

Rapport
aan
Dienst der Kustwegen
betreffende

RELATIES TUSSEN DE PARAMETERS IN DE METEOATLAS

1. INLEIDING

De meteoatlas bevat gemeten tijdreeksen van verscheidene parameters (onder andere: significante golfhoogtes H33 en HM0; gemiddelde golfperiode GTZ; laag-frequente golfenergie E10; golfrichtingen REM, RLF en RHF naargelang de beschouwde frequentieband van het golfspectrum; windsnelheid WVS en windrichting WRS) en dit voor verschillende locaties. Een vergelijking van deze meetwaarden is van belang om de (eventuele) relatie af te leiden tussen deze parameters.

Omwille van het groot aantal gegevens, de mogelijke aanwezigheid van uitbijters en de eventuele nonlineariteit is een goede voorstelling van de gegevens niet voor de handliggend en een analyse niet evident en soms weinig zinvol. Bijvoorbeeld: de toepassing van een regressie-analyse op windrichtingen is weinig zinvol omdat deze meetwaarden in een aantal gevallen duidelijke bi-modaliteit vertonen (d.w.z. dat de bivariate kansdichtheid verspreid is over twee populaties). Dit is een karakteristiek die niet kan samengevat worden door een éénduidige relatie tussen de parameters. Ook het eenvoudig fitten van een rechte lijn tussen parameters, waarbij zulke bi-modaliteit niet voorkomt, kan tot een foute voorstelling van de relatie leiden indien de relatie in werkelijkheid niet-lineair is en/of uitbijters voorkomen.

Om te voorkomen dat de grafische voorstellingen van de gegevens en de analyses tot foute conclusies zouden leiden, werd daarom een voorgaande studie uitgevoerd waarbij de karakteristieken van de relaties tussen de verschillende parameters werd onderzocht. In functie van deze karakteristieken werd daarna een grafische voorstelling en, waar mogelijk en zinvol, een samenvattende statistische analysemethode ontwikkeld. Dit rapport beschrijft deze studie en de voorstellingen en analyses die uiteindelijk werden weerhouden voor presentatie in de meteoatlas.

Het rapport is gestructureerd als volgt:

- Hoofdstuk 2 beschrijft de voorafgaande verwerking die nodig is om de meetwaarden uit de databank te selecteren. Deze selectie kan op twee wijzen worden uitgevoerd: 1. door extractie van simultane gegevens; 2. door extractie van orde-statistieken gemeten tijdens dezelfde observatieperiode;

- na voorstudie van de gegevens werd, afhankelijk van het type gegevens, uiteindelijk gekozen voor drie type voorstellingen die bondig worden samengevat in Hoofdstuk 3;
- Hoofdstukken 4, 5 en 6 beschrijven in meer detail elk van deze drie voorstellingswijzen en hun interpretatie. Deze paragrafen tonen ook geselecteerde voorbeelden van de analyse.

De besluiten van deze studie worden samengevat in Paragraaf 7.

2. EXTRACTIE VAN DE METINGEN: SIMULTANE METINGEN EN ORDE-STATISTIEKEN

Bij de extractie van de meetwaarden voor twee parameters X en Y worden steeds enkel die tijdstippen weerhouden waarbij zowel X en Y effectief gekend zijn. Zulke meetwaarden kunnen op twee wijzen worden beschouwd:

1. als **simultane metingen** waarbij de meting van X op tijdstip t wordt geassocieerd met de meting van Y op tijdstip t. In principe zou men ook tijdsverschuivingen tussen de twee tijdstippen kunnen beschouwen, indien vermoed wordt dat de hoogste correlatie zich voordoet bij zulk een tijdsverschuiving. In deze studie wordt zulk een tijdsverschuiving niet onderzocht, maar de methodes die verder worden uiteengezet zijn zonder wijziging van toepassing;
2. als meetwaarden van twee populaties die over een zelfde periode worden geobserveerd. In dit geval is het tijdstip van de meting van geen belang, maar is het de bedoeling om de respectieve kansverdeling van de beschouwde parameter op twee locaties X en Y te vergelijken. Daartoe dienen de **orde-statistieken** (d.w.z. de waarden die met een bepaalde frequentie worden overschreden) van X en Y te worden vergeleken.

Het aantal meetwaarden in de databank waarbij zowel X en Y gekend zijn is typisch zeer hoog (vaak 100,000 of meer) en een rechtstreekse grafische voorstelling en/of analyse van de oorspronkelijke meetwaarden is praktisch gezien onmogelijk. Daarom is een voorafgaande operatie nodig die enerzijds het aantal gegevens vermindert en anderzijds de oorspronkelijke karakteristieken van de meetwaarden zo volledig mogelijk weergeeft. Paragrafen 2.1 en 2.2 detailleren hoe deze datareductie gebeurt in het geval van de simultane metingen en in het geval van de orde-statistieken.

Onderzoek van de gegevens duidt verder aan dat voor 1 van de parameters beschouwd in deze studie (E10) een voorafgaande transformatie wenselijk is. Dit wordt verduidelijkt in Paragraaf 2.3. Dergelijke transformaties kunnen eventueel ook worden toegepast op de andere parameters om de grafische voorstellingen te verbeteren.

2.1. Data Reductie Voor Simultane Metingen

In het geval van de simultane metingen wordt aan elke meetwaarde X en Y een klasse k_x en k_y toegekend die overeenkomt met:

$$k_x = \text{int}\left(\frac{X}{\Delta}\right) \text{ en } k_y = \text{int}\left(\frac{Y}{\Delta}\right) \quad (1)$$

waarbij Δ een discretizatie-interval voorstelt dat wordt gekozen in functie van de beschouwde parameter. Op basis van deze voorstudie werd dit discretizatie-interval als volgt bepaald:

parameter	eenheid	Δ
significante golfhoogte (H33, HM0)	cm	2
laag-frequente golfenergie $E10^{1/2}$	cm	0.2
golfperiode (GTZ)	sec	0.1
richting (REM, RLF, RHF, WRS)	°	5
windsnelheid	m/sec	0.05

Voorgaande discretizatie-intervallen zijn gebaseerd op een compromis tussen enerzijds het zo goed mogelijk bewaren van het origineel karakter van de gegevens en anderzijds de noodzaak om het aantal te weerhouden gegevens te verminderen.

