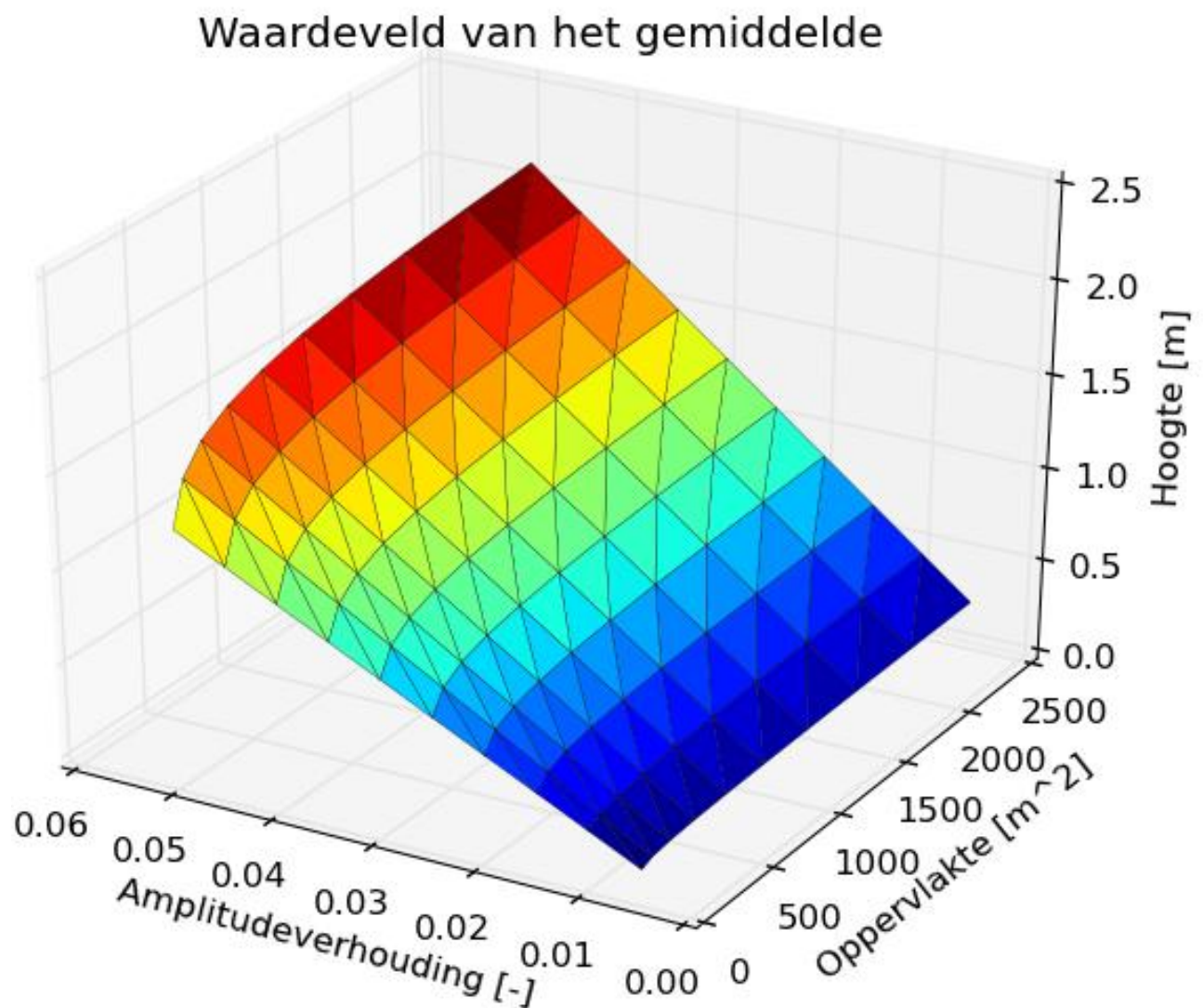


Extreme waarde verdeling van stochastische velden voor variërende parameters



Extreme waarde verdeling van stochastische velden voor variërende parameters

Door

C.J. Gijzenberg

BSc-eindwerk

Bachelor of Science
Civiele techniek
Technische universiteit Delft

Studentnummer: 4244486
Email: C.J.Gijzenberg@student.tudelft.nl
Datum: 27-10-2016

1^{ste} begeleider: Dr. Ir. P.C.J. Hoogenboom
2^{de} begeleider: Dr. Ir. M.A.N. Hendriks

Voorwoord

Dit rapport is gemaakt voor mijn BSc eindwerk om te kunnen afstuderen aan de technische universiteit in Delft. De aanleiding was een eerder BSc eindwerk, namelijk dat van T. van de Velde, dat ging over extreme waarde verdelingen van stochastische velden. De vraag aan mij was om te kijken wat er zou gebeuren als de variabelen zouden veranderen.

Het onderzoek is voornamelijk uitgevoerd met de programma's Python en EasyFit, en veel data zijn in het rapport zelf gezet. Met deze programma's is geprobeerd om een functie te maken die de maximale uitwijking van een schaal zou kunnen berekenen, zodat dit mogelijk later zou kunnen worden gekalibreerd voor constructieve doeleinden.

Het rapport is voornamelijk geschreven voor mensen die geïnteresseerd zijn in kansverdelingen en schaalconstructies. Het is handig om enige kennis te hebben van kansrekening.

*C.J. Gijzenberg
Delft, Oktober 2016*

Inhoud

Samenvatting.....	5
1. Inleiding	6
1.1 Inleiding	6
1.2 Doelstelling	6
1.3 Aanpak	6
2. Variabelen	7
2.1 Bereik van de amplitudeverhouding	7
2.2 Bereik van variabele I ; imperfectielengte	8
2.3 Bereik van variabele A ; oppervlakte	8
3. Maken en testen van de veldgenerator	9
3.1 Werking van de veldgenerator.....	9
3.2 Testen van de veldgenerator.....	9
4. Imperfectielengte van 5,5 meter	11
4.1 Verdelingstype.....	11
4.2 Het gemiddelde en de standaardafwijking	13
4.2.1 Het gemiddelde	13
4.2.2 De standaardafwijking	15
4.3 Terugkoppeling van het huidige model.....	17
5. Variabele imperfectielengte	18
5.1 Verdelingstype bij een variabele imperfectielengte	18
5.2 Waardevelden van de gemiddelde en de standaardafwijkingen.....	19
5.3 Één formule?	20
Conclusie.....	22
Aanbevelingen	23
Bronnen	24
Bijlage.....	25
Bijlage 1, De code van de veld generator	25
Bijlage 2, Code van Kolmogorov-Smirnov	26
Bijlage 3, Het stroomschema van hoofdstuk 4	27
Bijlage 4, Kolmogorov-Smirnov uitleg.....	28

Samenvatting

Extreme waarde verdeling van stochastische velden voor veranderende parameters gaat over de vraag of er een enkel verdelingstype samen met een functie en een aantal parameters, kan worden gegeven voor het berekenen van een maximale afwijking in een schaalconstructie, met een aantal gegevens zoals de oppervlakte.

Naar de volgende parameters is gekeken: de imperfectielengte, de amplitudeverhouding en de oppervlakte. De imperfectielengte is aangenomen tussen de 2,5 en 8,5 meter, de amplitudeverhouding is aangenomen tussen de 0,5% en de 5% en de oppervlakten die zijn onderzocht liggen tussen de 25 m^2 en 2.500 m^2 .

Er is een veldgenerator gemaakt, en deze kan velden van iedere grootte maken, en berekent voor iedere punt de hoogte, voor 400 punten per vierkante meter. De veldgenerator maakt gebruik van een sommatie van sinussen om ieder punt te berekenen. Het aantal sinussen hangt af van de grootste aangenomen imperfectielengte.

Met de veldgenerator zijn vervolgens 16 keer 5.000 velden gemaakt met een imperfectielengte van 5,5 meter, om te zien welk verdelingstype het beste zou passen. Dat bleken de bètaverdeling en de gegeneraliseerd extreme waarde verdeling te zijn. Ook bleken de lognormale verdeling en de gamma-verdeling goed te passen. Hierna zijn er 100 keer 100 velden gecreëerd bij dezelfde imperfectielengte, om een formule van het gemiddelde en de standaardafwijking te maken. De waarden van deze formule zijn vervolgens vergeleken met de eerder genoemde verdelingen en deze bleken nog goed te passen.

Naar een variabele imperfectielengte was nog niet gekeken, en met deze is ook gevarieerd. Eerst is er gekeken of er bij de variabele imperfectielengte nog steeds sprake was van een zelfde verdelingstype. Dit is gedaan door 16 keer 1.000 velden te creëren en deze te vergelijken. De vier eerder genoemde verdelingstypes leken nog steeds goed te passen. Met een zelfde type formule voor het gemiddelde en de standaardafwijking is ook gekeken naar een veranderende imperfectielengte, en dit keer met 4 keer 100 punten met 100 velden per punt. Uiteindelijk bleek dezelfde formule met andere (vier keer andere) constanten te passen. Er is ook geprobeerd om deze constanten met de imperfectielengte te verbinden, maar dit gaf geen goed resultaat.

Met andere woorden past er dus een aantal verschillende verdelingstypen, maar kan er geen passend model worden gemaakt voor de gemiddelde en de standaardafwijking van alle drie de variabelen.

De aanbevelingen voor een herhaal of vervolgonderzoek zijn als volgt: gebruik maken van een programma dat op meerdere punten verdelingen kan vergelijken, de veldgenerator aanpassen zodat er een aantal mogelijk fouten kunnen worden voorkomen, meer velden genereren om ruis te minimaliseren en beter omgaan met kleine oppervlakte in combinatie met grote imperfectielengte.

1. Inleiding

1.1 Inleiding

Dit BSc-eindwerk is een verdieping op een eerder BSc-eindwerk van T. van de Velde, *'Extreme waarde van stochastische velden'*. Hierin is gekeken naar het verdelingstype van de extreme waarden van stochastische velden, om dit toe te kunnen passen op vormimperfecties van schaalconstructies.

Aan de hand van meetresultaten van B. Elferink zijn er met behulp van een sommatie van sinussen, 10.000 stochastische velden gecreëerd en geanalyseerd. Er waren vijf typen verdelingen die de extreme waarden goed omschreven.

