



# Een grondige statistische studie van het Belgische kustklimaat

Elke Debrie

Promotoren:

Prof. dr. Els Goetghebeur

Prof. dr. David Dehenauw

MASTERPROEF VOORGELEGD TER BEHALING VAN DE  
ACADEMISCHE GRAAD VAN MASTER OF SCIENCE IN DE WISKUNDE,  
AFSTUDEERRICHTING TOEGEPASTE WISKUNDE

Academiejaar 2016-2017

# Dankwoord

Tijdens het schrijven van deze thesis heb ik veel bijgeleerd op vlak van statistiek en wetenschappelijk onderzoek. Ik ben er mij van bewust geworden dat er bij een statistische studie veel mogelijke onderzoekswegen zijn. In de voorbije maanden is mij eveneens duidelijk geworden dat het tot stand brengen van deze thesis niet mogelijk was geweest zonder de hulp van heel wat mensen.

Vooreerst wil ik mijn promotoren prof. Els Goetghebeur en prof. David Dehenauw bedanken voor de begeleiding op wiskundig en meteorologisch vlak. Zij waren steeds bereid om samen te zitten en mij verder op weg te helpen. Hun kennis en deskundige raadgevingen waren onontbeerlijk en hebben de fundamenteen gevormd van deze thesis. Ik wil hen bedanken voor hun bereidheid tot advies, zelfs over de presentatie.

Ik bedank het KMI en in het bijzonder Alex Dewalque voor het ter beschikking stellen van de zeesuurlijkse data in Koksijde en Middelkerke en Belgocontrol voor de uurlijkse data in Middelkerke. Verder wil ik Vicky Schroeders van de Meteorologische Wing bedanken voor de tijd die zij heeft genomen om een gedetailleerde beschrijving van de geschiedenis van het Meteo Station in Koksijde te geven en foto's van de anemometers te verzamelen.

Voor deze thesis gebruikte ik R, een programmeertaal en softwareomgeving voor statistische analyse, alsook de grafische userinterface RStudio, beiden open source. Ik ben de community van vrijwillige bijdragers achter RStudio dan ook erkentelijk voor het ontwikkelen van dit softwarepakket.

Tot slot, maar daarom zeker niet minder belangrijk, bedank ik mijn ouders en mijn vriend Bert. Zonder hun onvoorwaardelijke steun de voorbije vijf jaar zou ik niet staan waar ik nu sta.

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

Gent, 1 juni 2017  
Elke Debrie

# Inhoudsopgave

<b>1</b>	<b>Inleiding</b>	<b>8</b>
1.1	Inhoudelijke situering . . . . .	8
1.2	Klimaatanalyse in België . . . . .	9
1.3	Onderzoeksvragen . . . . .	9
1.4	Toepassingen . . . . .	9
1.4.1	Klimaatbeleid informeren . . . . .	9
1.4.2	Energiesector . . . . .	10
1.4.3	Hydrologische kennis . . . . .	11
1.4.4	Ventileren van luchtvervuiling . . . . .	11
1.4.5	Natuurbeheer en volksgezondheid . . . . .	11
1.4.6	Corrigeren van historische klimatologische gegevens . . . . .	11
<b>2</b>	<b>Gegevens over het Belgische kustweer</b>	<b>12</b>
2.1	Meetstations . . . . .	12
2.2	De variabelen: interpretatie en betekenis . . . . .	14
2.2.1	Registratiemomenten . . . . .	14
2.2.2	Windsnelheid en windstoten . . . . .	15
2.2.3	Windrichting . . . . .	15
2.2.4	Bewolgingsgraad . . . . .	16
2.2.5	Temperatuur . . . . .	17
2.2.6	Neerslag . . . . .	17
2.3	Hulpvariabelen . . . . .	17
2.3.1	Moment . . . . .	17
2.3.2	Seizoen en seizoenjaar . . . . .	17
2.3.3	Windrichting (hoek) . . . . .	18
2.3.4	Vershil Koksijde - Middelkerke . . . . .	18
2.3.5	Variabiliteit tussen opeenvolgende metingen . . . . .	19
2.3.6	Orthogonale componenten van de windsnelheid . . . . .	19

<b>3</b>	<b>Clean-up van de dataset</b>	<b>20</b>
3.1	Ontbrekende meetmomenten . . . . .	20
3.2	Windsnelheden . . . . .	20
3.2.1	Spectrum van gemeten waarden . . . . .	21
3.2.2	Ontbrekende waarden . . . . .	22
3.2.3	Foutieve outliers . . . . .	23
3.3	Temperatuur . . . . .	30
3.3.1	Spectrum van gemeten waarden . . . . .	30
3.3.2	Ontbrekende waarden . . . . .	30
3.3.3	Foutieve outliers . . . . .	30
3.4	Neerslag . . . . .	32
3.4.1	Spectrum van gemeten waarden . . . . .	32
3.4.2	Ontbrekende waarden . . . . .	32
3.4.3	Foutieve outliers . . . . .	33
<b>4</b>	<b>Verdeling van de windsnelheden</b>	<b>35</b>
4.1	Rayleighdistributie . . . . .	36
4.1.1	Definitie . . . . .	36
4.1.2	Wiskundige analyse van cumulatieve en kansdichtheidsfunctie . . . . .	37
4.1.3	Geschikt voor modellering van windsnelheden? . . . . .	39
4.2	Weibullverdeling . . . . .	41
4.2.1	Definitie . . . . .	42
4.2.2	Wiskundige analyse van de cumulatieve en kansdichtheidsfunctie . . . . .	42
4.3	Schatten van de vorm- en schaalparameter . . . . .	45
4.3.1	De momentmethode . . . . .	45
4.3.2	De maximum-likelihoodmethode . . . . .	45
4.3.3	De power-densitymethode . . . . .	47
4.4	Schatten van de Weibullparameters voor de windsnelheden in Koksijde en Middelkerke . . . . .	47
4.4.1	Schatting met de momentmethode . . . . .	47
4.4.2	Schatting met de maximum-likelihoodmethode . . . . .	48
4.4.3	Schatting met de power-densitymethode . . . . .	51
4.4.4	Overzicht van de resultaten . . . . .	51
4.5	Geschikt voor modellering van windsnelheden? . . . . .	51
4.6	Vergelijking . . . . .	55
4.6.1	Koksijde versus Middelkerke . . . . .	55
4.6.2	In de winter waait het meer . . . . .	55
4.7	Verdeling van rukwinden . . . . .	55

<b>5</b>	<b>Testen en methoden voor trendanalyse</b>	<b>58</b>
5.1	LOWESS . . . . .	58
5.2	Lineaire regressie . . . . .	59
5.3	Polynomiale regressie . . . . .	59
5.4	Niet-parametrische lineaire regressie . . . . .	59
5.5	Mann-Kendalltest en Sen's slope . . . . .	60
5.6	Bonferronicorrectie . . . . .	63
5.7	Vergelijking van de power . . . . .	65
5.7.1	Benchmarking: schatting kans op type-I-fout . . . . .	65
5.7.2	Testen van de power voor een verschil van 0,01 m/s per jaar . . . . .	66
5.8	Meer robuuste methoden . . . . .	66
<b>6</b>	<b>Trends in de windsnelheden</b>	<b>68</b>
6.1	Boxplots voor de windsnelheden in de verschillende seizoenen . . . . .	68
6.2	Jaar- en seizoensgemiddelden . . . . .	69
6.2.1	LOWESS . . . . .	70
6.2.2	Lineaire regressie . . . . .	72
6.2.3	Polynomiale regressie . . . . .	74
6.2.4	Niet-parametrische lineaire regressie . . . . .	75
6.2.5	Mann-Kendalltest en Sen's slope . . . . .	75
6.2.6	Samenvattende tabel en conclusies . . . . .	77
6.3	Maandgemiddelde windsnelheden . . . . .	77
6.4	Dagen met krachtige wind . . . . .	82
6.5	Dagen met zwakke tot matige wind . . . . .	85
6.6	Rukwinden . . . . .	86
6.7	Trendanalyse op de verdelingsparameters . . . . .	86
6.7.1	Visuele inspectie . . . . .	86
6.7.2	Analyse per jaar . . . . .	88
6.7.3	Analyse per seizoen . . . . .	88
6.8	Discussie . . . . .	88
<b>7</b>	<b>Kusttemperatuur</b>	<b>92</b>
7.1	Beschrijvende analyse . . . . .	92
7.1.1	Verdeling van de temperatuur . . . . .	92
7.2	Trendanalyse op de temperatuur . . . . .	93
7.2.1	Lineaire regressie . . . . .	94
7.2.2	Niet-parametrische lineaire regressie . . . . .	94
7.2.3	Mann-Kendalltest en Sen's slope . . . . .	96
7.3	Conclusie . . . . .	96