Binnen iedere klasse (k_x, k_y) worden vervolgens de gegevens gegroepeerd en worden volgende statistieken weerhouden:

- N: het aantal meetwaarden in de klasse;
- \bar{X} : het rekenkundig gemiddelde van de waarden X binnen de klasse;
- \bar{Y} : het rekenkundig gemiddelde van de waarden Y binnen de klasse.

Het zijn deze waarden die vervolgens worden uitgevoerd voor de verdere grafische en analytische verwerking. Merk op dat door het gebruik van de gemiddelde waarde eerder dan de klassemiddens de nauwkeurigheid verder wordt opgevoerd.

2.2. Data Reductie Voor Orde-Statistieken

Om de orde-statistieken te bepalen worden de weerhouden meetwaarden (X,Y) waarbij zowel X als Y gekend zijn, eerst gesorteerd van hoog naar laag respectievelijk voor X en Y. De gesorteerde waarden komen overeen met de orde-statistieken en worden hierna aangeduid als $X_{(i)}$ en $Y_{(i)}$, waarbij i de orde aanduidt. Voor $i=1$ vindt men de maximum waarden.

Om de data-reductie uit te voeren, wordt de reeks meetwaarden ($i, X_{(i)}, Y_{(i)}$) doorlopen en worden naast de maximum waarde ($i=1$), voor $i>1$ enkel die waarden i weerhouden wanneer hetzij

$$X_{(i-1)} - X_{(i)} > \Delta_x \quad (2)$$

hetzij

$$Y_{(i-1)} - Y_{(i)} > \Delta_y \quad (3)$$

Deze operatie verzekert opnieuw dat een voldoende resolutie wordt bereikt terwijl het aantal uit-te voeren gegevens beperkt blijft.

2.3. Voorafgaande Transformaties

In sommige gevallen kan het nuttig zijn om een voorafgaande transformatie uit te voeren. Op basis van deze voorstudie blijkt dit bijvoorbeeld het geval te zijn voor de laag-frequentie golfenergie E10 waarvan de waarden in de databank worden uitgedrukt in cm^2 . In deze oorspronkelijke schaal komen zeer hoge waarden voor zodat de keuze van een constant discretizatieinterval tot een groot aantal lege klassen ($n=0$) leidt bij de lagere waarden.

Daarom werd gekozen om als parameter de vierkantswortel van E10 te gebruiken. Deze parameter is dan in dezelfde schaal als de golfhoogte en blijkt ook beter gecorreleerd te zijn met de golfhoogte metingen.

Hoewel niet beschouwd in deze studie kan een dergelijke voorafgaande transformatie ook nuttig zijn voor de wind- en golfrichtingen. Uit deze studie blijkt dat de meest frequente richting vaak overeenkomt met een noordelijke of noordwestelijke richting (rond de 0°). Omdat deze richting op de grens valt van het beschouwde interval van 0 tot 360° , leidt dit tot een minder duidelijke grafische voorstelling. Daarom wordt voorgesteld om voor de uiteindelijke presentatie de richting 0° te definiëren als overeenkomend met de noordelijke richting. Dit kan gebeuren door de rekenkundige operatie:

$$X_{\text{referentie-noord}} = (X_{\text{oorspronkelijk}} + 180^\circ) \text{ modulus } 360^\circ \quad (4)$$

3. VERSCHILLENDE VOORSTELLINGSWIJZEN

Naargelang het type van de parameters (X,Y) die worden onderzocht worden volgende drie type-voorstellingen gebruikt:

1. een contourplot van de bivariate kansdichtheid:
Deze voorstellingswijze is aangewezen voor de onderlinge vergelijking van wind- en/of golfrichtingen. Een regressie-analyse van deze gegevens is te vermijden (zeker zonder een grafische voorstelling van de gegevens) omdat zulk een regressie-analyse het multimodale karakter van de gegevens verbergt. De methode zou ook kunnen toegepast worden voor de overige parameters;
2. een exploratieve en regressie-analyse van simultane gegevens:
Deze voorstellingswijze is aangewezen voor de onderlinge vergelijking van alle overige parameters (golfhoogte, golfperiode en windsnelheid). De voorstellingswijze omvat verschillende grafieken en analyses: ondermeer voorstelling van de scatterplot, bepaling van de kwantielwaarden van 1 parameter in functie van de andere parameter en, tenslotte, een regressie-analyse van deze kwantielwaarden. Bovendien wordt de exploratieve en de regressie-analyse toegepast zowel op de oorspronkelijke waarden als op de waarden na log-transformatie. De voorstellingswijze kan ook gebruikt worden om de eerder vermelde parameters te bekijken in functie van een golf- en/of windrichting (dit kan echter beter gebeuren door de marginale kansverdeling te bepalen voor specifieke richtingen). De relatie van wind- en/of golfrichtingen tot de eerder vermelde parameters is echter niet zinvol;
3. een empirische kwantiel-quantielplot:
Deze voorstellingswijze is aangewezen voor de onderlinge vergelijking van de orde-statistieken van de verschillende parameters, met uitzondering van wind- en/of golfrichtingen waarvoor het bepalen van orde-statistieken niet zinvol is. De orde-statistieken mogen niet behandeld worden door middel van een (klassieke) regressie-analyse omdat er in dit geval geen onderscheid is tussen de regressie van X naar Y enerzijds en de regressie van Y naar X anderzijds.