Voor het hele onderzoek stond het oppervlakte van de schaalconstructie vast; het was altijd 400 vierkante meter. De amplitudes en lengtes van de sinussen waren gebaseerd op de meetresultaten van B. Elferink, of de linearisering van deze meetresultaten.

De nieuwe vraag is dus wat er gebeurt als het oppervlakte wordt vergroot, of juist verkleind, en dit in combinatie met een andere amplitude/lengte verhoudingen.

1.2 Doelstelling

De doelstelling van dit onderzoek is het bepalen van invloed van de variabelen l , a en het schaaloppervlak A op de parameters van de extreme waarden verdeling; gemiddelde, standaardafwijking en verdelingstype.

1.3 Aanpak

Er moet eerst vooronderzoek worden gedaan naar het bereik van de verschillende variabelen.

Hierna wordt er met een veranderend oppervlakte gekeken naar het gemiddelde en de standaardafwijking van de maxima en minima, ten opzichte van een veranderende $l:a$ verhouding. Er zal ook gekeken worden of er één of meerdere verdelingstypen nodig zijn om de verdelingen te beschrijven, en of er een enkele formule kan worden gemaakt voor het gemiddelde en de standaardafwijking.

Dit moet vervolgens een aantal keer worden gedaan met verschillende imperfectielengtes.

Met deze data kan er een veld worden gecreëerd, met de A en $l:a$ verhouding op de assen, waar een gemiddelde en een standaardafwijking uit kunnen komen, die dan nog met een maximale lengte factor kunnen worden vermenigvuldigd, waarnaar er met een enkele verdeling de kans kan worden ingeschat op een maximale uitwijking.

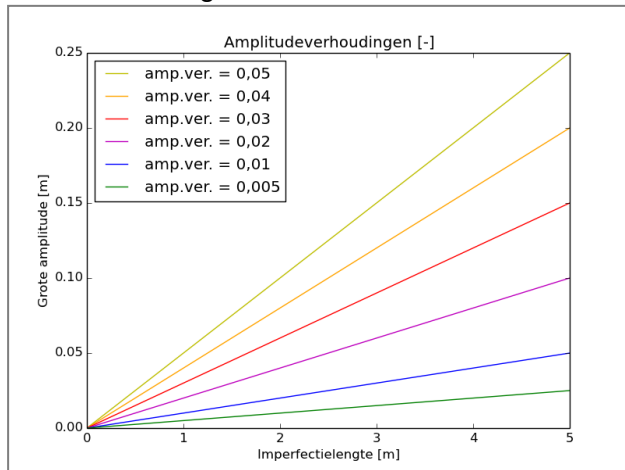
2. Variabelen

Er zijn drie variabelen die worden gebruikt en dus hier worden besproken. Het zijn de amplitudeverhouding, de imperfectielengte en de schaaloppervlakte. Het bereik van deze variabelen is met Dr. Ir. P.C.J. Hoogenboom besproken.

2.1 Bereik van de amplitudeverhouding

De amplitudeverhouding bestaat uit twee elementen: de amplitude en de lengte. Door de amplitude door de lengte te delen krijgt men de amplitudeverhouding.

De afbeelding hieronder is een visualisatie van een amplitudeverhouding, met alle amplitude voor iedere lengte.



Figuur 1: De grote van de amplitude samen met de imperfectielengte.

Korte opmerking:

Hoewel het lijkt dat er met de amplitudeverhouding ook de grootste hoek kan worden uitgerekend, kan dit echter niet; als er op een schaalconstructie geen grotere afwijking dan 1% kan worden geaccepteerd, dan moet de maximale amplitudeverhouding altijd kleiner zijn, omdat het een sommatie is van sinussen.

Ondergrens

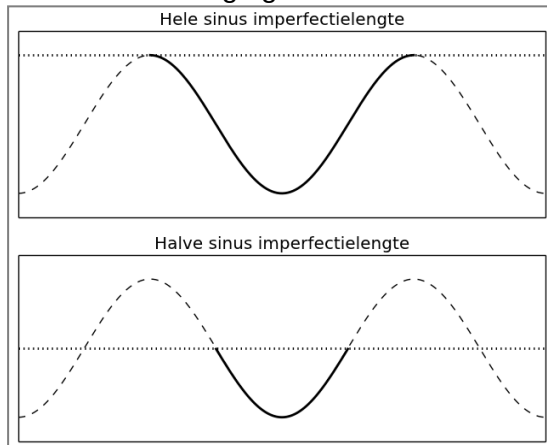
De ondergrens is gezet op 0,5% omdat de waarde die kleiner zijn niet echt interessant lijken te zijn om te onderzoeken. Dit zou later kunnen worden veranderd, omdat kleinere waarden zeker mogelijk zijn (Hoogenboom, 2016).

Bovengrens

De bovengrens is gezet op 5,0% omdat het onwaarschijnlijk lijkt dat een grotere amplitude mogelijk is zonder dat de schaal meteen stuk gaat. Vijf procent op een lengte van bijvoorbeeld vier meter is twintig centimeter, en aangezien de dikte van een (dunne)schaal meestal rond de tien centimeter ligt is dit te veel (Elferink, 2015).

2.2 Bereik van variabele l ; imperfectielengte

De variabele l staat voor de kniklengte van een vorm imperfectie. Hoewel het op een schaalconstructie een kromme lijn is, is het hier juist een rechte lijn die een de afstand van een halve sinusbeweging inhoud.



Figuur 2: Imperfectielengte (zwarte lijn), en de schaal (gestippelde lijn)

Korte opmerking:

De aangenomen vorm is een punt van discussie. Als een element aan weerskanten is ingeklemd kan de knikvorm niet worden beschreven als een halve sinus, maar moet het als een hele sinus worden beschreven. Maar is het element aan weerskanten niet moment vast opgelegd, dan kan het wel als een halve sinus worden beschreven.

Ondergrens

De ondergrens van de grootste imperfectielengte is 1 meter, omdat kleinere grootste imperfectielengtes niet reëel lijken (Hoogenboom, 2016). Uiteindelijk is er voor gekozen om 2,5 meter als kleinste aan te nemen.

Bovengrens

De bovengrens van de kniklengte hangt normaal af van de grote van de straal en de dikte van de schaalconstructie (Hoogenboom, 2016), maar in dit geval zal er geen bovengrens zijn. Er zal wel een bovengrens worden gekozen maar deze zal onafhankelijk van het oppervlak zijn, en dat is 8,5 meter geworden. Al wordt er wel geprobeerd om de lengte niet groter te maken dan de wortel van het oppervlak.

2.3 Bereik van variabele A ; oppervlakte

De variabele A staat voor het oppervlak van het stochastische veld, wat ongeveer gelijk staat met het oppervlak van een schaalconstructie.

Ondergrens

De ondergrens van het oppervlak is 25 vierkante meter, omdat er met minimaal 10.000 punten zal worden gerekend; 400 punten per vierkante meter. Als het aantal punten per vierkante meter zou worden vergroot naar bijvoorbeeld 4.000, dan kunnen er problemen ontstaan met het aantal punten bij de bovengrens. De computer bereikt dan zijn geheugenlimiet.

Bovengrens

De bovengrens is gezet op 2.500 vierkante meter, vanwege een aangenomen maximum van 1.000.000 punten die dan berekend moeten worden. Het zou kunnen dat in een later stadium deze oppervlakte nog wordt vergroot als blijkt dat de computer het aantal punten aan kan.

3. Maken en testen van de veldgenerator

3.1 Werking van de veldgenerator

De veldgenerator heeft een aantal variabelen en moet ook altijd een logisch veld produceren. Omdat er in een later stadium kan worden gekozen om de imperfectielengte aan te passen, wordt de lengte gevraagd; deze wordt vervolgens ingedeeld in stukken van 10 centimeter. Een lengte van 29 centimeter wordt dus in het 20-30 centimetervak gedeeld, en vervolgens wordt het gemiddelde van het vak genomen; 25 centimeter als langste imperfectielengte.

De methode van veldgeneratie is dezelfde die T. van de Velde gebruikt namelijk de volgende:

$$d = \sum a_i * \sin\left(\frac{x * \pi}{l_i} + \varphi_{x,i}\right) * \sin\left(\frac{y * \pi}{l_i} + \varphi_{y,i}\right)$$

Waarin:

d = waarde stochastische veld [m]

a_i = amplitude van de imperfectielengte [m]

l_i = imperfectielengte [m]

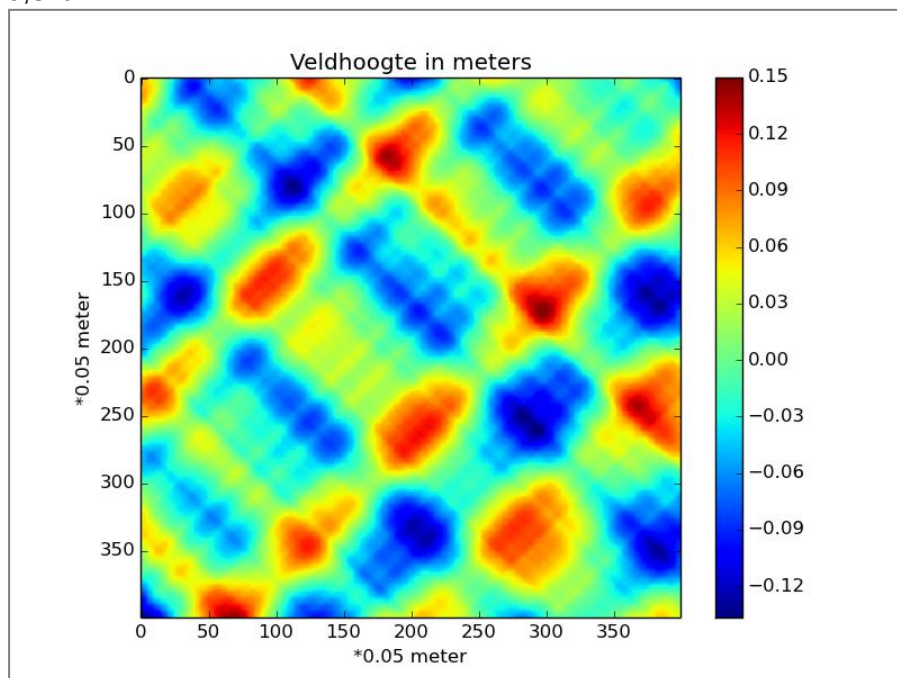
$\varphi_{x,i}$ = fase verschil van x [0 , 2π]

$\varphi_{y,i}$ = fase verschil van y [0 , 2π]

Voor de pythoncode van de veldgenerator zie Bijlage 1.

