratie de level curve van de hier gefitte copula met als gezamenlijke probabiliteit 90% en als levensduur 50 jaar gebruikt wordt in deze berekeningen.

Uit de berekende waarden in Tabel 4.5 blijkt enerzijds dat deze werkwijze duidelijkheid verschaft, anderzijds dat de keuze voor de situaties een grote impact kan hebben. Het verschil van de faalkans tussen de OR- en AND-situaties enerzijds en de SEC-situatie anderzijds is zeer groot. Dit viel al te verwachten op basis van Tabel 2.4 en Figuren 4.1 tot 4.4. De formules waarmee de kans op falen berekend wordt zijn gebaseerd op andere concepten: de copula tegenover de Kendall's Measure van een level curve van deze copula. Deze verschillende berekeningswijze komt ook naar voor in de figuur, waar voor eenzelfde combinatie van variabelen  $U$  en  $V$  een veel hogere retourperiode verkregen wordt voor SEC dan voor AND en OR. Aangezien de kans op falen in Tabel 4.5 telkens gebaseerd is op dezelfde combinatie van variabelen, komt dit effect hier ook heel duidelijk naar voor, met een kans op falen van 0.0316 voor de SEC-situatie ten opzichte van een kans op falen van 0.9948 voor de OR- en AND-situaties.

De juiste keuze van situatie is dus hoe dan ook een belangrijke opgave. In hoofdstuk 2 is aangegeven dat hier al onderzoek rond gedaan is, dit is echter nog steeds beperkt. Ook in deze studie zijn maar twee variabelen bestudeerd. Wat bijvoorbeeld nog interessant is, is de impact van de overstromingsduur en welke situatie het meest in-

teressant is wanneer deze duur in combinatie met bijvoorbeeld overstromingsvolume gebruikt wordt.



## HOOFDSTUK 5

# **CONCLUSIE**

Wegens de stijgende nood aan overstromingsmitigatie is het essentieel om de extremen te kennen die aanleiding geven tot overstromingen en hoe deze gedefinieerd kunnen worden. Dit kan door middel van retourperiodes die de grootte van een extreme waarde koppelen aan de gemiddelde tijd tussen twee gebeurtenissen met deze extreme waarde. Dit is een techniek die goed werkt in het univariate geval, maar in het bivariate geval volstaan de standaard probabilistische technieken niet om deze te berekenen. Om hier aan tegemoet te komen worden copula's gebruikt. Deze kunnen de gezamenlijke probabiliteit beschrijven zonder de noodzaak de marginale verdeling exact te kennen. Het probleem bij deze bivariate toepassing is dat er verschillende definities voor de retourperiode zijn. In dit werk zijn daarvan de OR-, AND-, COND- en SEC-situaties onderzocht en met elkaar vergeleken met behulp van data gebaseerd op waarnemingen op het KMI te Ukkel.

Door middel van het Probability Distributed Model werd een debietreeks gemaakt, waarmee op basis van enkele drempelwaarden overstromingsvariabelen berekend zijn. Op deze gegevens is de Frank copula toegepast, waarmee de verschillende types van retourperiodes berekend konden worden.

Uit het vergelijkend onderzoek tussen de situaties die gebruikt zijn voor het definiëren van de retourperiodes bleek dat het selecteren van de correcte waarden voor extreme gebeurtenissen een moeilijke opgave is. Uit dit onderzoek blijkt namelijk dat er verschillende problemen gepaard gaan met het gebruik van retourperiodes: ze zorgen al te vaak voor verwarring en worden in bivariate en multivariate gevallen te vaak door elkaar gebruikt. De kans op falen  $p_M$  voor een levensduur  $M$  van een hydraulische structuur lijkt daar een goed alternatief voor. De bivariate situaties, met onder andere de hier onderzochte OR, AND, COND en SEC, zijn echter wel noodzakelijk zijn om de extremen goed in te schatten. Het is dan wel belangrijk om ze juist toe te passen: afhankelijk van de toepassing verschillen de manieren waarop de variabelen met elkaar kunnen interageren en dus de situatie die geselecteerd dient te worden. Om deze stap goed uit te voeren, is nog veel onderzoek nodig: er is nog maar voor enkele toepassingen concreet omschreven welke situatie het meest relevant is. Dit kan op verschillende schaalniveau's gebeuren: van de toepassing als geheel (types

---

overstromingen) tot de variabelen die daarbij bepaald kunnen worden. Zo zijn bij dit onderzoek enkel overstromingsvolume en piekhoogte onderzocht, terwijl piekdebiet en overstromingsduur even relevante variabelen kunnen zijn voor een overstroming. Een vraag die daar uit volgt is of de concepten die nu bestaan voor bivariate toepassingen makkelijk uitgebreid kunnen worden naar multivariate toepassingen.