<b>8 De zeebries</b>	<b>99</b>
8.1 Beschrijvende analyse van windrichtingen . . . . .	99
8.2 Wat is de zeebries? . . . . .	100
8.3 De relevantie van trends in de zeebries . . . . .	102
8.4 Zeebries identificeren in de dataset: zeebriesdagen . . . . .	102
8.5 Resultaat van de analyse . . . . .	106
8.5.1 Aantal zeebriesdagen . . . . .	106
8.5.2 Gemiddelde sterkte van de zeebries . . . . .	107
8.6 Conclusie . . . . .	107
<b>9 Conclusies</b>	<b>109</b>
9.1 Samenvatting van het onderzoek . . . . .	109
9.2 Vergelijking met binnen- en buitenland . . . . .	110
9.2.1 Windsnelheid . . . . .	110
9.2.2 Rukwinden en stormdagen . . . . .	112
9.2.3 Temperatuur . . . . .	112
9.2.4 Zeebries . . . . .	113
9.3 Mogelijkheden voor verder onderzoek . . . . .	113
9.3.1 Homogenisering . . . . .	113
9.3.2 Koppelen van data voor analyse van de zeebries . . . . .	114
9.3.3 Variantie-evolutie van de windsnelheid . . . . .	114
9.3.4 Trendanalyse neerslag . . . . .	114
9.3.5 Complexere modellen . . . . .	114
<b>Bijlagen</b>	<b>115</b>
<b>Bijlage A English summary</b>	<b>115</b>
<b>Bijlage B Tabellen met meetmomenten</b>	<b>117</b>
<b>Bijlage C R-code</b>	<b>121</b>
C.1 Inleiding . . . . .	121
C.2 Gegevens over wind aan de Belgische kust . . . . .	121
C.2.3 Hulpvariabelen . . . . .	121
C.3 Clean-up van de dataset . . . . .	122
C.3.1 Ontbrekende meetmomenten . . . . .	122
C.3.2 Windsnelheden . . . . .	123
C.3.3 Temperatuur . . . . .	123

---

C.3.4	Neerslag . . . . .	124
C.4	Verdeling van de windsnelheden in Koksijde en Middelkerke . . . . .	124
C.4.1	Rayleighdistributie . . . . .	124
C.4.2	Weibullverdeling . . . . .	125
C.4.4	Schatten van de Weibullparameters voor de windsnelheden in Koksijde en Middelkerke . . . . .	126
C.4.5	Geschikt voor de modellering van windsnelheden? . . . . .	128
C.4.7	Verdeling van rukwinden . . . . .	128
C.5	Testen en methoden voor trendanalyse . . . . .	128
C.5.7	Vergelijking van de power . . . . .	128
C.6	Trends in de windsnelheden . . . . .	130
C.6.1	Boxplots voor de windsnelheden in de verschillende seizoenen . . . . .	130
C.6.2	Jaar- en seizoensgemiddelden . . . . .	130
C.6.3	Maandgemiddelde windsnelheden . . . . .	131
C.6.4	Dagen met krachtige wind . . . . .	131
C.6.5	Dagen met zwakke tot matige wind . . . . .	132
C.6.7	Trendanalyse op de verdelingsparameters . . . . .	132
C.6.8	Discussie . . . . .	133
C.7	Kusttemperatuur . . . . .	134
C.7.1	Beschrijvende analyse . . . . .	134
C.7.2	Trendanalyse op de temperatuur . . . . .	134
C.8	De zeebries . . . . .	134
C.8.1	Beschrijvende analyse van windrichtingen . . . . .	134
C.8.4	Zeebries identificeren in de dataset: zeebriesdagen . . . . .	134
C.8.5	Resultaat van de analyse . . . . .	135
C.9	Conclusies . . . . .	135

# Hoofdstuk 1

## Inleiding

In deze thesis zullen een aantal klimaatvariabelen aan de Belgische kust, meer bepaald in Koksijde en Middekerke, bestudeerd worden in de periode 1952 tot 2016. In een eerste hoofdstuk zal een overzicht gegeven worden van de beschikbare variabelen in de dataset, evenals hulpvariabelen die nodig bleken te zijn tijdens de statistische analyses. Daaropvolgend zal een clean-up van de dataset gedaan worden waarbij foutieve outliers gemotiveerd uit de dataset worden verwijderd. Vervolgens zal onderzocht worden welke kansverdeling de windsnelheden aan de Belgische kust het best kan beschrijven. In het hoofdstuk over trendanalyse zullen een aantal methoden besproken worden die geschikt zijn om trends in klimatologische tijdsreeksen te detecteren. Variabelen die onderzocht zullen worden, zijn jaar-, maand- en seizoensgemiddelde windsnelheden en jaar- en seizoensgemiddelde temperaturen. Eveneens zal aandacht besteed worden aan het aantal dagen met krachtige en zwakke tot matige wind. Tot slot zullen de vaststellingen die gedaan werden, vergeleken worden met analyses die eerder werden uitgevoerd op de klimatologische tijdsreeksen in Ukkel, waarvan de resultaten beschreven staan in het klimaatrapport van het KMI [8] en het MIRA [7].

### 1.1 Inhoudelijke situering

De laatste jaren bereiken onheilspellende berichten in verband met de klimaatopwarming ons steeds vaker o.a. via de media. De Redactie titelde in anderhalve maand o.a.

- *Reusachtige ijsberg dreigt af te scheuren van ijsplaat op Antarctica* (06/01/2017)
- *‘Klimaatverandering heeft grote gevolgen voor Europa’* (25/01/2017)
- *Klimaatrapport: ‘Droogte rond Middellandse Zee wordt zeer problematisch’* (25/01/2017)
- *Een superstorm die nu één keer om de 15 jaar plaatsvindt, zal jaarlijkse kost worden* (28/01/2017)
- *Opwarming aarde verandert leven in de bergen fundamenteel* (08/02/2017)
- *Opwarmende oceanen zijn twee procent zuurstof verloren* (16/02/2017)

De frequentie waarmee deze opduiken, zegt genoeg. De klimaatopwarming baart de mens zorgen en daarom gebeurt er veel onderzoek naar factoren die het klimaat beïnvloeden, de mechanismen waarlangs dit gebeurt en de manieren om deze opwarming tegen te gaan. Precies om een duidelijk en correct beeld te krijgen van het veranderende klimaat is de studie van klimatologische tijdsreeksen van groot belang.



## 1.2 Klimaatanalyse in België

Bijna alle wetenschappers zijn het erover eens dat een – mogelijks ernstige – klimaatverandering zijn intrede heeft gemaakt. Dit wordt almaar zichtbaarder door een aantal indicatoren, ook in België. Er werd bijvoorbeeld vastgesteld dat de gemiddelde temperatuur in de vier seizoenen gestegen is, met de lente op kop. De jaargemiddelde temperatuur in Ukkel ligt intussen ook al 2.4°C hoger dan in de tweede helft van de 18<sup>de</sup> en de eerste helft van de 19<sup>de</sup> eeuw. Ook de neerslaghoeveelheid in Ukkel vertoont een langzaam stijgende trend. De gemiddelde windsnelheid per jaar bleef tot ongeveer de jaren 1960 vrij stabiel, maar sindsdien is een daling ingezet. Het huidige jaargemiddelde blijkt ongeveer 10 à 15% lager te liggen dan pakweg 30 jaar geleden<sup>1</sup> [7].

Uit de klimaatrapporten van MIRA en KMI blijkt dat al vrij veel onderzoek gedaan werd naar de klimaatverandering in België, meer bepaald in het binnenland. Datareeksen van meetstations worden geanalyseerd om onder andere vragen te kunnen beantwoorden zoals: ‘Stijgt de gemiddelde temperatuur?’, ‘Zijn er meer hittegolven/koudegolven?’, ‘Regent het meer of minder?’, ‘Zijn er meer stormen?’,...

De conclusies hebben hierbij echter enkel betrekking op het binnenland. Analyses van klimatologische reeksen aan de Belgische kust zijn nog niet gedaan, of worden toch niet expliciet vermeld. Zouden de resultaten overeenkomen met deze in het binnenland of speelt het microklimaat aan de kust in die mate een rol dat de vaststellingen niet dezelfde zijn als in het binnenland? Waarnemingen toonden bijvoorbeeld al aan dat de kust koelere zomers heeft en zachtere winters, dat er minder neerslag is en meer wind [12]. Maar men kan op sommige dagen ook gewoon vaststellen dat het aan de kust mooi en zonnig is, terwijl het 5 km verderop in het binnenland volkomen bewolkt is.

In deze thesis zal daarom het Belgische kustklimaat, meer bepaald dat van Koksijde en Middelkerke, van naderbij bekeken worden en zullen we proberen een antwoord te vinden op een aantal vragen. Data zijn echter pas beschikbaar sinds de jaren 1950. Wanneer het klimaat bestudeerd wordt, worden uitspraken gedaan over een lange periode. De analyse die voor de kust wordt gedaan, zal bijgevolg eerder beperkt zijn.

## 1.3 Onderzoeksvragen

Gebruik makend van statistische methoden en op basis van meteorologische kennis zullen volgende zaken bestudeerd worden:

1. Welke verdeling volgen de gemeten windsnelheden aan de Belgische kust? Is er een trend in de tijd?
2. Vertonen de jaar- en seizoensgemiddelde temperaturen aan de kust een trend?
3. Is de zeebries de afgelopen 70 jaar toegenomen? Is er een verband tussen de stijging in temperatuur en de sterkte van de zeebries?