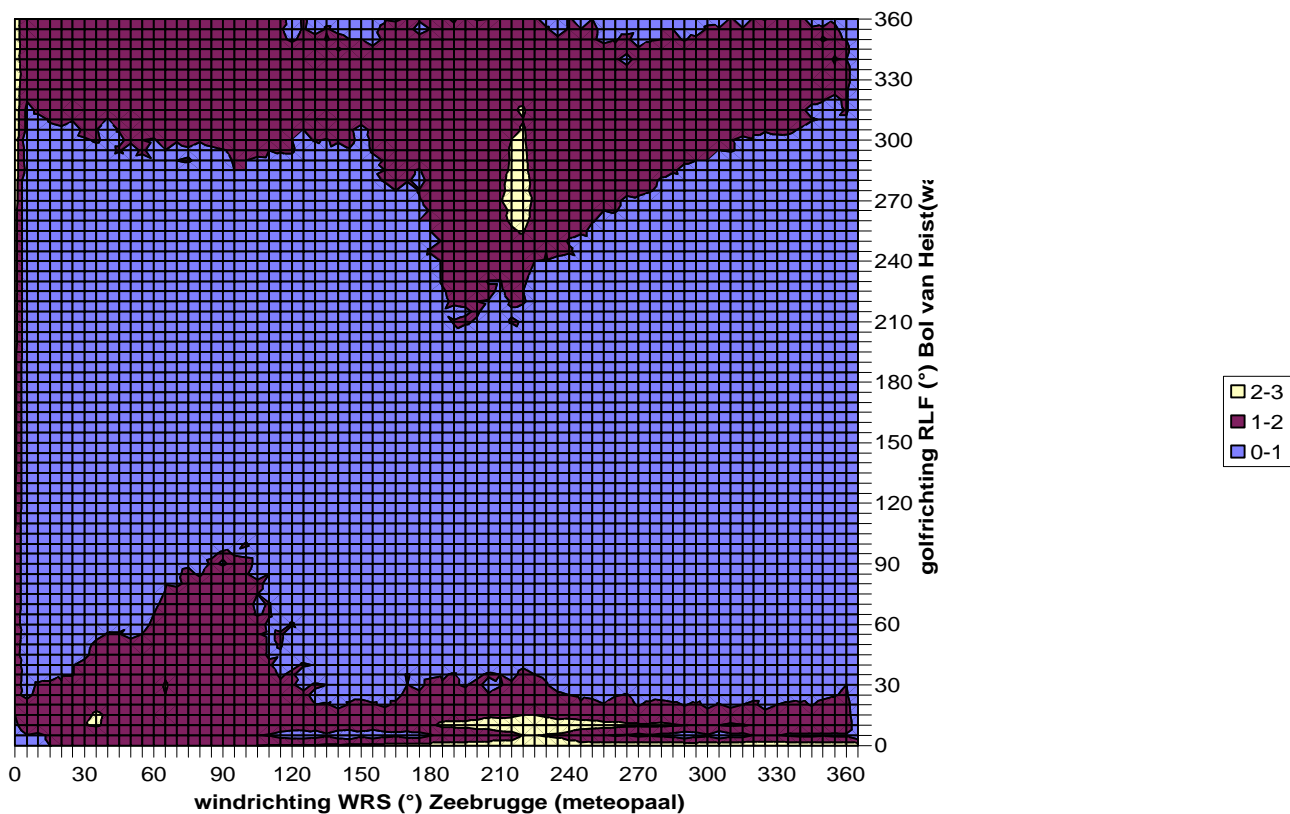
De aangewezen voorstellingswijzen worden in volgende tabel aangeduid in functie van de te-beschouwen parameters

		Y		
		richtingen	overige parameters	orde-statistieken alle parameters (exclusief richtingen)
X	simultane metingen	richtingen	contourplot	(regressie)
		overige parameters		regressie
	orde-statistieken	alle parameters (exclusief richtingen)		

4. CONTOURPLOT VAN DE BIVARIATE KANSDICHTHEID

Voor de beschrijving van de onderlinge relatie tussen wind- en/of golfrichtingen is een contourplot van de bivariate kansdichtheid het meest aangewezen. De volgende figuur toont een voorbeeld van een dergelijke contourplot voor de vergelijking van de windrichting ter plaatse van de meteopaal in Zeebrugge met de laag-frequente golfrichting ter plaatse van de Bol van Heist.

CONTOUR MAP VAN LOG10 AANTAL MEETWAARDEN



Deze figuur illustreert meteen de complexiteit van het verband tussen de richtingen: men merkt meteen 3 verschillende modi. Bij een noordoosterlijke windrichting (30°) is er een eerste piekwaarde van waarnemingen voor ongeveer dezelfde golfrichting bij de Bol van Heist. Voor diezelfde golfrichting vindt men echter het meest frequent een zuidwesterlijke windrichting (tussen de 210 en 240°). Bij diezelfde windrichting vindt men dan echter ook een tweede frequente voorkomende golfrichting in een noordwesterlijke richting. Zowel de wind- als de golfrichting vertonen dus verschillende modi. Zulke modi kunnen enkel duidelijk worden weergegeven in zulk een contourplot en daarom wordt enkel deze weergave (zonder verdere analyse) weerhouden voor de meteoatlas.

5. EXPLORATIEVE EN REGRESSIE-ANALYSE VAN SIMULTANE METINGEN

Voor de onderlinge vergelijkingen van simultane metingen (die niet overeenkomen met een richtingsmeting) is een regressie-analyse aangewezen (althans indien men geen multi-modaliteit vermoedt). Onderzoek van de gegevens toont echter dat zulk een analyse met de nodige zorg dient te worden uitgevoerd omdat:

1. de gegevens uitbijters kunnen bevatten;
2. de relatie niet-lineair kan zijn;
3. de dispersie rond de regressie-lijn kan variëren naargelang de beschouwde verklarende parameterwaarde.

Daarom werd gekozen om in de eerste plaats een aantal grafische voorstellingen te genereren die de gegevens voorstellen zonder enige bijkomende veronderstelling (exploratieve analyse) om daarna slechts een regressie-analyse uit te voeren gebruikmakend van een aantal statistieken die tijdens die exploratieve analyse worden afgeleid en robuust zijn ten overstaan van uitbijters.

5.1. Exploratieve Analyse

In deze sectie overlopen we de verschillende voorstellingswijzen gebruikt in de exploratieve analyse. Hiertoe gebruiken we als voorbeeld de vergelijking tussen de golfhoogte H33 en de golfperiode GTZ gemeten ter plaatse van de Akkaert waverider boei.