3.2 Testen van de veldgenerator

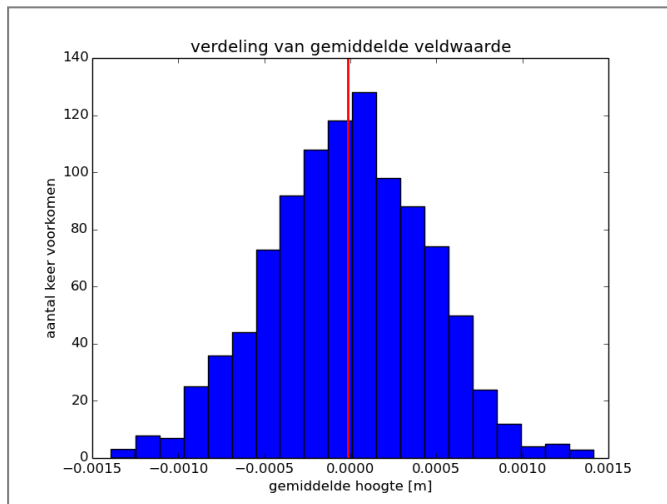
Om te testen of de generator goed werkt zal er een veld van 20 bij 20 meter worden gegenereerd, met een imperfectielengte van 5,5 meter, en een amplitudeverhouding van 0,5%.



Figuur 3: Het gegenereerde testveld.

Het lijkt er op dat het veld goed kan worden gegenereerd, zonder dat er sprongen zijn, of andere problemen zijn.

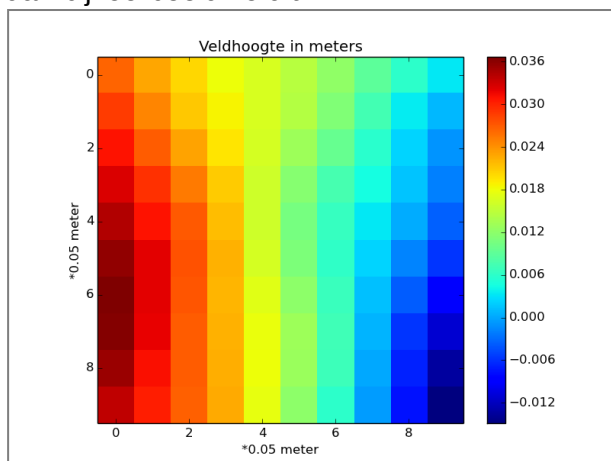
Verder moet het gemiddelde van het veld rond de nul liggen. Om te zien of dit klopt gaan er 1.000 velden met de hiervoor genoemde waarde worden gemaakt. Van ieder veld wordt de gemiddelde waarde berekend, en de verdeling van die gemiddelde waarde is op de volgende bladzijde te zien.



Figuur 4: De verdeling van de gemiddelde veldwaarde, met de rode lijn als gemiddelde daarvan.

Het gemiddelde zit op $-1,5 \cdot 10^{-5}$ meter, of wel -0,015 millimeter. Dit is voldoende klein om van te zeggen dat het gemiddelde op nul zit.

Doordat de imperfectielengte onafhankelijk is gehouden van het oppervlak kan dit, met kleine oppervlakten en grote imperfecties, problemen opleveren. Een enkele veld van 0,5 bij 0,5 meter met een imperfectielengte van 5,5 meter en een amplitudeverhouding van 0,5% ziet er dan bijvoorbeeld zo uit:



Figuur 5: Een kleine test veld om te laten zien dat bij kleine vlakken en grote imperfectielengte er geen goede testvelden ontstaan.

In eerste instantie lijkt er niks mee mis te zijn, totdat er naar de schaalbalk wordt gekeken; die gaat van -0,012 tot 0,036 meter, wat betekent dat het gemiddelde niet op nul zit. Dit betekent dat voor grote imperfectielengtes met kleine velden, de resultaten mogelijk verkeerd zijn. Verder is ook te zien hoe het programma de individuele vakjes inkleurt in combinatie met de aswaarden.

Als laatste is er nog iets aangepast. In het BSc-eindwerk van T. van de Velde wordt zowel de maximum als de minimum waarden meegenomen. Hier worden de maximum en minimum waarden ook opgezocht in ieder veld, maar de grootste (in absolute zin) wordt alleen meegenomen in de berekeningen. Zo wordt er voor gezorgd dat altijd de grootste waarde wordt mee genomen, en niet de op-drie-na of op-tien-na grootste waarde. Dit moet voornamelijk fouten in de kleine velden voorkomen.

4. Imperfectielengte van 5,5 meter

Eerst gaat er worden gekeken naar de verdelingstypen, het gemiddelde en de standaardafwijking bij een imperfectielengte van 5,5 meter. Zie ook bijlage 3, voor een schematische overzicht van dit hoofdstuk.

4.1 Verdelingstype

Nu kan er worden gekeken naar hoe de extreme waarde zich verdelen met de verschillende condities. De vraag is namelijk of er één enkele verdeling voor het aangenomen veld kan worden gebruikt, of dat voor verschillende delen er een ander verdelingstype moet worden gebruikt.

Er zal op zestien punten worden gekeken naar het verdelingstype, en dat zijn de punten waar de volgende lijnen kruisen; de lijnen op (0,5) , (2), (3,5) en (5) van de amplitudeverhoudings-as, en 25, 400, 1.225 en 2.500 vierkante meter van de oppervlakte-as. Verder worden er op ieder punt 5.000 velden gecreëerd, met 5,5 meter als grootste imperfectielengte.

De 5.000 maxima van deze velden worden geanalyseerd door het programma EasyFit (versie 5.6), en dit programma geeft vervolgens een lijst met verdelingstypen, en hoe goed deze passen. Dit 'passen' ofwel toetsen kan op verschillende manieren worden gedaan. De methode van Kolmogorov–Smirnov (voor de werking zie bijlage 4) wordt hiervoor gebruikt, omdat deze in het voorafgaande onderzoek ook is gebruikt, en het interessant is om te zien hoe dit vergelijkt.

Als de naar de eerste en dus beste fits wordt gekeken komt het volgende veld tevoorschijn:

Tabel 1: Best passende verdelingstypen per onderzocht punt.

Amplitude-verhouding	Oppervlakte			
	25 m ²	400 m ²	1225 m ²	2500 m ²
0,5 %	Johnson SB	Gen. Extreme Val.	Gen. Extreme Val.	Log-Logistic(3P)
2,0 %	Log-Pearson 3	Gen. Gamma (4P)	Gamma (3P)	Gen.Extreme Val.
3,5 %	Gen. Gamma (4P)	Johnson SB	Beta	Beta
5,0 %	Johnson SB	Gen. Gamma	Johnson SB	Beta

Dit is ver van een eenduidig veld, aangezien er acht verschillende verdelingstypen zijn. Als er wordt gekeken naar de beste vijf van ieder punt kan er misschien een scherper beeld worden geschept. De beste krijgt hier vijf punten, de tweede vier etc. Er zijn dus maximaal 80 punten te verdienen (totaal aantal punten is 240)

De hoogst scorende verdeling is dan de gegeneraliseerde extreme waarde verdeling met 41 punten. Als tweede komt de bètaverdeling met 40 punten.

Dit vertelt echter nog niet het hele verhaal; een enkel verdelingstype zou het makkelijkst zijn, maar als er voor één wordt gekozen moet er eigenlijk bekend zijn hoe goed een verdeling fit. Dus op de volgende bladzijde staan er vier tabellen met de Kolmogorov-Smirnov waarden van de gegeneraliseerde extreme waarde verdeling, de bètaverdeling, de lognormale verdeling en de gamma-verdeling. (de lognormale verdeling en gamma-verdeling verdienen respectievelijk 1 en 4 punten, maar het zijn meer bekende en gebruikte verdelingen en worden daarom ook genoemd)

Tabel 2: Kolmogorov-Smirnov waarden van de gegeneraliseerde extreme waarde verdeling bij een imperfectielengte van 5,5 meter. De functie gebruikt drie argumenten.

Gen. Extreme Value		Oppervlakte			
		25 m ²	400 m ²	1225 m ²	2500 m ²
Amplitude-verhouding	0,5 %	0,00976	0,00803	0,00658	0,00736
	2,0 %	0,00812	0,00614	0,00740	0,00605
	3,5 %	0,01123	0,00754	0,00747	0,00709
	5,0 %	0,00839	0,00911	0,00807	0,00635

Tabel 3: Kolmogorov-Smirnov waarden van de Bèta-verdeling bij een imperfectielengte van 5,5 meter. De functie gebruikt vier argumenten.

Bèta		Oppervlakte			
		25 m ²	400 m ²	1225 m ²	2500 m ²
Amplitude-verhouding	0,5 %	0,00847	0,00947	0,00809	0,00905
	2,0 %	0,00974	0,00572	0,00823	0,00884
	3,5 %	0,01094	0,00769	0,00700	0,00655
	5,0 %	0,00868	0,00800	0,00842	0,00530

Tabel 4: Kolmogorov-Smirnov waarden van de Lognormale verdeling bij een imperfectielengte van 5,5, meter. De functie gebruikt twee argumenten.

Lognormale		Oppervlakte			
		25 m ²	400 m ²	1225 m ²	2500 m ²
Amplitude-verhouding	0,5 %	0,02033	0,00999	0,00996	0,00956
	2,0 %	0,01655	0,01215	0,00697	0,01085
	3,5 %	0,02740	0,01073	0,00935	0,00804
	5,0 %	0,02376	0,01512	0,01098	0,00597

Tabel 5: Kolmogorov-Smirnov waarden van de Gamma-verdeling bij een imperfectielengte van 5,5 meter. De functie gebruikt twee argumenten.