## 1.4 Toepassingen

Hieronder worden een aantal toepassingen gegeven waarvoor onderzoek van klimaatvariabelen, zoals gedaan in deze masterthesis, nuttig zouden kunnen zijn.

### 1.4.1 Klimaatbeleid informeren

Klimaatverandering moet zowel lokaal, regionaal en continentaal in kaart gebracht worden om klimaatbeleid te informeren. Aan de hand van deze grondige studies kunnen instellingen zoals het IPCC aanbevelingen

---

<sup>1</sup>Testen die gebruikt werden om het gedrag van klimatologische tijdsreeksen te bepalen, waren de test van homogeniteit van Pettitt en de Mann-Kendalltest voor trendanalyse.

schrijven en monitoren, kan de EU klimaatmaatregelen definiëren als prioriteit en deze maatregelen kunnen dan door de lidstaten omgezet worden in beleid. Dit beleid situeert zich vooral in de domeinen energie, wonen, onderwijs (sensibilisering) en economie en dit beleid kan concreet worden in nieuwe wetten en regelgeving, belastingen, subsidiëring. Er is dus nood aan onderzoek naar het klimaat dat informatie vergaart en dat de klimaatsbeleidscyclus kan bijstaan en informeren.

## 1.4.2 Energiesector

Het belang van de inzet van windenergie nam toe naarmate duidelijk werd dat gebruik van fossiele brandstoffen bijdraagt aan de klimaatopwarming. Windenergie is bovendien de meest voorkomende en snelst groeiende energietechnologie [26]. Bij deze vorm van energieopwekking is niet alleen het rendement van belang, maar ook het hernieuwbaar karakter [63]. Door gebruik te maken van deze vorm van energie kan een groot deel van de  $CO_2$ -uitstoot teruggedrongen worden. Een grote hoeveelheid  $CO_2$  versterkt het broeikaseffect wat de energiebalans van de aarde verstoort. Dit werd nogmaals benadrukt op de klimaatconferentie in Parijs in 2015. De bouw van windmolenparken, die voorzien in duurzame en groene energie, zorgt dus mogelijks voor een belangrijk onderdeel van energietransitie. Op die manier heeft de inzet op windenergie een gunstige bijdrage met oog op klimaatverandering. Met als doel om de uitstoot van broeikasgassen te verminderen, werd door regeringen vanuit het Kyoto klimaatverdrag de bouw van windmolenparken dan ook gestimuleerd [61].

Een grondige en correcte windtrendanalyse is voor de windenergiesector dan ook van groot belang. Windmolens beginnen te draaien vanaf 3,5 m/s (12,6 km/h) en ze worden stilgelegd wanneer de windsnelheden 25 m/s (90 km/h) overschrijden [66]. Een analyse van de windsnelheid kan aangeven hoe het aantal dagen waarop de windsnelheid onder 2 tot 3 Beaufort blijft of 10 Beaufort overschrijdt, verwacht wordt te zullen veranderen. Ook op dagen waarop de windmolens draaien, zou de opgewekte energie aanzienlijk kunnen dalen als de windsnelheden terugvallen aangezien de windenergie varieert met het kwadraat van de snelheid [17, 28]. Wanneer gekozen wordt voor windenergie, is het dus van belang een goed zicht te hebben op de voorspelbaarheid van het windaanbod [63].

Een probleem dat hierbij opduikt, is het feit dat windsnelheden waarmee analyses worden gedaan over het algemeen gemeten worden op 10 m hoogte. Belgische windmolenparken in de Noordzee (anno 2017 Northwind, Belwind, C-Power en meer gepland of in aanbouw [13]) daarentegen hebben windmolens van 40 m tot 100 m hoog en bijgevolg zou een analyse van de wind op grote hoogte accurater zijn voor wat betreft het windmolenbeheer.

Volgens Peterson et al. [45] is het gerechtvaardigd om de windsnelheden op grote hoogte te extrapoleren uit de windsnelheden op 10 m, volgens de zogenaamde power law [45, 46]:

$$\frac{v(z_1)}{v(z_2)} = \left(\frac{z_1}{z_2}\right)^m. \quad (1.1)$$

Hierbij is  $v(z)$  de windsnelheid in m/s op hoogte  $z$  en  $m$  is een locatieafhankelijke exponent die afhangt van de ‘oppervlakteruwheid’ en de stabiliteit van de atmosfeer. Bij een stabiele atmosfeer bijvoorbeeld neemt  $m$  de waarde 0,143 aan [38]. Aan de hand van de power law is het dus mogelijk om de windsnelheden op een willekeurige hoogte – hier op hoogte van de windmolens – te extrapoleren [50, 63, 38]. Deze uitdrukking blijkt een vrij goede benadering te kunnen geven van de gemiddelde windsnelheid op de gewenste hoogte onder neutrale omstandigheden [62]. In [53] wordt een meer uitgebreide power law uitgewerkt en toegepast op windsnelheden in Istanbul.

Wanneer men op die manier beschikt over geschatte windsnelheden op de gewenste hoogte, kan men de vermogensdichtheid per eenheidsoppervlak schatten, aan de hand van de geschatte Weibullparameters (zie Hoofdstuk 4) op die zekere hoogte. Op die manier kan dus ook aan de hand van de trend van de Weibullparameters geschat worden hoeveel energie per jaar gemiddeld meer of minder geproduceerd wordt.

### 1.4.3 Hydrologische kennis

Een ander gevolg voor de verandering in windsnelheden is te vinden in het hydroklimaatstelsel. In een onderzoek van onder andere [37] werd een verband gevonden tussen de veranderingen in windsnelheden en verdampingssnelheden. Echter, veranderingen in verdampingssnelheden worden veroorzaakt door interactie tussen verschillende factoren zoals veranderingen in omgevingstemperatuur, relatieve vochtigheid, inkomende zonnestraling ... [11]. Om te onderzoeken of dit verband ook aan de Belgische kust te vinden is, zouden lokale modellen ontwikkeld moeten worden die bovengenoemde variabelen in rekening brengen en beschrijven. Wanneer windsnelheden afnemen, kan een verminderd vochttransport wel een voordeel zijn voor ecosystemen in het geval ook de neerslaghoeveelheden gemiddeld afnemen.

### 1.4.4 Ventileren van luchtvervuiling

In verschillende studies kon reeds worden vastgesteld dat windsnelheden dalen [8, 17, 50]. Wanneer windsnelheden dalen, kan dit gevolgen hebben voor het ventileren van luchtvervuiling in stedelijke gebieden. Een voorbeeld van een vervuiler die in steden vaak voorkomt zijn de ‘vluchtige organische stoffen’. Deze zijn onder andere aanwezig in uitlaatgassen van gemotoriseerde voertuigen. Deze vluchtige organische stoffen kunnen door UV-straling omgezet worden in ozon. Deze ozon, die aanwezig is in de troposfeer en beter bekend is als smog, kan een schadelijke impact hebben op mensen, dieren en vegetatie [56]. Een goed beeld van de evolutie van windsnelheden is dus van belang om een idee te hebben over hoe luchtvervuilende deeltjes in steden en daarrond worden verspreid. De *Gaussian plume formula* beschrijft in welke mate luchtverontreinigende stoffen efficiënt worden verspreid. Eén van de parameters die hierin cruciaal is, is de horizontale windsnelheid [38].

### 1.4.5 Natuurbeheer en volksgezondheid

Het klimaat in een regio bepaalt welke dieren en planten er kunnen gedijen. Veranderingen in de klimatologische toestand van een regio kunnen een grote impact hebben op de lokale fauna en flora. Bijvoorbeeld trekvogels, waadvogels en insecten die voor de mens schadelijke ziektes kunnen overbrengen (zoals de aziatische tijgermug [10]), wat een aanzienlijke impact kan hebben op de volksgezondheid.

### 1.4.6 Corrigeren van historische klimatologische gegevens

Het KMI is sinds enkele jaren bezig met een clean-up van de datasets die verzameld werden door meetstations verspreid over heel België. Door de enorme omvang van de vele datasets konden correcties nog maar doorgevoerd worden tot ongeveer 50 jaar terug. Gegevens van voor 1970 werden, voor Koksijde en Middelkerke althans, nog niet nagekeken en gecorrigeerd. Met deze thesis werd alvast een steentje bijgedragen om de datasets van Koksijde en Middelkerke van verkeerde waarden te ontdoen.

## Hoofdstuk 2

# Gegevens over het Belgische kustweer

Dit werk is een analyse van twee meteorologische datasets, met name deze uit Koksijde (1952 - 2016) en deze uit Middelkerke (1955 - 2016). De gegevens werden verkregen van het KMI (Koninklijk Meteorologisch Instituut) in Ukkel. Een uurlijkse dataset van Middelkerke met windsnelheden en windrichtingen werden verkregen via Belgocontrol.