5.1.1. Scatterplots van de gegevens

Een eerste voorstellingswijze bestaat erin om het scatterdiagramma van de weerhouden gegevens weer te geven. Het scatterdiagramma wordt zowel voor de originele waarden als voor de loggetransformeerde variabelen getoond. De beide diagrammen tonen essentieel dezelfde informatie, maar naargelang het bereik van de gegevens is het ene diagramma duidelijker dan het andere. Voor het gekozen voorbeeld toont een weergave in de oorspronkelijke waarden de beste resolutie. Het resultaat hiervan wordt getoond in de volgende figuur.

Voor iedere klasse (n , k_x , k_y) wordt de gemiddelde waarde van Y uitgezet in functie van de gemiddelde waarde van X . De symboolgrootte in de figuur is functie van het aantal meetpunten binnen een gegeven klasse volgens een logaritmische schaal (2 pts voor $n < 10$, 3 pts voor $n < 100$, ...).

Het is meteen duidelijk dat de drie aandachtspunten eerder aangevoerd (uitbijters, niet-lineariteit, variatie van de dispersie) in dit geval alle drie van toepassing zijn.

**Exploratief onderzoek simultane metingen
golfhoogte H33 Akkaert (waverider) versus golfperiode GTZ Akkaert (waverider)**



5.1.2. Voorwaardelijke kwantielwaarden

Een tweede serie van exploratieve voorstellingswijze bestaat erin de kwantielwaarden van de ene parameter te tonen in functie van de waarde van de andere parameter. Men noemt dit voorwaardelijke kwantielwaarden. Deze voorstellingswijze wordt toegepast zowel voor X naar Y als voor Y naar X, en dit voor en na logtransformatie.

Voor iedere klasse van de controleparameter (de parameter in de x-as) wordt de benedenkwartielwaarde LQ, de mediaanwaarde MQ en de bovenkwartielwaarde UQ van de responsieparameter (de parameter in de y-as) getoond. Gegeven de parameter-waarde in de x-as worden deze waarden respectievelijk in 75%, 50% en 25% van de metingen overschreden. Op basis van deze waarden kan men ook grenzen LT en UT bepalen die door Tukey zijn voorgesteld om uitbijters te identificeren.

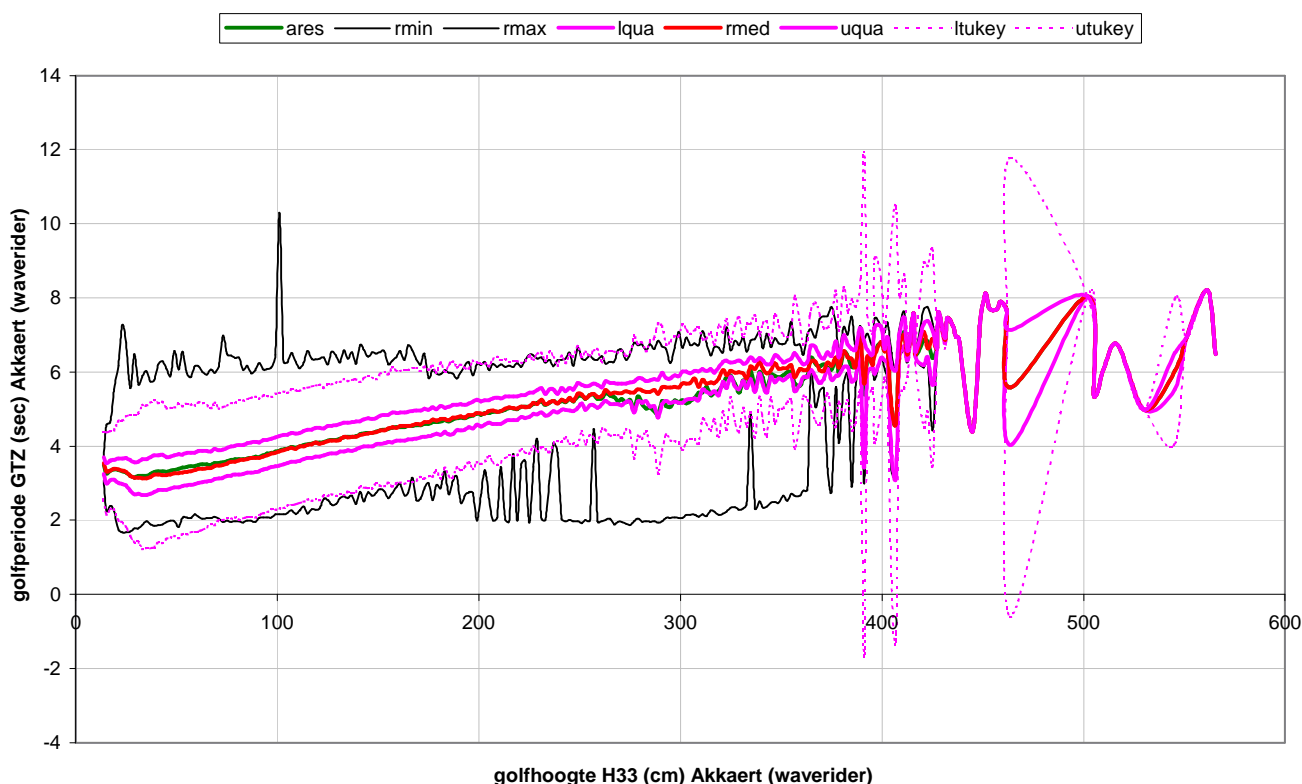
$$LT = LQ - 1.5 (UQ-LQ) \tag{5}$$

$$UT = UQ + 1.5 (UQ-LQ) \tag{6}$$

Waarden die buiten het interval (LT,UT) vallen worden meestal als uitbijters beschouwd. Om dit te beoordelen wordt in de figuur ook de minimum en maximum waarde in het interval aangeduid. Tenslotte wordt in de figuur eveneens het gemiddelde van de responsieparameter getoond.

Volgende figuur toont het resultaat voor de voorwaardelijke kwantielwaarden van GTZ (response variabele) in functie van H33 (controle variabele). De Tukey grenzen worden duidelijk overschreden: bij lagere waarnemingen van H33 zijn een aantal van de hogere metingen van GTZ eerder uitzonderlijk; voor waarnemingen van H33 tussen 2 en 4 meter is er een groep van uitbijters die reeds in het scatterdiagramma kon worden geïdentificeerd.