Om concreet naar ontwerp over te gaan, is verder onderzoek eveneens nodig: er zijn nog maar enkele ideeën gegeven hoe er met de grote variabiliteit aan ontwerpvariabelen omgegaan moet worden. Een ander aspect dat meer aandacht verdient, is de rol van de copula's hierin. In dit onderzoek bleek het niet nodig om de fits van verschillende copula's te vergelijken, maar in andere omstandigheden kan een goede keuze van copula wel een groot verschil uitmaken, aangezien deze de vorm van de isolijnen bepalen. Hiervoor moet er tussen genoeg copulafamilies gekozen worden, en dienen de technieken om de fit te bepalen robuust te zijn, wat eveneens verder onderzoek noodzaakt.

Door aan al deze stappen te voldoen, kunnen bi- en multivariate retourperiodes een zeer waardevol hulpmiddel zijn om extremen te bepalen. Gezien de Europese regelgeving en de steeds duidelijker aanwezig klimaatverandering is dit zelfs een noodzakelijk hulpmiddel: de inschatting van extremen zal alleen maar beter moeten worden.



# **BIBLIOGRAFIE**

- Bender, J., Wahl, T., Mudersbach, C., and Jensen, J. (2013). Flood frequency analysis and river confluences–univariate vs. multivariate extreme value statistics. *ICWRER 2013 proceedings*, 1056:216–328.
- Breymann, W., Dias, A., and Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3(1):1–14.
- Cabus, P. (2008). River flow prediction through rainfall–runoff modelling with a probability-distributed model (PDM) in Flanders, Belgium. *Agricultural Water Management*, 95(7):859–868.
- Chow, V. T., Maidment, D. R., and Mays, L. W. (1988). *Applied hydrology*. Editions McGraw-Hill, New York.
- Corbella, S. and Stretch, D. D. (2012). Multivariate return periods of sea storms for coastal erosion risk assessment. *Natural Hazards and Earth System Sciences*, 12(8):2699–2708.
- De Jongh, I. L. M., Verhoest, N. E. C., and De Troch, F. P. (2006). Analysis of a 105-year time series of precipitation observed at Uccle, Belgium. *International Journal of Climatology*, 26(14):2023–2039.
- De Michele, C. and Salvadori, G. (2003). A generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. *Journal of Geophysical Research: Atmospheres*, 108(D2):4067.
- De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., and Rosso, R. (2005). Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering*, 10(1):50–57.
- Deheuvels, P. (1979). La fonction de dépendence empirique et ses propriétés. Un test non paramétrique d'indépendence. *Bulletin de la classe des sciences, Académie Royale de Belgique, 5e série*, 65:274–292.
- Demarée, G. R. (2003). The centennial recording raingauge of the Uccle Plateau: Its history, its data and its applications. *Houille Blanche*, 4:95–102.

- Dodangeh, E., Shahedi, K., Shiau, J.-T., and MirAkbari, M. (2017). Spatial hydrological drought characteristics in Karkheh River basin, southwest Iran using copulas. *Journal of Earth System Science*, 126(6):80.
- Dung, N. V., Merz, B., Bardossy, A., and Apel, H. (2015). Handling uncertainty in bivariate quantile estimation—An application to flood hazard analysis in the Mekong Delta. *Journal of Hydrology*, 527:704–717.
- Durante, F. and Salvadori, G. (2010). On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2):143–161.
- Eberhart, R. and Kennedy, J. (1995). A new optimizer using Particle Swarm Theory. In *Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995. MHS'95.*, pages 39–43. IEEE.
- European Union (2007). Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. *Official Journal of the European Union*, L288:27–34.
- Fang, H.-B., Fang, K.-T., and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82(1):1–16.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N., and Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water resources research*, 40(1):W01101.
- Fermanian, J.-D. (2005). Goodness-of-fit tests for copulas. *Journal of multivariate analysis*, 95(1):119–152.
- Genest, C. and Favre, A.-C. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, 12(4):347–368.
- Genest, C., Favre, A.-C., Béliveau, J., and Jacques, C. (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research*, 43(9):W09401.
- Genest, C., Ghoudi, K., and Rivest, L.-P. (1998). “Understanding Relationships Using Copulas”, by Edward Frees and Emiliano valdez, January 1998. *North American Actuarial Journal*, 2(3):143–149.
- Genest, C., Quessy, J.-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2):337–366.