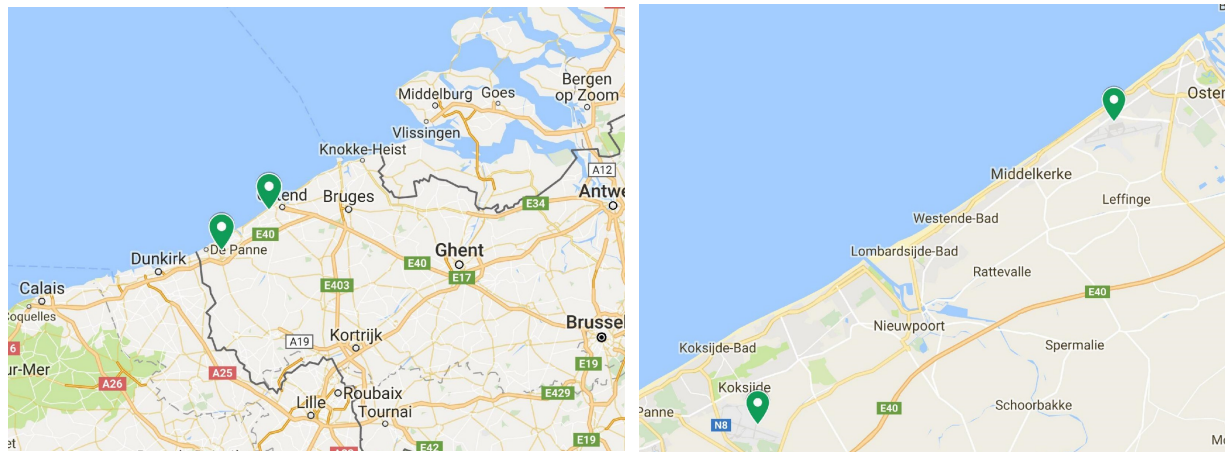
### 2.1 Meetstations

**Locatie** De data die in deze masterthesis worden geanalyseerd zijn geregistreerd in meetstations gelegen in Koksijde (luchtmachtbasis,  $51^{\circ}05'40.85''N, 02^{\circ}39'17.33''E$ ) en Middelkerke (luchthaven Oostende,  $51^{\circ}11'56''N, 02^{\circ}51'46''E$ ). De meetstations van Koksijde en Middelkerke liggen aan de kust, op 19,58 km van elkaar verwijderd. Het meetstation van Koksijde ligt 3,83 km verwijderd van de kustlijn, dat van Middelkerke ongeveer 1 km.

**Geschiedenis meetstation Koksijde** Het militair vliegveld van de Belgische luchtmacht bevindt zich in Koksijde. De eerste observaties uit Koksijde dateren van 1946. Hier zijn echter geen gegevens over beschikbaar in de dataset. Gegevens worden pas beschikbaar vanaf 1952, het moment waarop het burgerpersoneel het station verlaat en dit wordt overgenomen door gespecialiseerd personeel van de luchtmacht. Op 11 augustus 1958 werd het volledige meetstation naar een nieuwe plaats verhuisd, bovendien de plaats waar het meetstation tot op vandaag gevestigd is. Het volledige MeteoPark werd vernieuwd en onder andere een anemometer werd geïnstalleerd. In het begin van de jaren '60 verhuisde de anemometer naar zijn huidige locatie, net naast de landingsbaan. Vanaf de jaren '90 werden metingen niet langer enkel door waarnemers geregistreerd, maar werd de taak deels overgenomen door een computer verbonden met sensoren (persoonlijke communicatie met Vicky Schroeders per e-mail op 16 mei 2017). Sinds 20 april 2014 wordt de wind in Koksijde ultrasoon gemeten. Een ultrasone anemometer maakt gebruik van het principe dat de voortplantingssnelheid van geluid afhankelijk is van de windsnelheid. Bovendien kunnen met dit apparaat ook windrichtingen bepaald worden, evenals windstoten [60].



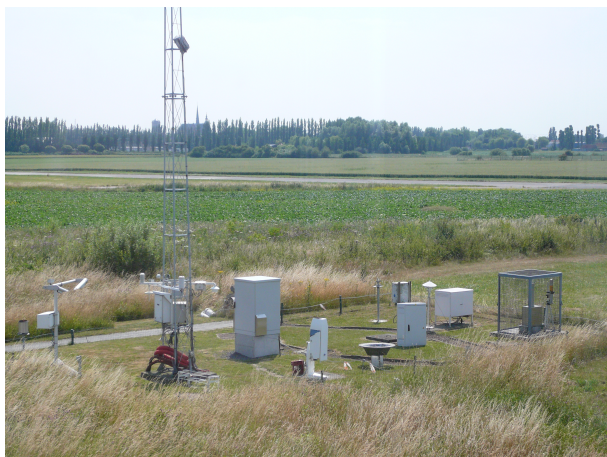
Figuur 2.1: Vroegere en huidige locatie van de anemometer op de luchtmachtbasis in Koksijde



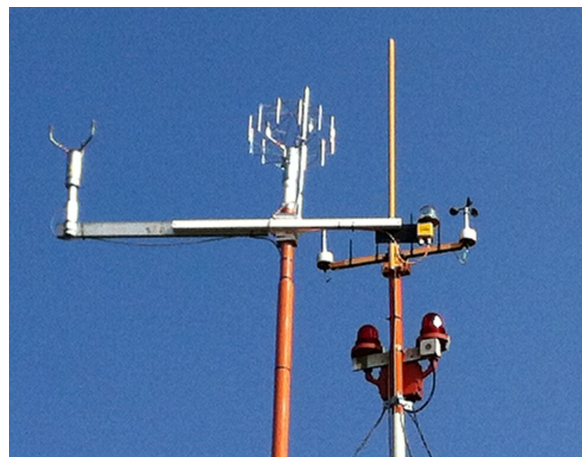
Figuur 2.2: Locatie meetstations Koksijde en Middelkerke

**Topografie** De topografie in Koksijde en Middelkerke is vergelijkbaar; beide gemeenten liggen aan de kust, op amper 20 km van elkaar verwijderd, waar het landschap vlak is. Verder weten we ook dat de topografie van de meetomgeving in zowel Koksijde als Middelkerke niet in die mate veranderd is dat er invloed is op de gemeten windsnelheden in de loop der jaren (persoonlijke communicatie met David Dehenauw per e-mail op 30 oktober 2016), ook al is de anemometer verzet in de jaren '50 en '60. Beide waarnemingsplaatsen zijn open, wat betekent dat er geen hinderlijke invloed is van obstakels zoals gebouwen of hoge bomen die zich in de buurt van de meetstations bevinden. Op Figuur 2.2 is de locatie van beide meetstations aangeduid. Het is van groot belang voor een correcte trendanalyse van de windsnelheid dat de omgeving niet veranderd is. Veranderingen in 'oppervlakterutheid', zoals verstedelijking of ander landgebruik, kunnen de windsnelheid namelijk significant beïnvloeden [27, 57, 34, 50]. Vautard et al. [57] stelden vast dat 25 tot 60% van de gevonden dalende trend in windsnelheid in het noordelijk halfrond te verklaren is door een toename van de oppervlakterutheid, die veroorzaakt kan worden door verstedelijking, bebossing of veranderend landgebruik [50].

Uit de geschiedenis van het meetstation in Koksijde weten we wel dat de anemometer twee keer verplaatst is geweest, namelijk in 1958 en in het begin van de jaren '60. Bovendien werden de windsnelheden vanaf de jaren '90 op een andere manier gemeten dan voordien. Deze veranderingen in locatie en meetapparatuur zouden eventueel een invloed kunnen hebben op de gemeten windsnelheden of op de accuraatheid ervan. Een test die kan gebruikt worden om dit na te gaan is de test van Pettitt – zie ook paragraaf 9.3.1.



(a) MeteoPark van Meteo Wing in Koksijde



(b) Opstelling van een ultrasonische windmeter (drie korte rechtopstaande staafjes uiterst links) naast een windmeter met coupellen (kleine zwarte schepjes uiterst rechts).

Figuur 2.3: Foto's meetstation Koksijde

**Meetapparatuur** De meettoestellen die geïnstalleerd zijn, doen een groot aantal meteorologische waarnemingen, hoofdzakelijk aan de grond op 1,5 m hoogte [31], van onder andere temperatuur, luchtvochtigheid, luchtdruk, neerslag en bewolking. Ook wind wordt door deze toestellen geregistreerd, maar dit op een hoogte van 10 m. Foto's van de aanwezige meetapparatuur op de luchtmachtbasis in Koksijde staan in Figuur 2.3.