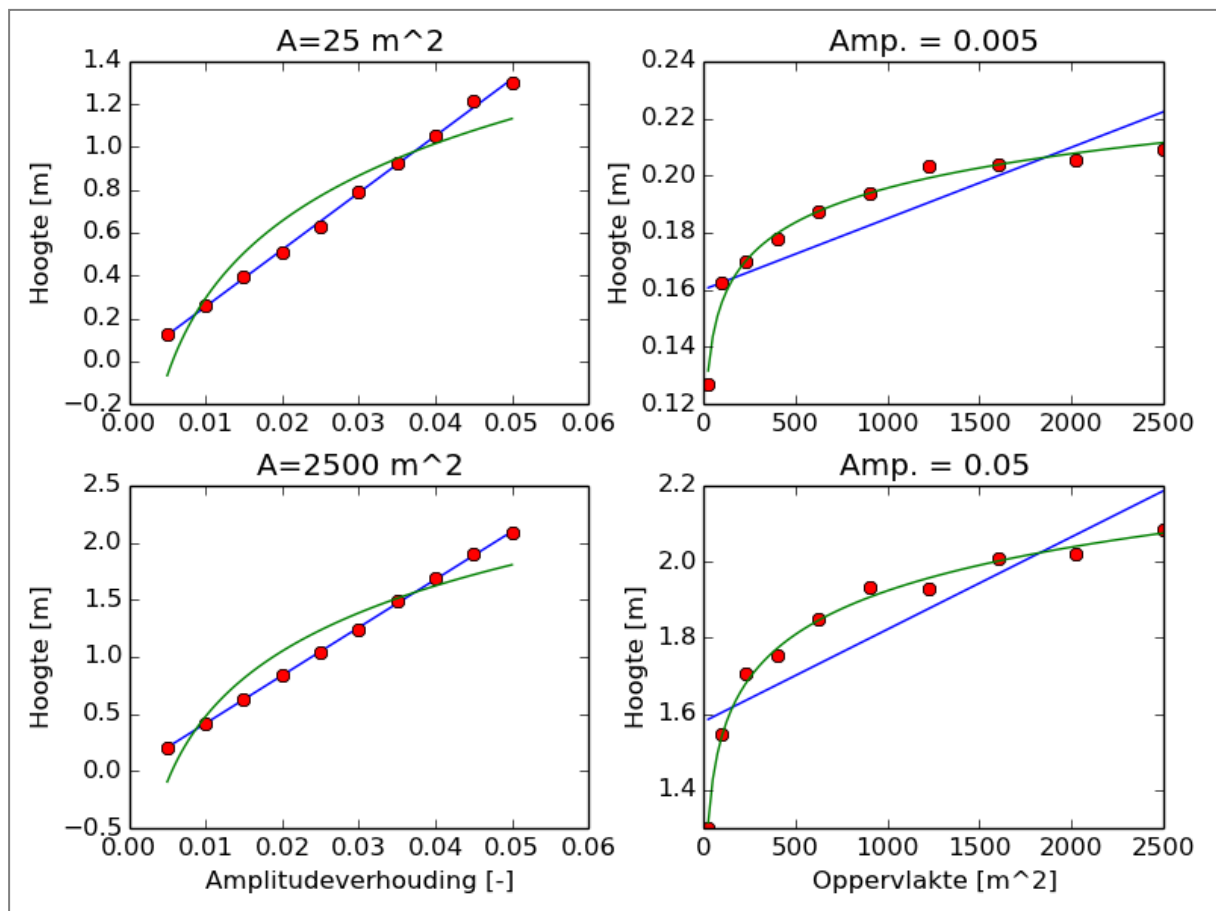
Gamma		Oppervlakte			
		25 m ²	400 m ²	1225 m ²	2500 m ²
Amplitude-verhouding	0,5 %	0,01074	0,01051	0,01820	0,01727
	2,0 %	0,00935	0,00853	0,01053	0,01164
	3,5 %	0,01158	0,01026	0,01099	0,01194
	5,0 %	0,01206	0,00671	0,01578	0,01347

4.2 Het gemiddelde en de standaardafwijking

Nu er bekend is welke verdelingstypen er kunnen worden gebruikt, is het handig als de gemiddelde en de standaardafwijking bekend zijn. Deze veranderen echter door het veld heen, dus moet daar een wiskundig veld voor worden gemaakt.

Om te beginnen wordt er onderzocht of de gemiddelde en de standaardafwijking lineair zijn, en dit wordt bekeken door naar de buitenste grenzen van het veld te kijken. Op ieder van de vier grenzen worden tien punten gekozen (met dezelfde tussen afstand op de amplitude as), en op ieder punt wordt er 100 keer een veld gecreëerd (enige constante is de imperfectie lengte: 5,5 meter), waar uit het gemiddelde en de standaardafwijking worden berekend. Met deze tien punten wordt vervolgens een plot gemaakt en er wordt een lineair model voor gemaakt, met de kleinste-kwadratenmethode.

4.2.1 Het gemiddelde



Figuur 6: De regressie lijnen; lineair (blauw), logaritmische (groen), van het gemiddelde voor de uiterste van het veld, met behulp van de kleinste-kwadratenmethode.

Zoals te zien is in figuur 6 lijken de blauwe lijnen niet helemaal goed te passen in het veld waar de oppervlakte varieert. Er zijn twee grote problemen met de lineaire variant namelijk:

1. Het gemiddelde kan een positieve waarde aanhouden bij een negatieve oppervlakte
2. Het gemiddelde zal naar oneindige grootte stijgen, hoewel dit beperkt zou moeten worden door de som van alle imperfectielengten vermenigvuldigd met de amplitudeverhoudingen.

Hierna is er voor gekozen om een andere type regressielijn uit te zoeken om te zien of deze beter past, namelijk de logaritmische. Er is ook gekeken naar een arctangens maar deze was zo slecht dat die niet verder vermeld wordt. Met behulp van de gemiddelde kwadratische fout

(MSE) en de gemiddelde absolute fout (MAE) worden de logaritmische met de lineaire vergeleken:

Tabel 6: De MeanSquareError-waarden en MeanAbsoluteError-waarden van de twee lijnen per grafiek, voor het gemiddelde.

Lijn	Lineair		Logaritmische	
	MSE	MAE	MSE	MAE
Lijn A, A = 25 m ²	0,00185071	0,01034723	0,14671090	0,10678986
Lijn J, A = 2500 m ²	0,00131369	0,00952176	0,34915113	0,16446369
Lijn 1, amp = 0,005	0,00184509	0,01034165	0,00009447	0,00221341
Lijn 10, amp = 0,05	0,14618035	0,09952828	0,00345856	0,0159519

De conclusie is dat het gemiddelde zich, met verschillende oppervlakten, laat leiden door het natuurlijke logaritme van de oppervlakte, en lineair als de amplitudeverhouding verandert. Er zou een simpele formule kunnen worden gemaakt door de beide functies bij elkaar op te tellen, maar dit geeft echter een ver van zuiver resultaat. Als er een extra term met zowel x en y wordt gebruikt wordt het een stuk zuiverder. Dit geeft de volgende formule:

$$\mu(x,y) = A * y + B * \ln(x) + C * \ln(x) * y + D$$

Met:

μ = gemiddelde hoogte [m]

x = oppervlakte [m²]

y = amplitudeverhouding [-]