**Exploratief onderzoek simultane metingen
golfhoogte H33 Akkaert (waverider) versus golfperiode GTZ Akkaert (waverider)**



Men merkt ook op dat de voorwaardelijke kwantielwaarden soms erg variabel zijn: dit gebeurt wanneer het aantal waarnemingen van de responsievariabele voor de gekozen parameterwaarde klein is. Eliminatie van deze variatie is enkel mogelijk door middel van een regressie-analyse die hierna wordt besproken.

5.2. Regressie-analyse

Om de invloed van de uitbijters te beperken wordt de regressie-analyse uitgevoerd uitgaande van de kwantielwaarden eerder bepaald in de exploratieve analyse. De inputgegevens voor de regressie-analyse bestaat bijgevolg uit volgende reeks van waarden:

- X_j : de j 'de waarde van de predictieparameter. Deze waarde komt overeen met het gemiddelde binnen het discretizatieinterval eerder gekozen bij de extractie van de gegevens;
- N_j : het totaal aantal metingen van de responsieparameter wanneer de predictieparameter binnen het discretizatieinterval valt;
- LQ_j , MQ_j en UQ_j : de benedenkwartielwaarde, de mediaanwaarde en de bovenkwartielwaarde van de metingen van de responsieparameter wanneer de predictieparameter binnen het discretizatieinterval valt.

De nauwkeurigheid van de kwantielwaarden is invers proportioneel met de vierkantswortel van N_j . Bijgevolg dient een gewogen regressie te worden uitgevoerd met gewichten overeenkomstig het aantal N_j . Bovendien kan er niet van worden uitgegaan de regressie van de mediaanwaarde lineair verloopt over het ganze bereik van de predictieparameter. Daarom wordt een stuksgewijze lineaire regressie uitgevoerd. Dit gebeurt in drie stappen. In de eerste stap worden de breekpunten bepaald. In de tweede stap wordt de stuksgewijze lineaire regressie uitgevoerd. In de laatste stap wordt de standaarddeviatie van de dispersie rond de gefitte lijn bepaald. Hierbij wordt gebruik gemaakt van de boven- en benedenkwartielwaarde om de invloed van uitbijters te vermijden. Om toe te laten dat de standaarddeviatie functie is van de predictieparameter wordt opnieuw een stuksgewijze lineaire regressie uitgevoerd. In deze regressie worden in eerste instantie dezelfde breekpunten verondersteld als voor de fit van de mediaanwaarde. De significantie van de hellingverandering bij ieder breekpunt wordt echter getest en indien niet significant wordt het breekpunt verwijderd en de fit van de standaarddeviatie herhaald.

De drie stappen worden hierna in meer detail uiteengezet gevolgd door een aantal voorbeelden van de resultaten van de analyse.

5.2.1. Automatisch selectie van de breekpunten

De breekpunten van de stuksgewijze lineaire lijn worden bepaald door de predictieparameterwaarden te ordenen van hoog naar laag en vervolgens te onderzoeken hoever men (startend van de hoogste waarde) met goed resultaat een lineaire lijn kan fitten. Dit gebeurt door voor de mediaanwaarden binnen het beschouwde segment van predictiewaarden X_U tot X_L een gewogen kwadratische regressiefit uit te voeren. Omdat de regressie gewogen dient te worden uitgevoerd, worden pseudovariaten gebruikt waarbij zowel Y als alle verklarende X waarden en

het intercept met de vierkantswortel van N worden vermenigvuldigd. Om de significantie van de kwadratische term te beoordelen, dienen de kwadratische verklarende pseudo-predictieparameters $\sqrt{N}X^2$ orthogonaal te worden gekozen ten opzichte van de lineaire component $\sqrt{N}X$ en het intercept \sqrt{N} . Dit gebeurt door eerst een regressie uit te voeren van $\sqrt{N}X^2$ naar de twee andere waarden en het residu E te beschouwen als bijkomende predictievariabele. Vervolgens wordt de regressie uitgevoerd van $\sqrt{N}Y$ naar \sqrt{N} , $\sqrt{N}X$ en E. Deze regressie wordt uitgevoerd voor opeenvolgende uitbreidingen van het predictieinterval (X_U , X_L), waarbij in eerste instantie minimaal 20 discretizatie-intervallen worden gebruikt, tot de regressiecoëfficiënt geassocieerd met E als significant wordt bevonden (bij een significantieniveau van 10%). Wanneer dit gebeurt wordt de laatst toegevoegde predictiewaarde beschouwd als een potentieel breekpunt en wordt de procedure herhaald, nu startend vanaf dit laatste breekpunt en zonder de voorgaande waarden te beschouwen. Deze procedure wordt herhaald tot hetzij het volledig bereik van de predictiewaarden is beschouwd of maximaal 15 breekpunten zijn geïdentificeerd.

5.2.2. Fit van de stuksgewijze lineaire regressielijn

Gebruikmakend van voorgaande (voorlopige) lijst van breekpunten wordt vervolgens de stuksgewijze lineaire regressielijn gefit. Dit gebeurt opnieuw gebruikmakend van de pseudo-variaten: naast het gemeenschappelijk intercept \sqrt{N} en de pseudo-variant $\sqrt{N}X$ wordt voor het k'de segment de variaten $\sqrt{N}(X-X_j)$ voor $j < k$ toegevoegd. Vervolgens wordt de significantie van de hellingverandering in ieder segment nagegaan en de minst significante schatting wordt verwijderd, waarna de procedure wordt herhaald met het overeenkomstig breekpunt verwijderd. Wanneer alle geschatte termen significant zijn (bij een significantieniveau van 10%) wordt de procedure stopgezet.