## 2.2 De variabelen: interpretatie en betekenis

De variabelen die in beide datasets aanwezig zijn, zijn de volgende:

1. Jaar (1952/1955 - 2016)
2. Maand (01 - 12)
3. Dag (01 - 31)
4. Uur (00, 06, 12, 18 UTC)
5. Code van het meetstation (6400, 6407)
6. Station (KOKSIJDE, MIDDELKERKE)
7. Windsnelheid (m/s)
8. Windrichting (°)
9. Windstoten (m/s)
10. Totale mate van bewolking (octa)
11. Mate van lage bewolking (octa)
12. Mate van hoge bewolking (octa)
13. Temperatuur (°C)
14. Minimumtemperatuur (°C)
15. Maximumtemperatuur (°C)
16. Neerslag (mm)

	YEAR	MONTH	DAY	HOUR	CODE	STATION	Wind speed (m/s)	Wind direction (°)	Wind peak speed (m/s)	Total nebulosity (octa)	Low-Middle nebulosity (octa)	High nebulosity (octa)	Temperature (°C)	Temperature min (°C)	Temperature max (°C)	Precipitation (mm)
1	1952	01	01	00	6400	KOKSIJDE	2.0	100	NA	0	0	0	2.0	NA	NA	NA
2	1952	01	01	06	6400	KOKSIJDE	7.1	250	NA	8	0	8	3.0	2.0	NA	0.4
3	1952	01	01	12	6400	KOKSIJDE	8.1	280	NA	2	2	0	5.0	NA	NA	NA
4	1952	01	01	18	6400	KOKSIJDE	5.0	280	NA	1	1	0	3.0	NA	6	2.0
5	1952	01	02	00	6400	KOKSIJDE	7.1	260	NA	8	0	8	4.0	NA	NA	NA
6	1952	01	02	06	6400	KOKSIJDE	8.1	290	NA	8	0	8	9.0	2.0	NA	5.0
7	1952	01	02	12	6400	KOKSIJDE	7.6	280	NA	8	7	1	7.0	NA	NA	NA
8	1952	01	02	18	6400	KOKSIJDE	5.6	260	NA	8	4	4	6.0	NA	9	7.0
9	1952	01	03	00	6400	KOKSIJDE	4.0	250	NA	3	3	0	2.0	NA	NA	NA
10	1952	01	03	06	6400	KOKSIJDE	3.5	330	NA	8	4	4	4.0	2.0	NA	6.0
11	1952	01	03	12	6400	KOKSIJDE	5.0	360	NA	5	5	0	5.0	NA	NA	NA
12	1952	01	03	18	6400	KOKSIJDE	5.0	320	NA	3	3	0	5.0	NA	6	0.1
13	1952	01	04	00	6400	KOKSIJDE	5.6	330	NA	1	1	0	4.0	NA	NA	NA
14	1952	01	04	06	6400	KOKSIJDE	6.1	330	NA	0	0	0	4.0	3.0	NA	1.0

Figuur 2.4: Uitzicht van de dataset in RStudio

### 2.2.1 Registratiemomenten

Er zijn gegevens beschikbaar sinds 1 januari 1952 in Koksijde en sinds 1 mei 1955 in Middelkerke en dat elke 6 uur, meer bepaald om 0, 6, 12 en 18 UTC tot en met 31 maart 2017. De dataset beschikt dus over een groot arsenaal aan gegevens.

Universal Time Coordinated is een standaardtijd die gebaseerd is op de atoomklok en bijna gelijk aan de Greenwich Mean Time (GMT). In België (en andere delen van Europa) loopt de klok in de winter 1 uur voor op de standaardtijd en in de zomer 2 uur (ten gevolge van winter- en zomertijd). Voor de data betekent dit dus dat metingen gedaan worden om 1, 7, 13 en 19 uur lokale tijd wanneer het winteruur geldt en om 2, 8, 14 en 20 uur lokale tijd als de zomertijd geldt.

Op een totaal van 186 219 meetmomenten (Koksijde en Middelkerke samen) ontbreken er slechts 44 (de rijen komen niet voor in de dataset). Vooral in 2004 ontbreken er waarden zowel in Koksijde als Middelkerke. Deze ontbrekende momenten zijn terug te vinden in Tabellen B.1 en B.2 in Bijlage B.

In de dataset die door Belgocontrol ter beschikking werd gesteld zijn uurlijkse windgegevens uit Middelkerke beschikbaar sinds 1 juli 1958. Metingen werden reeds gedaan vanaf 1 mei 1955, maar tot juli 1958 zijn metingen pas beschikbaar elke drie uur (0, 3, 6, 9, 12, 15, 18 en 21 UTC). Gegevens die in deze dataset beschikbaar zijn, zijn de windsnelheden en de windrichtingen.

### 2.2.2 Windsnelheid en windstoten

Windsnelheden worden gemeten op 10 m boven de grond.

Elke tien minuten wordt de gemiddelde windsnelheid over die voorbije tien minuten berekend. Dit komt neer op 6 gemiddelde windsnelheden per uur. Het gemiddelde van deze 6 gemiddelden geeft het uurlijks gemiddelde over dit voorbije uur. In de dataset zijn echter slechts om de zes uur windsnelheden beschikbaar, waardoor gegevens van vijf opeenvolgende uren niet worden gebruikt. Voor de windstoten wordt het maximum genomen over de voorbije zes uur (persoonlijke communicatie met David Dehenauw per e-mail op 18 mei 2017).

#### De schaal van Beaufort

De schaal van Beaufort wordt gebruikt om de windintensiteit te definiëren. Hierbij gaat het over de gemiddelde windsnelheid gemeten over een periode van ten minste 10 minuten, niet over rukwinden [29]. De schaal van Beaufort loopt van 0 (windstil) tot 12 (orkaan). De opdeling in vijf intervallen die in de windrozen verderop in deze thesis wordt gebruikt, is gebaseerd op deze schaal.

	Omschrijving	Beaufortschaal	Knopen	km/h	m/s
	Windstil	0	0 tot 0,39	Minder dan 1	0 tot 0,2
	Zwakke wind	1 – 2	0,58 tot 6,41	1 tot 11	0,3 tot 3,3
	Matige wind	3 – 4	6,61 tot 15,36	12 tot 28	3,4 tot 7,9
	Vrij krachtige wind	5	15,55 tot 20,8	29 tot 38	8 tot 10,7
	Krachtige wind	6	20,99 tot 26,83	39 tot 49	10,8 tot 13,8
	Harde wind	7	27,02 tot 33,24	50 tot 61	13,9 tot 17,1
	Stormachtige wind	8	33,43 tot 40,24	62 tot 74	17,2 tot 20,7
	Storm	9	40,43 tot 47,43	75 tot 88	20,8 tot 24,4
	Zware storm	10	47,62 tot 55,21	89 tot 102	24,5 tot 28,4
	Zeer zware storm	11	55,40 tot 63,37	103 tot 117	28,5 tot 32,6
	Orkaan	12	meer dan 63,37	meer dan 117	meer dan 32,6

### 2.2.3 Windrichting

De windrichting is de richting van waaruit de luchtstroom komt [59]. Dit betekent dat bij een noordoostelijke wind de luchtverplaatsing gebeurt van het noordoosten naar het zuidwesten. De gemeten windrichtingen in de datasets vertonen een discreet spectrum met sprongen van  $10^\circ$ . Hoeken op een kompas worden afgelezen vanaf het noorden in wijzerzin. In de datasets wordt deze conventie gevolgd, waardoor  $360^\circ$  een noordelijke wind voorstelt.



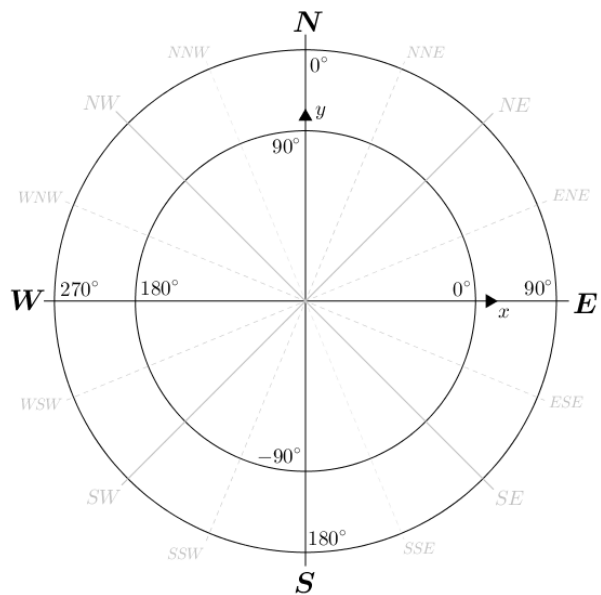
Hoek in dataset (°)	Hoek op goniometrische cirkel (°)	Windstreek	Hoek in dataset (°)	Hoek op goniometrische cirkel (°)	Windstreek
<b>360</b>	<b>90</b>	<b>N</b>	<b>180</b>	<b>-90</b>	<b>S</b>
10	80	N	190	-100	S
20	70	NNE	200	-110	SSW
30	60	NNE	210	-120	SSW
40	50	NE	220	-130	SW
50	40	NE	230	-140	SW
60	30	ENE	240	-150	WSW
70	20	ENE	250	-160	WSW
80	10	E	260	-170	W
<b>90</b>	<b>0</b>	<b>E</b>	<b>270</b>	<b>180</b>	<b>W</b>
100	-10	E	280	170	W
110	-20	ESE	290	160	WNW
120	-30	ESE	300	150	WNW
130	-40	SE	310	140	NW
140	-50	SE	320	130	NW
150	-60	SSE	330	120	NNW
160	-70	SSE	340	110	NNW
170	-80	S	350	100	N