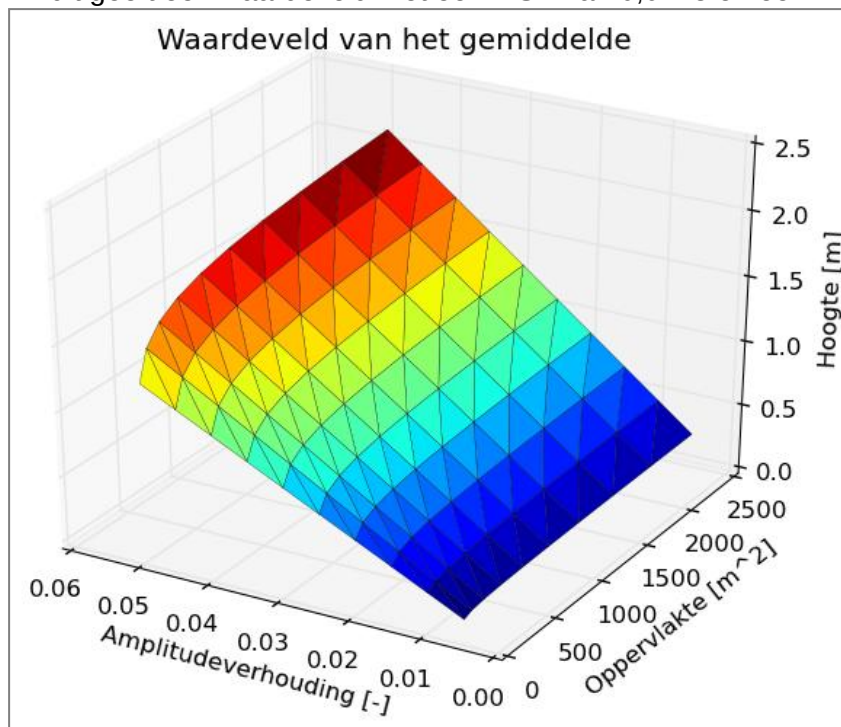
A = constante [m], = 16,213

B = constante [m⁻¹], = 0,00171

C = constante [m⁻¹], = 3,249

D = constante [m], = -0,01264

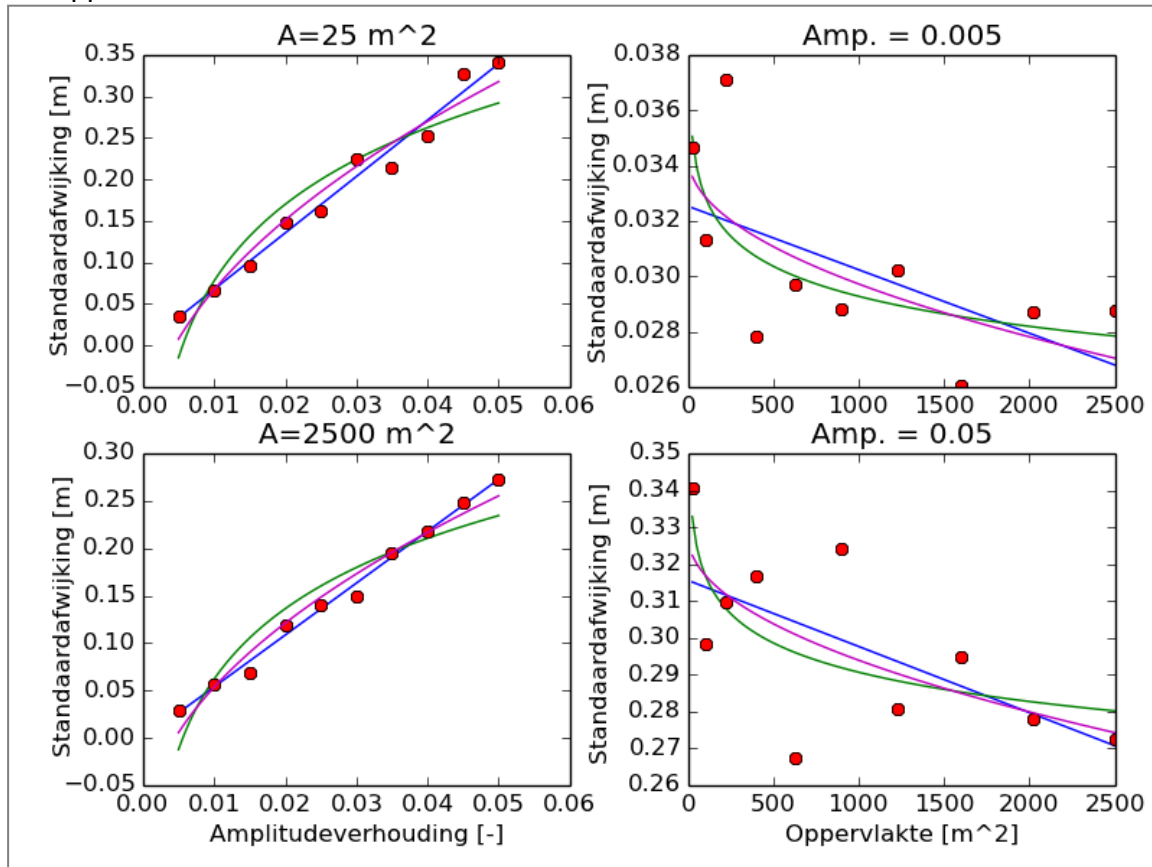
En dit geeft een waardeveld met een MSE van 0,0275 en een MAE van 0,01284.



Figuur 7: Een impressie van het waardeveld van het gemiddelde.

4.2.2 De standaardafwijking

Voor de standaardafwijking is gekozen om alleen de lineaire, de logaritmische en de wortel te onderzoeken. De wortel is gekozen om te vergelijken met de eerste twee, in de velden waar het oppervlak varieert.



Figuur 8: De regressie lijnen; lineair (blauw), logaritmisch (groen), wortel(magenta), van de standaardafwijking voor de uiterste van het veld, met behulp van de kleinste-kwadraten methode.

De standaardafwijking, als de amplitudeverhouding wordt aangenomen als vast (dus de rechter grafieken van figuur 8), is erg ontstabiel, hoewel de waarde relatief klein zijn. Er is een tweede keer een zelfde aantal velden gecreëerd, en daarbij bleek de instabiliteit nog steeds aanwezig te zijn.

Tabel 7: De MeanSquareError-waarden en MeanAbsoluteError-waarden van de drie lijnen per grafiek, voor de standaardafwijking.

Lijn	Lineair		Logaritmisch		Wortel(x)	
	MSE	MAE	MSE	MAE	MSE	MAE
Lijn A, A = 25 m ²	0,00006379	0,00212729	0,00005097	0,00165875	0,00005454	0,00194664
Lijn J, A = 2500 m ²	0,00318123	0,0136766	0,00275986	0,01346213	0,00293991	0,01343674
Lijn 1, amp = 0,005	0,00209543	0,01135222	0,01146959	0,02940352	0,00439229	0,01814672
Lijn 10, amp = 0,05	0,00054333	0,00547276	0,00682393	0,02219791	0,00209019	0,01120630

Tabel 7 laat zien dat de keuze niet makkelijk is. Toch zal hier ook voor een regressielijn worden gekozen. Het natuurlijke logaritme lijkt toch de beste keus te zijn, en past eigenlijk ook het beste bij de lijn van het gemiddelde; de grootste waarde zullen steeds vaker worden bereikt, en dus zal de standaardafwijking afnemen.

De conclusie is dus dat de standaardafwijking zich, met verschillende oppervlakten, laat leiden door het natuurlijke logaritme van de oppervlakte, en dat de amplitudeverhouding lineair werkt. Dit geeft de volgende formule:

$$\sigma(x, y) = A * y + B * \ln(x) + C * \ln(x) * y + D$$

Met:

σ = standaardafwijking [m]

x = oppervlakte [m²]

y = amplitudeverhouding [-]

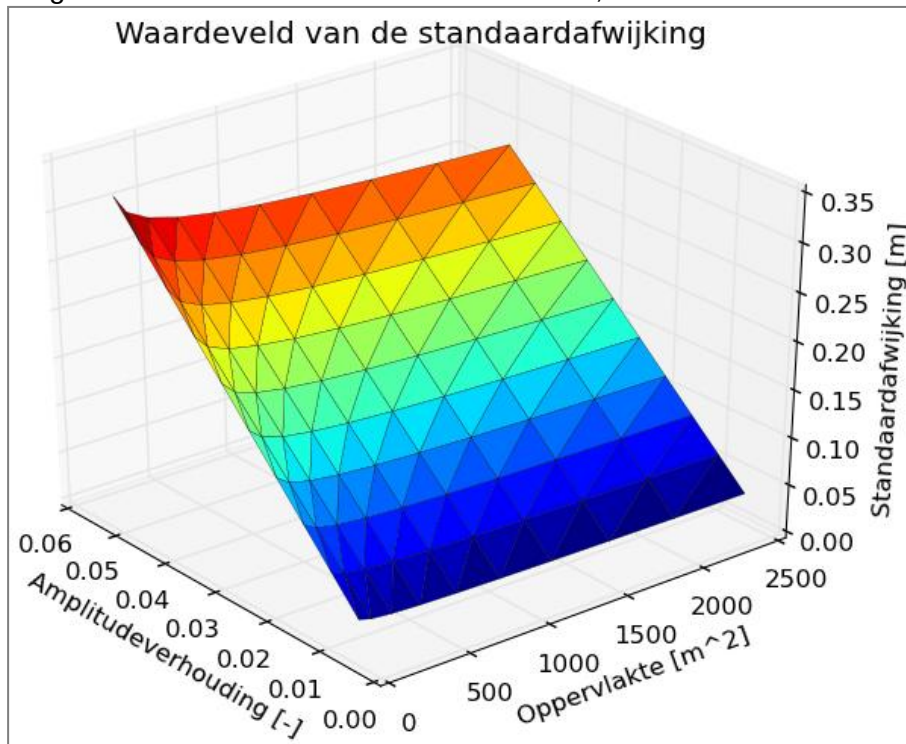
A = constante [m], = 7,564

B = constante [m⁻¹], = -0,001425

C = constante [m⁻¹], = -0,255879

D = constante [m], = 0,00837

Dit geeft een waardeveld met een MSE van 0,0141 en een MAE van 0,0085.



Figuur 9: Een impressie van het waardeveld van de standaardafwijking.

4.3 Terugkoppeling van het huidige model

Met het maken van het huidige model zijn er opnieuw een aantal fouten bij gekomen, doordat er gemiddelden zijn genomen, en afrondingen zijn gemaakt. Nu gaan de velden worden vergeleken met de waarden en verdelingstypen van de 16 eerder gecreëerde velden. Deze keer wordt er alleen naar de fit gekeken van de lognormale en gamma-verdeling omdat deze maar twee variabelen nodig hebben, en er maar twee variabelen zijn. De benodigde formules zijn uit het programma Easyfit gehaald, en de pythoncode staat in bijlage 2.

Tabel 8: Kolmogorov-Smirnov waarden van de Lognormale verdeling bij een imperfectielengte van 5,5 meter, met de veldwaarde van het gemiddelde en de standaardafwijking.

Lognormal		Oppervlakte			
		25 m ²	400 m ²	1225 m ²	2500 m ²
Amplitude-verhouding	0,5 %	0,0912	0,0321	0,0265	0,0186
	2,0 %	0,0169	0,0294	0,0099	0,0106
	3,5 %	0,0228	0,0169	0,0089	0,0168
	5,0 %	0,0243	0,0138	0,0158	0,0179

Tabel 9: Kolmogorov-Smirnov waarden van de Gamma-verdeling bij een imperfectielengte van 5,5 meter, met de veldwaarde van het gemiddelde en de standaardafwijking

Gamma		Oppervlakte			
		25 m ²	400 m ²	1225 m ²	2500 m ²
Amplitude-verhouding	0,5 %	0,0809	0,0218	0,0237	0,0223
	2,0 %	0,0114	0,0304	0,0140	0,0096
	3,5 %	0,0198	0,0180	0,0118	0,0144
	5,0 %	0,0230	0,0151	0,0144	0,0173

Op een paar uitschieters na lijken de modellen nog goed te passen. In het algemeen is de fout wel iets groter geworden, maar dat was verwacht, maar blijft nog binnen de 5% marge.

5. Variabele imperfectielengte

In dit hoofdstuk zal de laatste variabele gaan variëren. De imperfectielengte is tot nu toe telkens 5,5 meter geweest en zal nu worden veranderd door het optellen of aftrekken van 1,5 of 3,0 meter.

5.1 Verdelingstypen bij een variabele imperfectielengte

Om de verdelingstypen van de vier nieuwe lagen te bepalen gaat er worden gekeken naar een totaal van zestien punten op deze lagen. Iedere laag zal dus vier punten hebben, en in ieder punt zullen er 1.000 velden worden gecreëerd en daar zullen telkens de grootste absolute waarden uit worden onthouden. Het aantal velden per punt is kleiner om tijd te besparen. De punten worden enigszins 'midden' op het vlak gekozen.