5.2.3. Fit van de standaarddeviatie

Voor de schatting van de standaarddeviatie wordt uitgegaan van de interkwartielafstand UQ-LQ die eerder is bepaald in de exploratieve analyse. Voor een normaal verdeelde toevalsvariable komt de standaarddeviatie overeen met $(UQ-LQ)/1,35$ hetgeen toelaat om voor ieder predictieinterval een eerste niet-parametrische schatting van σ te maken. Zulk een schatting is echter enkel betrouwbaar wanneer het aantal gegevens N voor het predictieinterval voldoende groot is en de discretizatie die is gebruikt voor de extractie niet tot een nul-standaarddeviatie leidt (omdat alle gegevens binnen eenzelfde interval vallen). Daarom wordt voor $N < 10$ of een σ -schatting gelijk aan 0, volgende alternatieve schatter van σ gebruikt:

$$\sigma = \sqrt{N} \sqrt{\frac{2}{\pi}} |MQ - MQ_{fit}| \quad (6)$$

Deze benaderende schatter veronderstelt dat MQ_{fit} , de mediaanwaarde bepaald op basis van de stuksgewijze lineaire regressie, de exacte waarde voorstelt en het statistisch resultaat betreffende de standaardfout van de empirische mediaanwaarde MQ bepaald op basis van N meetwaarden.

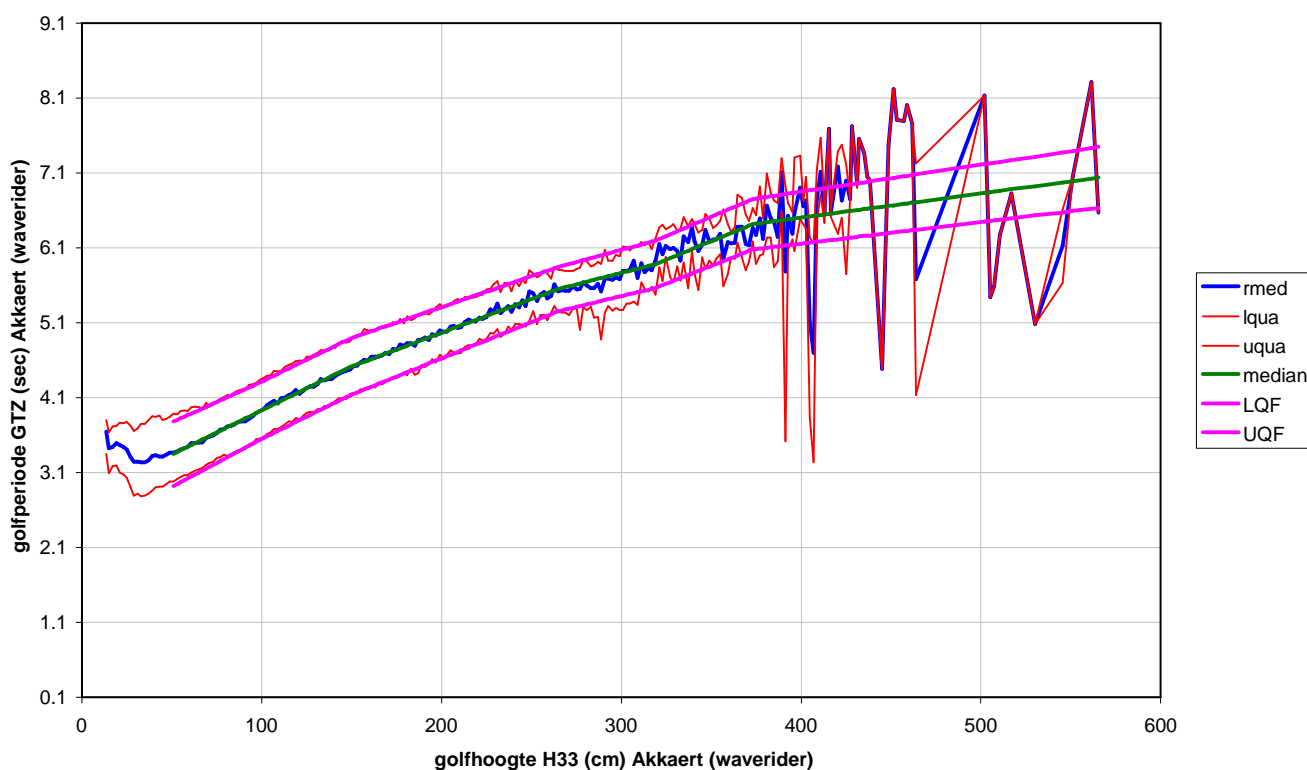
Startend met de breekpunten eerder bepaald voor de fit van de mediaanwaarden, wordt vervolgens een stuksgewijze lineaire regressiefit uitgevoerd van σ in functie van de predictiewaarden, waarbij opnieuw niet significante breekpunten (bij een significantieniveau van 10%) worden geëlimineerd.

5.2.4. Voorbeelden van de regressie-analyse

De volgende figuur toont de toepassing van voorgaande regressie-analyse op de simultane metingen van GTZ in functie van H33 ter plaatse van Akkaert. Zowel de empirische kwantielwaarden als de gefitte mediaanwaarde en de kwantielwaarden die worden afgeleid op basis van de geschatte standaarddeviatie ($LQ=MQ-0,65\sigma$, $UQ=MQ+0,65\sigma$) worden getoond.

De fit van de mediaan toont een relatief groot aantal breekpunten hetgeen aanduidt dat de relatie significant niet-lineair is. De standaarddeviatie daarentegen toont een meer gelijkmatig verloop hoewel ook in dit geval verschillende breekpunten worden gevonden (hoewel deze breekpunten statistisch significant zijn, tonen de resultaten duidelijk dat de waarde van σ quasi-constant is en overeenkomt met ongeveer 0.7 sec