Tabel 2.5: Windrichtingen in de dataset en op de goniometrische cirkel en de overeenkomstige windstreken

Op sommige plaatsen in de dataset staat er een windrichting van  $0^\circ$ . Deze is niet opgenomen in de tabel, maar staat voor windstille omstandigheden. In de datasets staan slechts 37 entries (27 in Koksijde en 10 in Middelkerke) met windrichting  $0^\circ$ . In de dataset van Koksijde komen bovendien twee entries voor waarbij de windsnelheid niet nul is, hoewel de windrichting aangeeft dat het windstil zou geweest zijn. Voor het vervolg zullen we deze waarden echter niet aanpassen of weglaten. Het is aannemelijker dat het hier gaat over een verkeerd genoteerde windrichting: windsnelheden worden namelijk gerapporteerd aan de hand van uurgemiddelden over de voorbije zes uur en gegeven dat hier een windsnelheid verschillend van nul gerapporteerd wordt, doet dit vermoeden dat met een grotere kans de windrichting verkeerd is. Bovendien ligt de focus in deze thesis op de windsnelheden en niet op de windrichtingen en daarom zal de windrichting niet op NA gezet worden.

## 2.2.4 Bewolkingsgraad

Voor elk moment wordt in de dataset weergegeven hoe bewolkt het op dat moment was. De mate van bewolking wordt in de meteorologie uitgedrukt in octa's en duidt aan welke hoeveelheid van de hemelkoepel bedekt is met zichtbare wolken, in achtsten. Zo betekent een bewolkingsgraad van 8 octa's 'volledig bewolkt', één van 4 duidt op een halfbewolkte hemel en 0 octa's geeft aan dat het blauwe hemel is [30]. Tabel 2.7 geeft weer hoe de bewolkingsgraad precies wordt ingedeeld [30].

Figuur 2.6: Kompas met windrichtingen vanaf het noorden in wijzerzin en goniometrische cirkel met hoeken vanaf het positieve deel van de  $x$ -as in tegenwijzerzin.



Term	Mate van bewolking in octa's	Omschrijving
Heldere hemel	0	Wolkenloze hemel
Lichtbewolkt	1–2	Er bevinden zich stapelwolken (cumulus) of hoge sluierwolken (cirrus) aan de hemel
Halfbewolkt	3–5	Afwisselend opklaringen en wolkenvelden
Zwaarbewolkt	6–7	Weinig of geen opklaringen; lage wolken (nimbostratus, stratus) of hoge, dikke wolkenluiers (cirrostratus)
Betrokken	8	Volledig bewolkt

Tabel 2.7: Termen die de mate van bewolking uitdrukken

In de dataset wordt bovendien een onderscheid gemaakt tussen lage en middelhoge (0 tot 6 km), hoge (6 tot 13 km) en totale bewolking. De bewolkingsgraad werd hierbij louter discriptief bekeken, maar zal voor (trend)analyses niet gebruikt worden.

### 2.2.5 Temperatuur

Voor elk meetmoment wordt de temperatuur (°C) óp het uur van rapportering weergegeven. Bovendien wordt ook de minimum- en maximumtemperatuur van de voorbije 24 uur gegeven. De minimumtemperatuur van de voorbije 24 uur wordt gerapporteerd om 6 UTC en de maximumtemperatuur gemeten in de voorbije 24 uur wordt gegeven om 18 UTC. Temperaturen worden sinds 1982 gemeten tot op 1 decimaal nauwkeurig. Vóór die datum werden slechts gehele getallen geregistreerd.

### 2.2.6 Neerslag

Neerslaghoeveelheid wordt in de dataset om de 12 uur gerapporteerd, om 6 UTC en om 18 UTC. Neerslaghoeveelheden zijn tot op een halve millimeter nauwkeurig. Wanneer  $x$  mm neerslag wordt gemeten, komt dit overeen met  $x$  liter water per vierkante meter.

## 2.3 Hulpvariabelen

Tijdens de analyses die volgen, bleek het zinvol een aantal bijkomende variabelen te definiëren, die zijn afgeleid van de in 2.2 beschreven variabelen.

### 2.3.1 Moment

Een eerste variabele die nuttig is, is de variabele **moment**. Hierbij wordt voor elk meetmoment de datum en het uur samengevoegd. Voor elk observatiemoment worden de variabelen **YEAR**, **MONTH**, **DAY** en **HOOR** samengevoegd. Deze hulpvariabele zal vooral bruikbaar zijn bij het detecteren van de zeebriesdagen in de dataset: hierbij zal de doorsnede van twee datasets genomen worden volgens de variabele **moment**.

### 2.3.2 Seizoen en seizoenjaar

In de variabele **SEIZOEN** wordt voor elk meetmoment bepaald in welk seizoen de meting werd gedaan. Deze variabele zal vooral van belang zijn wanneer analyses per seizoen gemaakt worden. Hierbij gaat het over de

astronomische seizoenen<sup>1</sup>.

Wanneer per jaar (van 1952 t.e.m. 2016) de gemiddelde windsnelheid per seizoen moet berekend worden, wordt op een aantal hindernissen gestoten. Een eerste veronderstelling die gemaakt wordt, is dat de lente het eerste seizoen van het tropische jaar is. Het jaar is ingedeeld in vier seizoenen:

- lente (21 maart t.e.m. 20 juni)
- zomer (21 juni t.e.m. 20 september)
- herfst (21 september t.e.m. 20 december)
- winter (21 december t.e.m. 20 maart)

De lente, zomer en herfst vallen volledig in hetzelfde kalenderjaar. De winter daarentegen is gelegen in twee kalenderjaren. Er moet dus beslist worden of de winter, startend in december van jaar  $X$  en eindigend in maart van jaar  $X + 1$  zal bekeken worden als de winter van jaar  $X$  of die van jaar  $X + 1$ . Gebaseerd op het MIRA klimaatrapport [7] wordt beslist om elke winter te laten tellen als de winter van het vroegste jaar. Dit betekent concreet dat de winter van 1993 loopt van 21 december 1993 tot 20 maart 1994. Met het oog op een efficiënte verwerking werden aan de dataset kolommen **SEIZOEN** (met als waarden lente (L), zomer (Z), herfst (H) en winter (W)) en **SEIZOENJAAR** (zelfde gegevens als bij **SEIZOEN**, maar waarbij ook de jaartallen worden vermeld) toegevoegd. Voor de lente, zomer en herfst moeten we alleen maar de waarden voor **YEAR** en **SEIZOEN** concatenen. Voor de winter moeten we aangeven dat januari, februari en maart (t.e.m. 20 maart) behoren tot het voorgaande jaar.

### 2.3.3 Windrichting (hoek)

In de variabele **HOEK** worden de gegeven windrichtingen omgezet naar hoeken zoals deze op de goniometrische cirkel zouden gelezen worden. Tabel 2.5 geeft de intervallen met bijhorende windrichtingen weer. Belangrijk om op te merken is dat de windrichtingen niet overeenkomen met de manier waarop hoeken op een goniometrische cirkel worden gelezen. Hoeken op een goniometrische cirkel worden afgelezen vanaf het positieve deel van de  $x$ -as in tegenwijzerzin. Zo komt een hoek van  $40^\circ$  in de dataset overeen met een hoek van  $50^\circ$  op de goniometrische cirkel. Om elk van de hoeken uit de dataset te noteren zoals die zouden afgelezen worden op de goniometrische cirkel, waarbij de zin van de  $x$ -as het oosten aanduidt en die van de  $y$ -as het noorden, kan de volgende transformatie worden doorgevoerd:

$$\begin{array}{ccc} \text{Windrichtingen in dataset} & \rightarrow & \text{Hoeken op goniometrische cirkel} \\ \alpha & \mapsto & -\alpha + 90^\circ \end{array}$$

Deze transformatie is vooral belangrijk bij het ontbinden van de windvector in zijn componenten. In Tabel 2.5 staan de transformaties opgesomd.

### 2.3.4 Verschil Koksijde - Middelkerke

Eén van de nieuwe variabelen, **windverschil**, geeft het verschil aan in windsnelheid tussen beide meetstations ( $\text{wind}_K - \text{wind}_M$ ). Hierbij geeft een negatieve waarde voor het verschil aan dat het meer waaide in Middelkerke dan in Koksijde en vice versa.

Voor de analyse van de temperatuur worden dezelfde variabelen aangemaakt. **tempverschil** geeft hierbij het verschil in temperatuur aan tussen beide meetstations.

<sup>1</sup>Naast de astronomische seizoenen zijn er ook de meteorologische seizoenen. Deze beginnen steeds aan het begin van een nieuwe maand. Zo valt het begin van de meteorologische lente op 1 maart, de zomer op 1 juni, de herfst op 1 september en de winter op 1 december.