Tabel 10: De beste verdelingen, en de Kolmogorov-Smirnov waarden van de eerder gebruikte verdelingen, bij variërende oppervlakten, amplitudeverhouding en maximale imperfectielengte.

Opp. & amp.verh.	Imp.lengte	Beste fit	G.E.V.	Bèta	Lognormal	Gamma
400 m ² , 0,02	2,5 m	Johnson SB	0,01089	0,01047	0,01243	0,01649
	4,0 m	Burr (4p)	0,01681	0,01983	0,01873	0,01837
	7,0 m	G.E.V.	0,01235	0,01283	0,01977	0,01395
	8,5 m	Gamma	0,01567	0,01624	0,02091	0,01533
400 m ² , 0,035	2,5 m	Dagnum (4P)	0,02633	0,02907	0,03088	0,03862
	4,0 m	Burr (4p)	0,01408	0,01439	0,03086	0,02042
	7,0 m	Log-Pearson 3	0,01294	0,01443	0,01908	0,01605
	8,5 m	Gen. Gam. (4p)	0,01812	0,01780	0,02869	0,01903
1225 m ² , 0,02	2,5 m	G.E.V.	0,01802	0,01928	0,02887	0,03443
	4,0 m	Burr (4p)	0,01731	0,01663	0,01604	0,01797
	7,0 m	Pearson 6	0,01985	0,02163	0,02060	0,02976
	8,5 m	Burr (4p)	0,01654	0,01590	0,01442	0,02037
1225 m ² , 0,035	2,5 m	Pearson 6	0,01509	0,01412	0,01585	0,02220
	4,0 m	Johnson SB	0,02488	0,02874	0,02823	0,03610
	7,0 m	Gamma	0,01858	0,01982	0,01777	0,01744
	8,5 m	Burr (4p)	0,02253	0,02677	0,02179	0,03042

Zoals in tabel 10 te zien is passen de 'simpele' lognormale en gamma-verdelingen nog steeds maar de bètaverdeling en gegeneraliseerde extreme waarde verdeling passen nog altijd beter. Toch is de grootte van de fout voor de lognormale en gamma-verdeling nog niet zo groot dat ze direct zijn te verwerpen. Verder valt op dat de 4 parameter Burr-verdeling nu een aantal keer als beste past, en dus zou er kunnen worden gekeken of deze verdeling misschien een betere zou zijn voor het hele model. Dit wordt echter niet gedaan omdat het een 4 parameter verdeling is.

5.2 Waardevelen van de gemiddelde en de standaardafwijkingen

Om de waardevelen te berekenen worden er vier nieuwe velden gemaakt van ieder 100 punten. Ieder punt staat voor het gemiddelde en de standaardafwijking van 100 velden in dat punt. Dezelfde formules voor het gemiddelde en de standaardafwijking worden gebruikt als eerder vermeld.

Dit betekent dat er voor beide velden de vier constanten worden uitgerekend, en deze worden in de volgende tabel 11 vermeld samen met de MSE en de MAE.

Tabel 11: De acht constanten, de MSE en de MAE voor de vier verschillende imperfectielengte.

Imperfectielengte [m]	2,5	4,0	7,0	8,5
Gemiddelde, constante A [m]	6,324	10,953	20,662	20,768
Gemiddelde, constante B [m^{-1}]	$9,334 \cdot 10^{-4}$	$-3,409 \cdot 10^{-5}$	$4,676 \cdot 10^{-3}$	$-1,230 \cdot 10^{-2}$
Gemiddelde, constante C [m^{-1}]	1,044	2,094	4,640	6,930
Gemiddelde, constante D [m]	$-6,126 \cdot 10^{-3}$	$-1,357 \cdot 10^{-3}$	$-2,442 \cdot 10^{-2}$	$7,526 \cdot 10^{-2}$
Gemiddelde MSE	0,00147	0,00994	0,05830	0,15465
Gemiddelde MAE	0,00287	0,00716	0,01828	0,02830
Standaardafwijking, c. A [m]	2,219	5,035	10,817	14,328
Standaardafwijking, c. B [m^{-1}]	$-9,494 \cdot 10^{-4}$	$-1,867 \cdot 10^{-4}$	$-2,831 \cdot 10^{-3}$	$-2,687 \cdot 10^{-3}$
Standaardafwijking, c. C [m^{-1}]	$-1,268 \cdot 10^{-1}$	$-2,552 \cdot 10^{-1}$	$-3,161 \cdot 10^{-1}$	$-3,735 \cdot 10^{-1}$
Standaardafwijking, c. D [m]	$6,485 \cdot 10^{-3}$	$1,220 \cdot 10^{-3}$	$1,545 \cdot 10^{-2}$	$1,883 \cdot 10^{-2}$
Standaardafwijking MSE	0,00137	0,00688	0,03447	0,11959
Standaardafwijking MAE	0,00259	0,00642	0,01385	0,02508

Met de waarden van tabel 11 en de zestien punten (die gebruikt zijn voor het controleren van de verdelingstype), kan nu worden gecontroleerd of de verdelingstypen nog voldoende passen.

Tabel 12: De Kolmogorov-Smirnov waarden voor Lognormale verdeling met de veldwaarde van tabel 11.

Imperfectielengte	2,5 m	4,0 m	7,0 m	8,5 m
A = 400, amp. = 0,02	0,01680	0,02147	0,03350	0,03244
A = 400, amp. = 0,035	0,02865	0,04288	0,02958	0,02858
A = 1225, amp. = 0,02	0,03182	0,03114	0,03438	0,01749
A = 1225, amp. = 0,035	0,02189	0,04218	0,02723	0,02939

Tabel 13: De Kolmogorov-Smirnov waarden voor Gamma-verdeling met de veldwaarde van tabel 11.

Imperfectielengte	2,5 m	4,0 m	7,0 m	8,5 m
A = 400, amp. = 0,02	0,01588	0,01658	0,03789	0,02358
A = 400, amp. = 0,035	0,02539	0,03751	0,02796	0,01707
A = 1225, amp. = 0,02	error	0,02293	0,04380	0,01562
A = 1225, amp. = 0,035	error	0,04999	0,01763	0,03816

De verdelingstypen lijken over het algemeen redelijk goed te passen, maar er is één ding dat opvalt; namelijk de twee errors in tabel 13. De oorzaak is niet geheel bekend, maar na een aantal testen lijkt het erop dat het te maken heeft met een integraal, die x^{a-1} moet berekenen.

Het probleem zit hem in de grootte van a, namelijk als deze waarde groter is dan 100 dan is er grote kans op een stackoverflow, met als gevolg dat er geen waarde wordt berekend.

5.3 Één formule?

Het effect van de imperfectielengte op de 'constante' A, B, C en D, is tot nu toe nog niet in een continue formule gepast. Er is ook een wens om de constante eenheidsloos te maken. Na enige analyse zijn de volgende formules opgesteld:

$$\mu(x, y, L) = a_m * L * y + b_m * \ln\left(\frac{1}{L}\right) * \ln(x) + c_m * \ln\left(\frac{1}{L}\right) * \ln(x) * y + d_m * L$$

$$\sigma(x, y, L) = a_s * L * y + b_s * \ln\left(\frac{1}{L}\right) * \ln(x) + c_s * \ln\left(\frac{1}{L}\right) * \ln(x) * y + d_s * L$$

Met:

μ = gemiddelde hoogte [m]

σ = standaardafwijking [m]

x = oppervlakte [m²]

y = amplitudeverhouding [-]

$a_m, b_m, c_m, d_m, a_s, b_s, c_s$ en d_s als constante [-]

De waarden van deze constanten hebben echter één nadeel; ze zorgen er voor dat het model niet goed meer past. Hiermee wordt bedoeld dat Kolmogorov-Smirnov waarden in de meeste lagen naar enkele tientallen procenten stegen.

Om een preciezere functie te creëren is vervolgens de wens om de constanten eenheidsloos te maken weggehaald. Eerst is gekeken hoe de imperfectielengte zich gedraagt als de andere variabelen worden vastgezet. De imperfectielengte lijkt een lineaire invloed te hebben, net zoals de amplitudeverhouding. Dit geeft de volgende formules:

$$\mu(x, y, L) = a_m * L * y + b_m * L * \ln(x) + c_m * L * \ln(x) * y + d_m * L + e_m * y + f_m * \ln(x) + g_m * \ln(x) * y + h_m$$

$$\sigma(x, y, L) = a_s * L * y + b_s * L * \ln(x) + c_s * L * \ln(x) * y + d_s * L + e_s * y + f_s * \ln(x) + g_s * \ln(x) * y + h_s$$

Met:

μ = gemiddelde hoogte [m]

σ = standaardafwijking [m]

x = oppervlakte [m²]

y = amplitudeverhouding [-]

$a_m, b_m, c_m, d_m, e_m, f_m, g_m, h_m, a_s, b_s, c_s, d_s, e_s, f_s, g_s$ en h_s als constante [voor eenheden en waarde zie tabel 14]

Tabel 14: De waarden van de acht constanten voor het gemiddelde en de standaardafwijking.

	Gemiddelde	Standaardafwijking
a [-]	9,546*10 ⁻¹	-3,070*10 ⁻²
b [m ⁻²]	-1,451*10 ⁻³	-4,080*10 ⁻⁴
c [m ⁻²]	2,573	2,000
d [-]	9,314*10 ⁻³	2,594*10 ⁻³
e [m]	-1,659	-6,232*10 ⁻²
f [m ⁻¹]	6,978*10 ⁻³	6,281*10 ⁻⁴
g [m ⁻¹]	8,314*10 ⁻¹	-3,007
h [m]	-4,508*10 ⁻²	-4,198*10 ⁻³

Tabel 15: Kolmogorov-Smirnov waarden voor Lognormale verdeling met de constanten waarden van tabel 14.