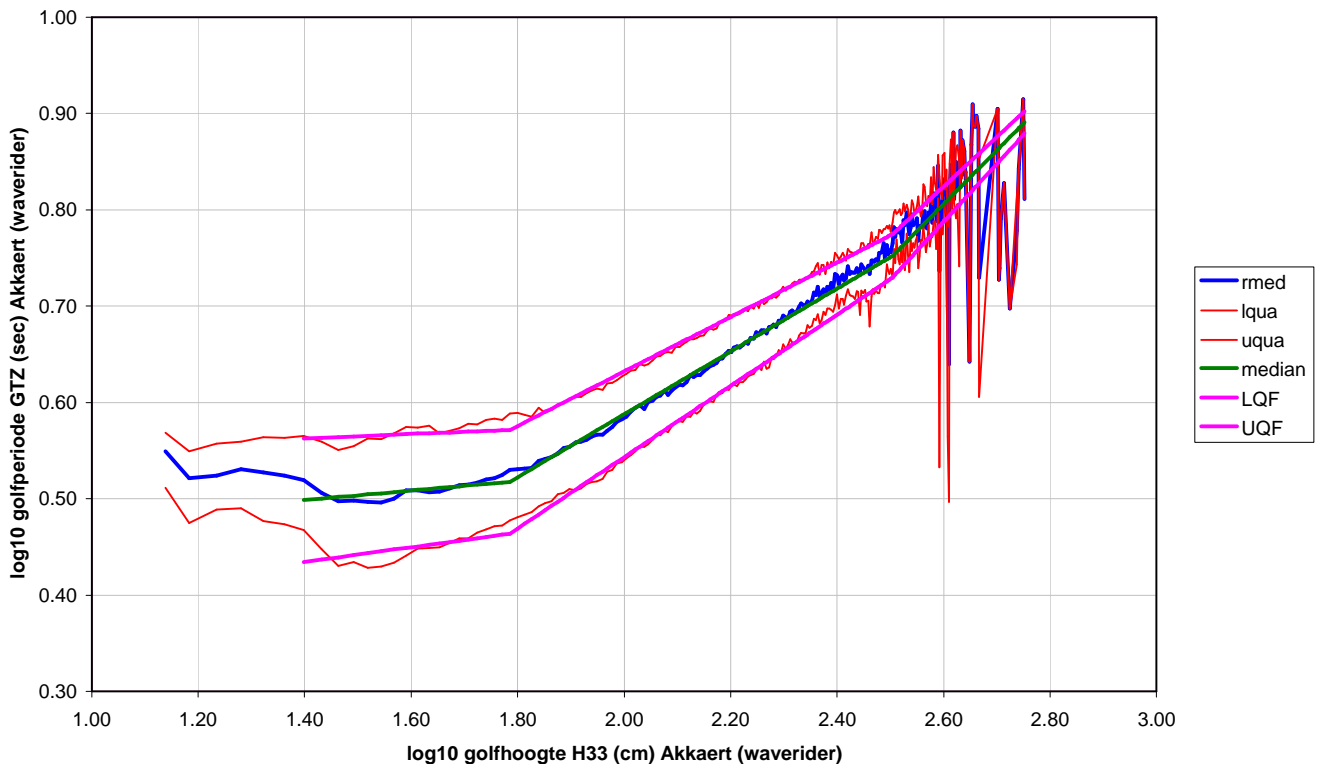
**Gewogen regressie-fit van een lineaire spline
golfhoogte H33 Akkaert (waverider) versus golfperiode GTZ Akkaert (waverider)**



(als onderdeel van het resultaat, worden ook de numerische resultaten bepaald die eveneens in de meteoatlas kunnen worden geraadpleegd).

Ter vergelijking wordt hierna de analyse getoond na logtransformatie van de parameters.

**Gewogen regressie-fit van een lineaire spline
golfhoogte H33 Akkaert (waverider) versus golfperiode GTZ Akkaert (waverider)**



Voor de mediaanwaarde leidt dit duidelijk tot een beter resultaat. De resultaten voor het laatste segment, bij de hogere waarden van H33, is duidelijk minder betrouwbaar en gebaseerd op slechts enkele waarnemingen. Bij de lagere waarden wordt een duidelijk niet-lineair verband vastgesteld. Voor het middenstuk vindt men:

$$\text{mediaan} (\log_{10} \text{GTZ}) = -0,06 + 0.33 \log_{10} \text{H33}$$

of equivalent

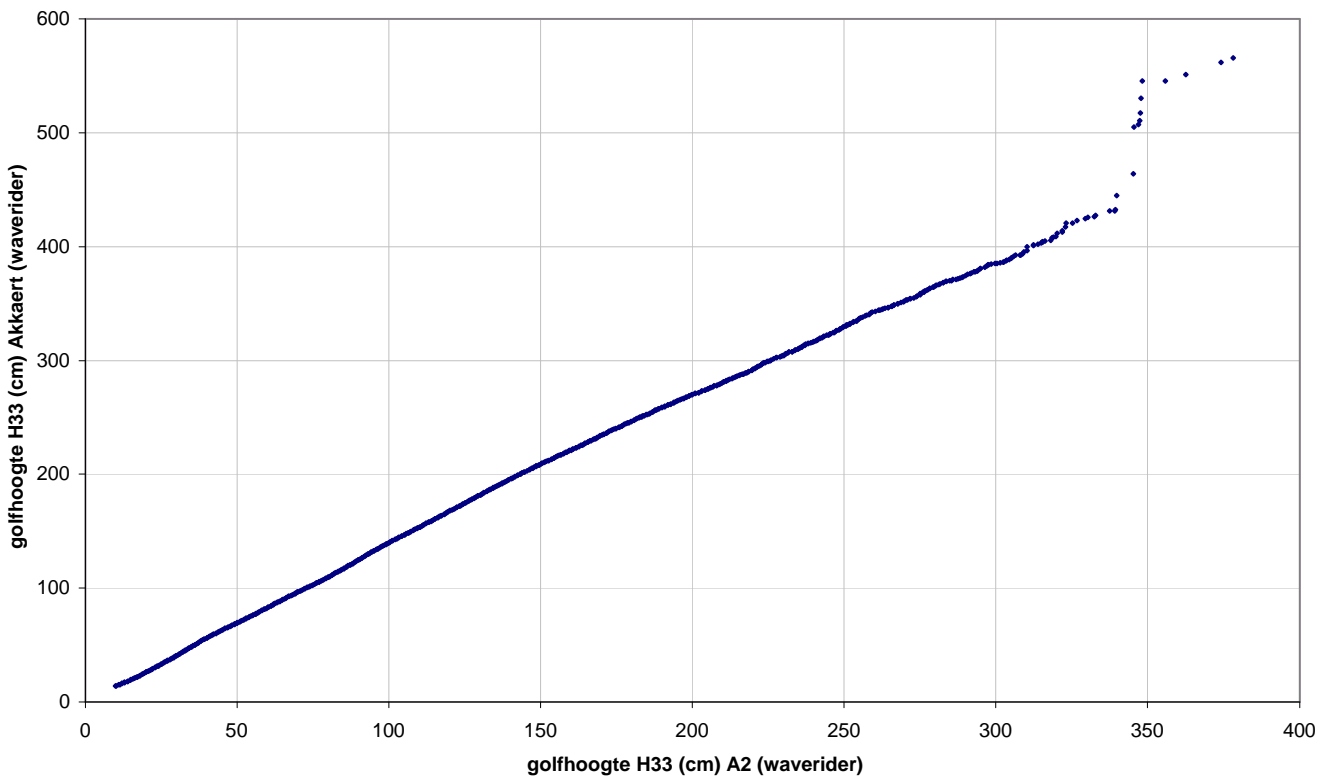
$$\text{mediaan} (\text{GTZ}) = 0,87 \text{H33}^{0.33}$$