### 2.3.5 Variabiliteit tussen opeenvolgende metingen

Bij het beoordelen van de waarschijnlijkheid dat een waarde fout is, is het aangewezen een maat te hebben voor de verschillen tussen deze waarde ( $w_i$ ) en de omliggende waarden ( $w_{i-1}$  en  $w_{i+1}$ ). Een goede maat daarvoor is  $(w_i - w_{i-1})^2 + (w_i - w_{i+1})^2$ . Hoe groter deze som wordt, des te groter is de waarschijnlijkheid dat de buitensporige afwijking van geregistreeerde windsnelheid in de tijd toe te schrijven is aan een meetfout. Concreet betekent dit dus dat de geregistreeerde windsnelheid  $w_i$  op tijdstip  $i$  in grote mate verschilt van de meting ervoor (dus op tijdstip  $i - 1$ ) en/of van de meting erna (dus op tijdstip  $i + 1$ ). Deze variabele wordt in de dataset aangeduid met `windschilkoks` en `windschilmid`.

`tempschilkoks` en `tempschilmid` worden op analoge manier gemaakt als `windschilkoks` en `windschilmid`, maar dan voor de temperaturen.

### 2.3.6 Orthogonale componenten van de windsnelheid

De twee orthogonale componenten  $u$  en  $v$  van de windvector worden bekomen door deze windvector loodrecht te projecteren op de  $x$ - en  $y$ -as van een orthogonaal assenstelsel. Hierbij ligt de  $u$ -component volgens de oost-westas (positieve  $u$  is oosten, negatieve  $u$  is westen) en de  $v$ -vector volgens de noord-zuidas (positieve  $v$  is noorden, negatieve  $v$  is zuiden).

## Hoofdstuk 3

# Clean-up van de dataset

Vooraleer te starten met de analyse van de dataset lijkt het zinvol om deze eerst van naderbij te bestuderen. Zo is het interessant om op zoek te gaan naar mogelijks foutieve waarden, outliers en ontbrekende gegevens zodat met een opgeschoonde dataset kan worden verdergewerkt.

### 3.1 Ontbrekende meetmomenten

In de dataset van Koksijde zijn er een aantal momenten waarop geen metingen werden geregistreerd. In de dataset ontbreekt de overeenkomstige rij helemaal en wordt gesprongen van de ene meting naar de volgende meting 12 uur later in plaats van 6 uur later. Deze lijst van ontbrekende meetmomenten staat weergegeven in Tabellen B.1 en B.2 in Bijlage B.

In de dataset van Koksijde zijn er 23 ontbrekende rijen (0,024%) en in Middelkerke zijn dat er 22 (0,024%). Het gaat hierbij niet over dezelfde rijen.

We vragen ons af of er al dan niet een patroon zit in het voorkomen van de missing data. Kijken we hierbij enkel naar het moment in de tijd waarop deze extremen voorkomen, dan vinden we geen patroon. Er is geen vast aantal dagen gelegen tussen twee opeenvolgende afwezige gegevens. De toestand van het weer op deze dagen vertoont evenmin een patroon. Dit is te zien in Tabel B.5 in Bijlage B. De laatste drie kolommen zeggen respectievelijk iets over de bewolgingsgraad, de temperatuur en de neerslag. Ook hierin is geen constante te vinden: het is soms bewolkt, soms niet, missing data komen zowel voor op warme als op koude dagen, zowel op dagen met veel als met weinig neerslag. Op het eerste zicht is er dus geen patroon te vinden in de missing data. Omdat dit geen studie is in verband met missing data, zullen we hier ook niet verder op ingaan.

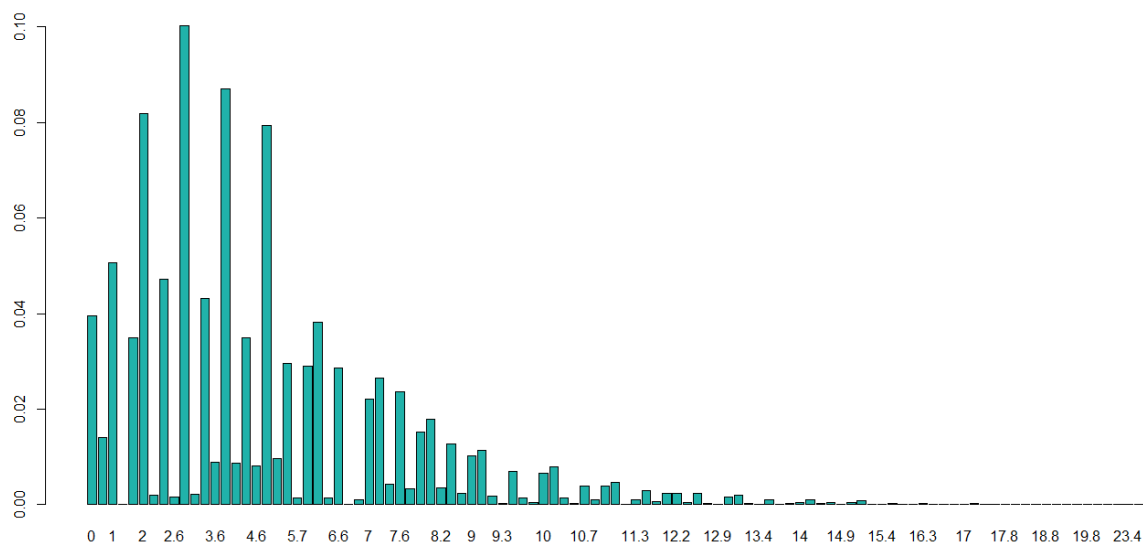
### 3.2 Windsnelheden

Vroeger werden waarnemingen handmatig ingegeven. Het is dus goed mogelijk dat af en toe een verkeerde waarde werd ingevoerd. Ook werden windsnelheden tot vóór 1996 in knopen gemeten. Later ging men over naar m/s en werden ook de in knopen gemeten windsnelheden omgezet naar m/s (persoonlijke communicatie met Wim Demol, Belgocontrol op 12 april 2017). Zo zou het kunnen dat een windsnelheid van 9 knopen werd bedoeld, maar door een onnauwkeurigheid 90 knopen werd ingegeven. Hierdoor bevindt zich nu in de dataset bijvoorbeeld een windsnelheid van 46,4 m/s in plaats van 4,6 m/s. De laatste decennia hebben KMI-medewerkers datasets nagekeken en gecorrigeerd, waardoor dergelijke fouten bij metingen tussen 1970 en 2013 niet meer te vinden zijn. Vroeger werd deze controle niet gedaan waardoor extreme waarden zich nog steeds in de dataset bevinden (persoonlijke communicatie met David Dehenauw per e-mail op 19 februari 2017). Gegevens vanaf 2013 werden ook nog niet gecorrigeerd.

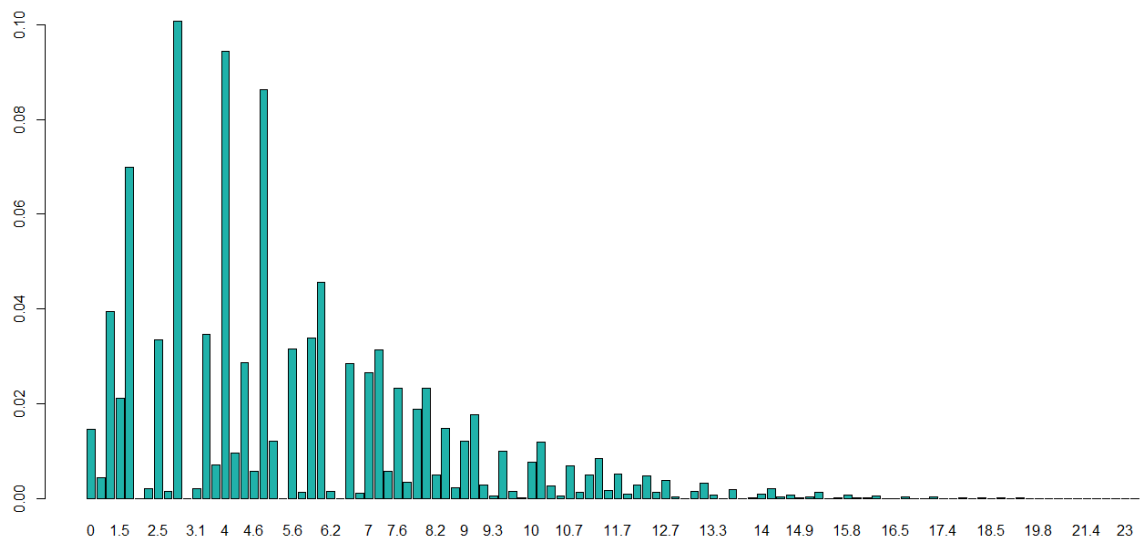
### 3.2.1 Spectrum van gemeten waarden

#### Frequentie via staafdiagram

In eerste instantie is het interessant om te bekijken welke verschillende windsnelheden in de dataset voorkomen. Een eerste globaal beeld wordt gegeven in de staafdiagrammen in Figuur 3.1a en 3.1b. Deze diagrammen verschillen van histogrammen omdat zij voor *elke waarde* die in de dataset voorkomt, aangeven hoeveel keer deze waarde voorkomt. De horizontale as toont dus alle voorkomende waarden, waardoor ze niet uniform gespatieerd zijn, zoals het geval is bij een histogram (voor de histogrammen, zie Figuur 4.1). De hoogte van de staven komt wel overeen met de relatieve frequentie. Op deze plots valt onmiddellijk op dat er erg veel verschillende waarden voorkomen.