Imperfectielengte	2,5 m	4,0 m	5,5 m	7,0 m	8,5 m
A = 400, amp. = 0,02	0,31441	0,06858	0,07375	0,03570	0,05854
A = 400, amp. = 0,035	0,32290	0,06985	0,06078	0,02955	0,05466
A = 1225, amp. = 0,02	0,40008	0,08529	0,08575	0,05339	0,04858
A = 1225, amp. = 0,035	0,42388	0,11129	0,08370	0,02261	0,05864

De conclusie die uit tabel 15 kan worden getrokken, is dat er met deze formule geen correct model kan worden gemaakt waarbij de imperfectielengte ook een variabele is. Hoewel het op sommige stukken goed lijkt te passen is de fout op andere plaatsen meer dan veertig procent. Er is alleen naar de lognormale verdeling gekeken omdat de gamma-verdeling weer een aantal stack-overflows had.

Conclusie

De variabelen l (imperfectielengte), a (amplitude) en het schaaloppervlak A hebben allemaal invloed op de extreme waarde in een veld. De invloed van de imperfectielengte en de amplitude volgen een rechte lijn, en bij een natuurlijke logaritme volgt de schaaloppervlakte een rechte lijn.

Er bestaat niet één beste verdelingstype van de extreme waarde in een punt, maar er zijn wel een aantal die zeer goed passen, namelijk de bètaverdeling (vier argumenten) en de gegeneraliseerde extreme waarde verdeling (drie argumenten). Ook zou tot op zekere hoogte gebruik kunnen worden gemaakt van een 'simpelere' lognormale verdeling (twee argumenten) of een gamma verdeling (twee argumenten).

Het maken van een passend model voor het vinden van het gemiddelde is niet gelukt. Zolang de imperfectielengte als vast werd aangenomen zijn er passende waarde gevonden, maar als de imperfectielengte variabele werd, dan ontstonden er grote fouten. Dit geldt eveneens voor het model van de standaardafwijking.

Aanbevelingen

Het lijkt erop dat het probleem een stuk complexer is dan dat het eerst leek. Allereerst omdat er eigenlijk geen perfect verdelingstype uit is voortgekomen, wel een aantal op elkaar lijkende verdelingstypen die lokaal passen. Het is ook te zien aan het feit dat er geen volledige correcte formule kon worden gemaakt.

Als eerste de verdelingstypen. Voor een individueel punt is het, met het programma Easyfit, makkelijk om te vergelijken welk verdelingstype het beste is, maar als er naar velden van punten moet worden gekeken kan dit niet. Er moet dus een programma komen die met meerdere punten een goede verdeling kan kiezen.

Een betere formule voor het gemiddelde en de standaardafwijking zou kunnen worden bereikt door een verandering in de veldgenerator. Het probleem zit hem namelijk in het feit dat de huidige generator de hoogte iedere 5 centimeter berekent, en dat de lengte van de sinussen altijd deelbaar is door 5 centimeter. Dit betekent dat als een piek niet exact op de goede locatie ligt, het hoogste punt mogelijk nooit berekend wordt.

Er zijn twee oplossingen voor dit probleem. De eerste oplossing zou kunnen zijn dat de generator kleinere stappen neemt, (dus meer dan 400 punten per vierkante meter), en de tweede oplossing zou kunnen zijn dat de lengte van de sinus altijd getal is met drie of meer cijfers achter de komma.

Om een betere formule te maken zouden er ook meer velden kunnen worden gegenereerd. De huidige waarden komen uit vijf keer honderd punten, en voor ieder punt zijn er honderd velden gegenereerd. Er blijkt echter nog zo'n 7% ruis in te zitten, wat mogelijk voorkomen kan worden door meer velden te creëren.

Er zou ook naar de maximale imperfectielengte kunnen worden gekeken. Zoals al is vastgesteld kunnen grote imperfectielengten het beeld vertekenen bij kleine velden, maar dit is niet in het model meegenomen. De velden van 25 m^2 en met imperfectielengten groter dan 5 meter zijn dus mogelijk incorrect, maar ze worden wel mee genomen in de formule (al zijn het weinig punten die hiermee te maken hebben).

Als laatste zou het natuurlijk ook interessant zijn om te kijken naar het gebied buiten de formule, en of de formule terug is te koppelen naar de echte wereld. De waarden van de formule zijn namelijk niet teruggekoppeld aan de werkelijkheid, en moeten mogelijk nog geschaald worden vanwege de gebruikte methode; als de imperfectielengte groter werd, dan werd het aantal sinussen ook groter dat werd opgeteld voor het maken van een veld.

Bronnen

EasyFit (versie 5.6) [Computerprogramma]. MathWave Technologies.

Elferink, B. (2015). *Shell imperfections. A study of the shape and magnitude of geometrical imperfections in thin concrete shell structures*. (MSc-eindwerk) Civiele Techniek, Technische universiteit Delft, Delft.

Hoogenboom, Dr. Ir. P.C.J. (2016). uit diverse gespreken.

Velde, T. van de (2016). *Extreme waarde van stochastische velden*. (BSc-eindwerk) Civiele Techniek, Technische universiteit Delft, Delft.

Bijlage 1, De code van de veld generator

```

1 from pylab import *
2 import numpy
3 import csv
4
5
6 def great(A,B):          #kiest de absolute grootste waarde
7     num1 = abs(A)
8     num2 = abs(B)
9     if (num1<num2):
10        return num2
11    if (num2<num1):
12        return num1
13
14 def create_field(A,l,angle, beurt,out_put_num):      #A voor lengte oppervlak, l voor langste lengte, en angle voor de hoek tussen a:l
15     field = numpy.zeros((A*20,A*20))
16     X_as = linspace(0,A-0.05,(A*20))
17     Y_as = numpy.transpose([X_as])
18     lengte = int((l+0.05)*10.0)
19     array_lengte = linspace(0.05,((lengte*0.1)-0.05),(lengte))
20     array_hoek = array_lengte*angle
21     X_phase = 2*pi*numpy.random.rand(lengte)
22     Y_phase = 2*pi*numpy.random.rand(lengte)
23     for i in range(lengte):
24         OSP = array_hoek[i]*(sin(((X_as*pi)/array_lengte[i])+X_phase[i])*sin(((Y_as*pi)/array_lengte[i])+Y_phase[i]))
25         field = field + OSP
26     print 'ronde', beurt
27     if (out_put_num == 0):
28         return field
29     if (out_put_num == 1):
30         return numpy.amax(field),numpy.amin(field)
31     if (out_put_num == 2):
32         C = great(numpy.amax(field),numpy.amin(field))
33         return C
34
35 def save_numbers(A,x):          #slaat een veld op in een document
36     with open(x, 'w') as csvfile:
37         a = csv.writer(csvfile,delimiter=';', quotechar=' ', quoting=csv.QUOTE_MINIMAL)
38         for i in range(len(A)):
39             number=A[i]
40             #beta = ' '
41             a.writerow(number)
42
43
44 def create_gemenst(A,l,angle,times,out_put_num1,out_put_num):          #om meerdere veld van gemiddelde en standaardafwijking te geven
45     lijst= numpy.zeros((times,2))
46     for i in range(times):
47         lijst[i]= create_field(A,l,angle, [i],out_put_num1)
48     lijst_2 = numpy.transpose(lijst)
49     alfa = lijst_2[0]
50     beta = lijst_2[1]
51     A1 = numpy.mean(alfa)
52     A2 = numpy.std(alfa)
53     B1 = numpy.mean(beta)
54     B2 = numpy.std(beta)
55     if (out_put_num == 1):
56         return A1,A2,B1,B2
57     if (out_put_num == 2):
58         return A1, A2
59     if (out_put_num==3):
60         return alfa
61     if (out_put_num==4):
62         return beta
63     #print 'hello'
64
65 def show_field(A,l,angle):          #om data in een hoogte veld te vertonen
66     z= create_field(A,l,angle,0,0)
67     plt.imshow(z, interpolation='nearest')
68     plt.title('Veldhoogte in meters')
69     plt.xlabel('x*0.05 meter')
70     plt.ylabel('y*0.05 meter')
71     plt.colorbar()
72     show()
73
74 def show_bargr(A,l,angle,times,typ1,typ):          #om data in een bar graph te laten zien
75     z = create_gemenst(A,l,angle,times,typ1,typ)
76     plt.hist(z, bins=10,range=(numpy.amin(z),numpy.amax(z)))
77     show()
78
79 #volgende stuk was alleen voor 1000 velden in een punt.
80 #poins = create_gemenst(35,8.5,0.035,1000,2,3)
81 #save_numbers(poins, 'C:\Users\CJ.Gijzenberg\Documents\Kees-Jaap\studie\spambep\spam.csv')
82
83
84 #volgende stuk was om 100 keer 100 velden te maken met een bepaald imperfectie lengte
85 keer = 10
86 mean=numpy.zeros((keer,keer))
87 stand= numpy.zeros((keer,keer))
88 ampv= linspace(0.005,0.05,keer)
89 opv = linspace(5,50,keer)
90 for al in range(keer):
91     for als in range(keer):
92         mean[al,als], stand[al,als]=create_gemenst(opv[als],2.5,ampv[al],100,2,2)
93     print 'ronde', al, als
94 save_numbers(mean, 'C:\Users\CJ.Gijzenberg\Documents\Kees-Jaap\studie\spambep\spam.csv')
95 save_numbers(stand, 'C:\Users\CJ.Gijzenberg\Documents\Kees-Jaap\studie\spambep\spam1.csv')
96