5.2.5. EMPIRISCHE KWANTIEL-KWANTIEL PLOT

Voor de voorstelling van het verband tussen de orde-statistieken wordt geopteerd voor de voorstelling van de kwantielwaarde (hetgeen overeenkomt met de orde-statistiek) van één variabele ten overstaan van de kwantielwaarde van de andere variabele. De volgende figuur toont een voorbeeld van zulk een resultaat voor de H33 metingen ter plaatse van de A2 -waveriderboei en de Akkaert-Waveriderboei.

De kwantiel-kwantielplot maakt een rechtstreekse vergelijking mogelijk van de statistische beschrijving van de zeeklimaatparameter ter plaatse van de twee locaties voor eenzelfde meetperiode. Een analoog resultaat zou worden bekomen door de kwantielwaarden die gefit zijn voor de marginale verdelingen uit te zetten ten opzichte van elkaar. De marginale verdelingen worden echter bepaald op basis van alle meetwaarden ter plaatse van iedere lokatie en zijn niet representatief voor exact dezelfde meetperiode. Een dergelijke figuur kan bijvoorbeeld gebruikt worden om een ontwerpwaarde ter plaatse van 1 lokatie om te rekenen naar een ontwerpwaarde van een andere lokatie.

**Kwantiel-kwantielplot
golfhoogte H33 A2 (waverider) versus golfhoogte H33 Akkaert (waverider)**



Bij de beoordeling van de plot dient men rekening te houden dat de hoogste kwantielwaarden slechts enkele waarnemingen weergeven en men dus typisch een grotere variatie mag verwachten.

De kwantiel-kwantielplot wordt niet verder analytisch behandeld ten eerste omdat zulk een analyse weinig nuttig is (het parametrisch verband tussen de kwantielen kan beter worden afgeleid door middel van de schattingen van de marginale kansverdelingen), maar ook omdat zulk een analyse niet triviaal is. De verschillende orde-statistieken zijn sterk afhankelijk en bovendien kan men geen eenvoudige regressie-analyse toepassen, vermits dit een verschillend resultaat zou opleveren naargelang de regressie wordt uitgevoerd van X naar Y of van Y naar X. Bovendien merkt men dat, althans voor het ganze bereik van de gegevens, de relatie niet-lineair is. Er is nagegaan dat dit ook het geval is na log-transformatie.

De kwantiel-kwantielplot moet daarom beschouwd worden als een eenvoudige rechtstreekse weergave van de aanpassing van de statistische beschrijving van de gegevens van één lokatie naar een andere lokatie. Zulke informatie is nuttig, bijvoorbeeld voor de calibratie van mathematische golfmodellen.

6. SAMENVATTING

Deze studie betreft het opzetten van zinvolle voorstellingswijzen en analysetechnieken om de relatie tussen verschillende hydrometeorologische parameters weer te geven. Volgende 3 voorstellingswijzen blijken aangewezen:

- voor de onderlinge vergelijking van de simultane metingen van wind-en/of golf-richtingen blijkt enkel een voorstelling van de bivariate kansdichtheid zinvol te zijn. Enkel indien deze bivariate kansdichtheid unimodaal is (hetgeen zelden voorkomt) is een regressie-analyse zinvol. Indien dit niet het geval is, dan zou zulk een regressie-analyse overeenkomen met het gemiddelde van twee of meer afzonderlijke relaties en een totaal fout beeld geven van de werkelijke relatie;
- voor de vergelijking van de simultane meetwaarden van de overige parameters (significante golfhoogte metingen, golfperiode karakteristieken, spectrum karakteristieken en windsnelheid) is de toepassing van een regressie-analyse mogelijk. Bij deze toepassing dient echter rekening te worden gehouden met de aanwezigheid van uitbijters (die frequent blijken voor te komen), de non-lineariteit van de relatie en de variatie van de spreiding rond de regressie-lijn. Hiertoe is een gewogen regressie-analyse methode ontwikkeld die wordt toegepast op samenvattende quantielwaarden (om de invloed van uitbijters te vermijden) en die een stuksgewijze lineaire relatie tussen de twee parameters schat (om de non-lineariteit correct weer te geven). De segmenten waarover lineariteit mag worden verondersteld worden automatisch bepaald in deze procedure en ook de variatie van de spreiding wordt weergegeven. De methode wordt toegepast zowel voor de metingen in hun oorspronkelijke schaal en na log-transformatie. Typisch wordt de meest eenduidige relatie (d.w.z. met het minste aantal segmenten en met de kleinste variatie van de spreiding) gevonden na de logtransformatie. De toegepaste methode wordt in detail beschreven in Hoofdstuk 4;
- de onderlinge relatie tussen de orde-statistieken van de verschillende parameters wordt op directe wijze voorgesteld door de orde-statistieken in functie van elkaar uit te zetten. De interpretatie en het eventueel gebruik van deze curven wordt uiteengezet in Hoofdstuk 5.