- Genest, C. and Rémillard, B. (2005). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Technical report, Technical Report G-2005-51, GERAD, Montreal, Canada.
- Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American statistical Association*, 88(423):1034–1043.
- Gräler, B., van den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., Baets, B. D., and Verhoest, N. E. C. (2013). Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, 17(4):1281–1296.
- Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, 3(9):816.
- IPCC (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2):81–93.
- Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814–861.
- Liebscher, E. (2008). Construction of asymmetric multivariate copulas. *Journal of Multivariate analysis*, 99(10):2234–2250.
- Luke, Y. L. (1969). *Special functions and their approximations*, volume 2. Academic press.
- Merz, B., Aerts, J. C. J. H., Arnbjerg-Nielsen, K., Baldi, M., Becker, A., Bichet, A., Blöschl, G., Bouwer, L. M., Brauer, A., Cioffi, F., et al. (2014). Floods and climate: emerging perspectives for flood risk assessment and management. *Natural Hazards and Earth System Sciences*, 14(7):1921.
- Moore, R. J. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences*, 11(1):483–499.
- Nelsen, R. B. (2003). Properties and applications of copulas: A brief survey. In Dhaene, J., Kolev, N., and Morettn, P. A., editors, *Proceedings of the First Brazilian Conference on Statistical Modeling in Insurance and Finance*, pages 10–28. University Press USP, Sao Paulo.
- Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd. New York: Springer Science Business Media.

- Ntegeka, V. and Willems, P. (2008). Trends and multidecadal oscillations in rainfall extremes, based on a more than 100-year time series of 10 min rainfall intensities at Uccle, Belgium. *Water Resources Research*, 44(7):W07402.
- Panchenko, V. (2005). Goodness-of-fit test for copulas. *Physica A: Statistical Mechanics and its Applications*, 355(1):176–182.
- Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proc. R. Soc. Lond. A*, 193(1032):120–145.
- Pham, M. T. (2016). *Copulas-based stochastic modelling of evapotranspiration time series conditioned on rainfall as design tool in water resources management*. PhD thesis, Department of Forest and Water management, Ghent University.
- Qian, L., Wang, H., Dang, S., Wang, C., Jiao, Z., and Zhao, Y. (2018). Modelling bivariate extreme precipitation distribution for data-scarce regions using Gumbel–Hougaard copula with maximum entropy estimation. *Hydrological Processes*, 32(2):212–227.
- Renard, Benjamin & Lang, M. (2007). Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology. *Advances in Water Resources*, 30(4):897–912.
- Restrepo-Posada, P. J. and Eagleson, P. S. (1982). Identification of independent rainstorms. *Journal of Hydrology*, 55(1-4):303–319.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.
- Sadegh, M., Ragno, E., and AghaKouchak, A. (2017). Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework. *Water Resources Research*, 53(6):5166–5183.
- Salvadori, G. and De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40(12):W12511.
- Salvadori, G. and De Michele, C. (2007). On the use of copulas in hydrology: theory and practice. *Journal of Hydrologic Engineering*, 12(4):369–380.
- Salvadori, G. and De Michele, C. (2010). Multivariate multiparameter extreme value models and return periods: A copula approach. *Water Resources Research*, 46(10):W10501.
- Salvadori, G. and De Michele, C. (2013). Multivariate extreme value methods. In AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., and Sorooshian, S., editors, *Extremes in a Changing Climate*, pages 115–162. Springer.