```

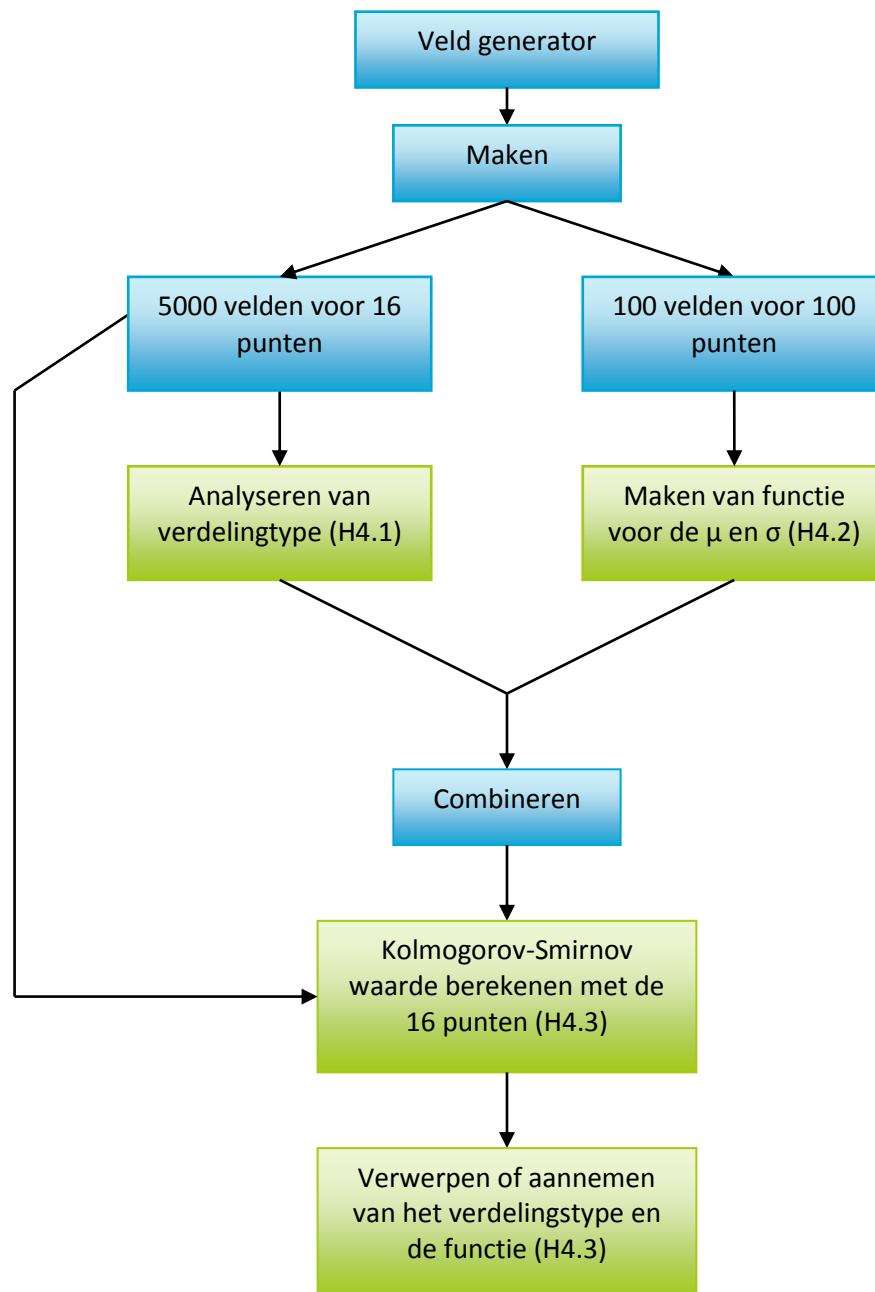
Bijlage 2, Code van Kolmogorov-Smirnov

```

1 from pylab import *
2 import numpy
3 import csv
4 import scipy.stats
5 import scipy.integrate as integrate
6
7 alfa=loadtxt('C:\Users\CJ.Gijzenberg\Documents\Kees-Jaap\studie\spambep\max1_20_8.5_3.5_1000.csv',delimiter=';')
8
9
10 L=8.5
11 y=0.035
12 x=20**2
13 num = 1000
14
15 A3 = sort(alfa)
16
17 for i in range(num):
18     if i<(num-1):
19         if (A3[i]==A3[i+1]):
20             print 'double trouble'
21 print 'done'
22
23 number1=True      #aan/uit knop voor de lognormale verdeling
24 number2=False     #aan/uit knop voor de gamma verdeling
25
26 G1 = 9.546e-01
27 G2 = -1.451e-03
28 G3 = 2.573
29 G4 = 9.314e-03
30 G5 = -1.659
31 G6 = 6.978e-03
32 G7 = 8.314e-01
33 G8 = -4.508e-02
34
35 S1 = -3.070e-02
36 S2 = -4.080e-04
37 S3 = 2.000
38 S4 = 2.594e-03
39 S5 = -6.232e-02
40 S6 = 6.281e-04
41 S7 = -3.007
42 S8 = -4.198e-03
43
44
45 #A1 en A2 zijn de geteste formules voor de velder
46 A1 = (G1*log(x)*y*L)+(G2*log(x)*L)+(G3*y*L)+(G4*L)+(G5*log(x)*y)+(G6*log(x))+(G7*y)+G8
47 A2 = (S1*log(x)*y*L)+(S2*log(x)*L)+(S3*y*L)+(S4*L)+(S5*log(x)*y)+(S6*log(x))+(S7*y)+S8
48
49 lmu = log(A1/sqrt(1+((A2**2)/(A1**2))))
50 lsi = sqrt(log(1+((A2**2)/(A1**2))))
51 gal = (A1**2)/(A2**2)
52 gbe = (A2**2)/A1
53
54 if number1==True:
55     kolmo = 0.0
56     kolmolast = 0.0
57     for j in range(len(A3)):
58         alta = j+1.0
59         tt = alta/num
60         lp = ((log(A3[j])-lmu)/lsi)
61         part = integrate.quad(lambda x: exp(-0.5*x*x),0,lp)
62         lapl = (1/(sqrt(2*pi)))*part[0]
63         kolmo= abs(lapl-0.5)-tt
64         if (kolmolast<kolmo):
65             kolmolast = kolmo
66     print 'lognormal:',kolmolast
67
68 if number2==True:
69     gamma1 = integrate.quad(lambda x: (x**(gal-1))*exp(-x),0,1000)
70     gamcom = gamma1[0]
71     kolmo2 = 0.0
72     kolmo2last = 0.0
73     for k in range(len(A3)):
74         belta = k+1.0
75         tt2 = belta/num
76         gamma2 = integrate.quad(lambda x: (x**(gal-1))*exp(-x),0,(A3[k]/gbe))
77         gaminc = gamma2[0]
78         gamcdf = gaminc/gamcom
79         kolmo2 = abs(gamcdf-tt2)
80         if (kolmo2last<kolmo2):
81             kolmo2last = kolmo2
82     print 'gamma:',kolmo2last
83

```

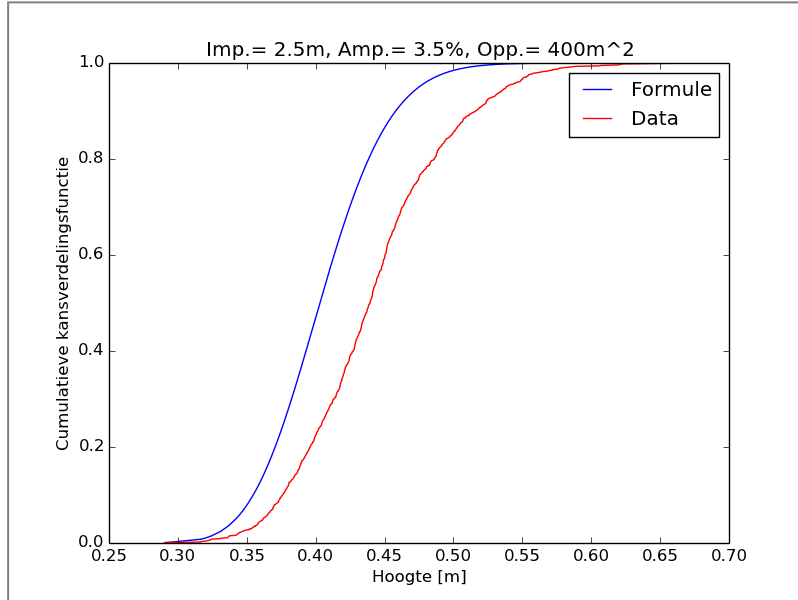
Bijlage 3, Het stroomschema van hoofdstuk 4



Bijlage 4, Kolmogorov-Smirnov uitleg

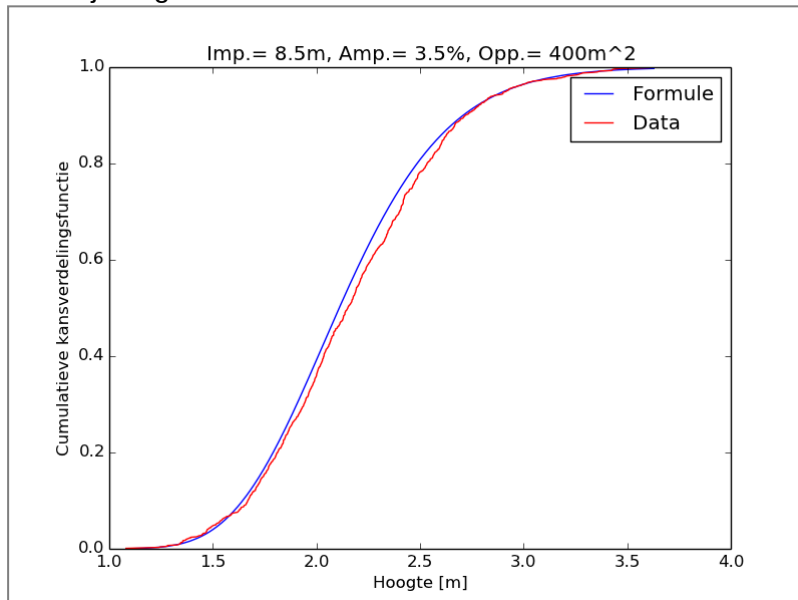
De Kolmogorov-Smirnov waarde wordt vaak in dit BSc-eindwerk gebruikt om te zien hoe goed een functie past, en de werking is eigenlijk heel simpel.

De waarde wordt namelijk verkregen door de absolute waarde te berekenen van de cumulatieve waarde van een formule/verdelingstype min de cumulatieve waarde van een data set. (Dit is effectief de grootste verticale afstand tussen de lijnen)



Afbeelding 1: Een 'slechte' Kolmogorov-Smirnov waarde

De afbeelding hierboven is een voorbeeld van een slechte fit, de grootste verticale afstand is namelijk ongeveer 32%. Hieronder een voorbeeld van een betere fit (namelijk 5,4%)



Afbeelding 2: Een 'goede' Kolmogorov-Smirnov waarde