- Salvadori, G., De Michele, C., Kottegoda, N. T., and Rosso, R. (2007). *Extremes in nature: an approach using copulas*, volume 56. Springer Science & Business Media.
- Salvadori, G., Durante, F., and De Michele, C. (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences*, 15(11):3293–3305.
- Salvadori, G., Durante, F., De Michele, C., Bernardi, M., and Petrella, L. (2016). A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities. *Water Resources Research*, 52(5):3701–3721.
- Scaillet, O. (2007). Kernel-based goodness-of-fit tests for copulas with fixed smoothing parameters. *Journal of Multivariate Analysis*, 98(3):533–543.
- Schaller, N., Kay, A. L., Lamb, R., Massey, N. R., van Oldenborgh, G. J., Otto, F. E. L., Sparrow, S. N., Vautard, R., Yiou, P., Ashpole, I., Bowery, A., Crooks, S. M., Haustein, K., Huntingford, C., Ingram, W. J., Jones, R. G., Legg, T., Miller, J., Skeggs, J., Wallom, D., Weisheimer, A., Wilson, S., Stott, P. A., and Allen, M. R. (2016). Human influence on climate in the 2014 southern England winter floods and their impacts. *Nature Climate Change*, 6(6):627–634.
- Serinaldi, F. (2015). Dismissing return periods! *Stochastic Environmental Research and Risk Assessment*, 29(4):1179–1189.
- Shiau, J.-T. (2003). Return period of bivariate distributed extreme hydrological events. *Stochastic Environmental Research and Risk Assessment*, 17(1-2):42–57.
- Shiau, J.-T. (2006). Fitting drought duration and severity with two-dimensional copulas. *Water Resources Management*, 20(5):795–815.
- Shiau, J.-T., Feng, S., and Nadarajah, S. (2007). Assessment of hydrological droughts for the Yellow River, China, using copulas. *Hydrological Processes*, 21(16):2157–2163.
- Shiau, J.-T. and Modarres, R. (2009). Copula-based drought severity-duration-frequency analysis in Iran. *Meteorological Applications*, 16(4):481–489.
- Singh, V. P. and Zhang, L. (2007). IDF curves using the Frank Archimedean copula. *Journal of Hydrologic Engineering*, 12(6):651–662.
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.
- Szolgay, J., Gaál, L., Bacigál, T., Kohnová, S., Hlavčová, K., Vyleta, R., Parajka, J., and Blöschl, G. (2016). A regional comparative analysis of empirical and theoretical flood peak-volume relationships. *Journal of Hydrology and Hydromechanics*, 64(4):367–381.

- Trenberth, K. E., Fasullo, J. T., and Shepherd, T. G. (2015). Attribution of climate extreme events. *Nature Climate Change*, 5(8):725–730.
- Vandenberghe, S. (2012). *Copula-based models for generating design rainfall*. PhD thesis, Department of Forest and Water management, Ghent University.
- Vandenberghe, S., Verhoest, N. E. C., Onof, C., and De Baets, B. (2011). A comparative copula-based bivariate frequency analysis of observed and simulated storm events: A case study on Bartlett-Lewis modeled rainfall. *Water Resources Research*, 47(7):W07529.
- Verhoest, N., Troch, P. A., and De Troch, F. P. (1997). On the applicability of Bartlett-Lewis rectangular pulses models in the modeling of design storms at a point. *Journal of Hydrology*, 202(1-4):108–120.
- Vernieuwe, H., Vandenberghe, S., De Baets, B., and Verhoest, N. E. C. (2015). A continuous rainfall model based on vine copulas. *Hydrology and Earth System Sciences*, 19(6):2685.
- Verstraeten, G., Poesen, J., Demarée, G., and Salles, C. (2006). Long-term (105 years) variability in rain erosivity as derived from 10-min rainfall depth data for Ukkel (Brussels, Belgium): Implications for assessing soil erosion rates. *Journal of Geophysical Research*, 111:D22109.
- VMM (2018). Waterinfo.be. <https://www.waterinfo.be/default.aspx?path=NL/Rapporten/Downloaden>. Laatst geraadpleegd op 14 maart 2018.
- Wahl, T., Jain, S., Bender, J., Meyers, S. D., and Luther, M. E. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Climate Change*, 5(12):1093.
- Wang, C., Chang, N.-B., and Yeh, G.-T. (2009). Copula-based flood frequency (COFF) analysis at the confluences of river systems. *Hydrological Processes*, 23(10):1471–1486.
- Willems, P. (2013). Revision of urban drainage design rules after assessment of climate change impacts on precipitation extremes at Uccle, Belgium. *Journal of Hydrology*, 496:166–177.
- Wist, H. T., Myrhaug, D., and Rue, H. (2004). Statistical properties of successive wave heights and successive wave periods. *Applied Ocean Research*, 26(3-4):114–136.
- Yue, S. and Rasmussen, P. (2002). Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14):2881–2898.

## BIBLIOGRAFIE

---

Zhang, L. and Singh, V. P. (2007). Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology*, 332(1-2):93–109.