(a) Koksijde (1952 – 2016)



(b) Middelkerke (1955 – 2016)

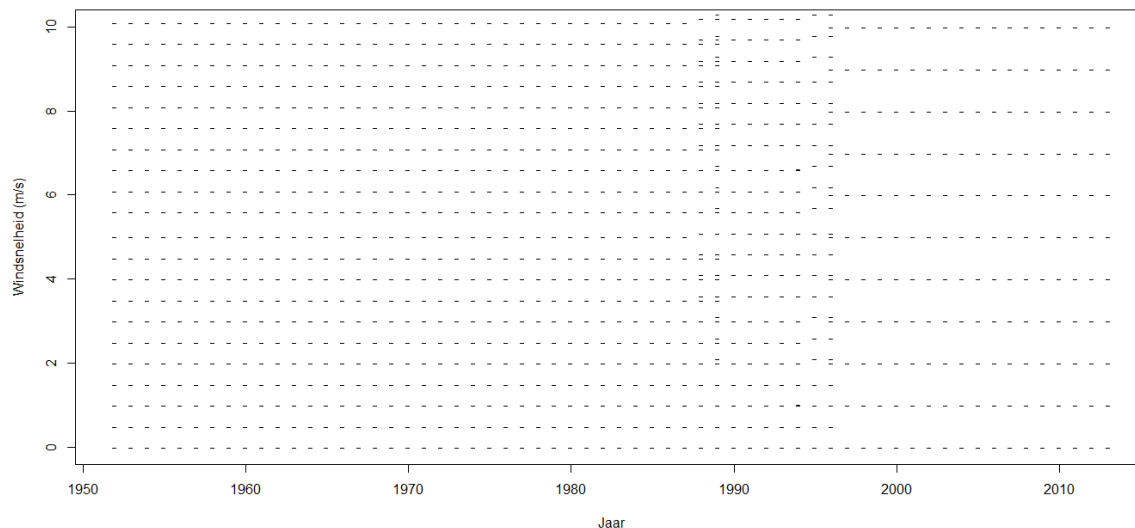
Figuur 3.1: Staafdiagram van de windsnelheden in Koksijde en Middelkerke

## Unieke waarden

In Tabel B.3 in Bijlage B is te zien dat bij de gemeten windsnelheden een aantal vreemde waarden staan: een windsnelheid van 6,63 m/s (7 juli 1994), 1,02 m/s (8 juli 1994) en 11,22 m/s (17 februari 1995) staan eenmaal genoteerd in de dataset van Koksijde; in deze van Middelkerke komen eenmaal de waarden 3,06 m/s (4 mei 1994) en 2,04 m/s (24 februari 1995) voor.

**Van knopen naar m/s vanaf 1 juli 1996** Uit communicatie met Wim Demol van Belgocontrol werd duidelijk dat de windsnelheden vóór 1997 in knopen werden gemeten in plaats van in m/s. Het is pas na 1997 dat windsnelheden in m/s worden gemeten. In de dataset werden echter wel alle windsnelheden die oorspronkelijk in knopen werden gemeten, omgezet naar m/s. De voornoemde unieke waarden zijn dus niet noodzakelijk foutief. Het is wel mogelijk dat door tikfouten toch een verkeerde waarde werd ingegeven. Deze waarden zijn echter geen outliers die de analyse sterk zouden kunnen beïnvloeden, dus voor het vervolg worden deze waarden gewoon meegenomen.

Uit staafdiagrammen in Figuur 3.1a en 3.1b en Tabel B.3 in Bijlage B blijkt dat sommige windsnelheden (1, 2, 3, 4, 5, ... m/s) bijna dubbel zoveel gemeten worden dan andere windsnelheden (1.5, 2.5, 3.5, 4.5, ... m/s). Algemeen is te zien dat gehele waarden systematisch vaker worden gemeten dan decimale waarden. Uit een plot van de gemeten waarden in de tijd (zie Figuur 3.2) is duidelijk op te maken dat het meettoestel enkel en alleen gehele waarden weergeeft vanaf 1 juli 1996, zowel in Koksijde als in Middelkerke. Dit zou ons op het eerste zicht doen vermoeden dat de gevoeligheid van het meettoestel varieert. Dit heeft echter alles te maken met het feit dat tot en met 1996 windsnelheden in knopen werden gemeten (persoonlijke communicatie met Wim Demol, Belgocontrol op 12 april 2017). Mogelijks is het zinvol om op de data methoden toe te passen om deze eventuele *digit preference* te analyseren en indien nodig te corrigeren. We laten dit hier echter buiten beschouwing en werken verder met de geregistreerde waarden.



Figuur 3.2: Plot met de gemeten windsnelheden. Vanaf 1 juli 1996 wordt enkel nog in gehele waarden gemeten. Het is op dit keerpunt dat windsnelheden in m/s werden gemeten en niet langer in knopen. Dit spectrum wordt zowel in Koksijde als in Middelkerke gezien.

### 3.2.2 Ontbrekende waarden

In de dataset zijn niet voor alle meetmomenten windsnelheden geregistreerd. Tabel 3.3 geeft weer hoeveel ontbrekende waarden voor de windsnelheid zich in de dataset van Koksijde en Middelkerke bevinden. Gegevens

zijn in Koksijde beschikbaar vanaf 1 januari 1952, in Middelkerke is dat pas vanaf 1 mei 1955. Het theoretisch te verwachten aantal meetmomenten in Koksijde en Middelkerke ligt iets hoger dan het werkelijk aantal meetmomenten. Dit is te verklaren doordat op sommige plaatsen een volledige ‘rij’ gegevens ontbreekt zoals aangegeven in sectie 3.1.

	Theoretisch aantal meetmomenten	Werkelijk aantal meetmomenten	NA's
Koksijde	95 324	95 301 (−0,028%)	66 (0,069%)
Middelkerke	90 456	90 434 (−0,033%)	10 (0,011%)
Totaal	185 780	185 735 (−0,031%)	76 (0,041%)

Tabel 3.3: Aantal nulmetingen en niet-geregistreerde metingen in de dataset.

Wat betreft de rukwinden, beschikken de datasets van Koksijde en Middelkerke over een beperkt aantal gegevens. Voor Koksijde werden windstoten pas geregistreerd vanaf 1985. Maar ook dan duurde het nog tot 1994 vooraleer er een aanvaardbaar aantal gegevens beschikbaar was. Het aantal beschikbare waarden voor de rukwinden gedurende de laatste 30 jaar staan weergegeven in de Tabel 3.4.

Periode	Aantal beschikbare metingen	Percentage bruikbare data
1952 – 1984	0	0% data
1985 – 1987	407	9,29% data
1988 – 1993	3317	37,83% data
1994 – 2000	10 001	97,79% data
2001 – 2003	3889	88,79% data
2004 – 2016	14 562	99,66% data

Tabel 3.4: Tabel met beschikbare rukwindgegevens in verschillende perioden in Koksijde

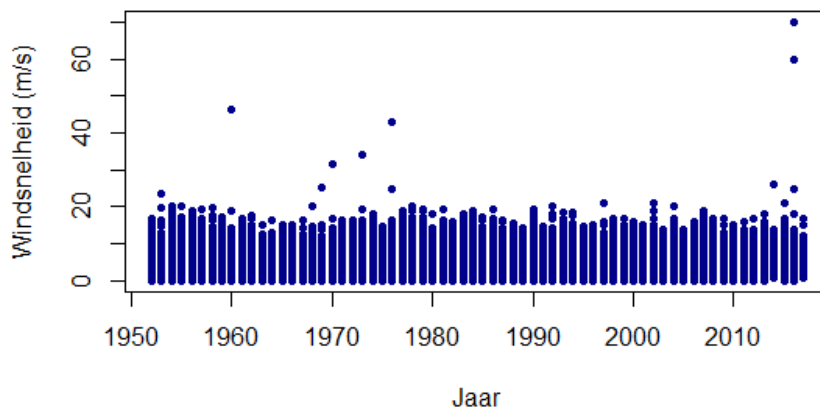
### 3.2.3 Foutieve outliers

**Koksijde** Bij het uitzetten van alle windsnelheden in Koksijde blijken verdacht hoge waarden voor te komen in de dataset (zie Figuren 3.5 en 3.6). Hierbij werden de windsnelheden enerzijds samen geplott, anderzijds werden deze ook per seizoen weergegeven. In deze figuren is te zien dat er een aantal zeer hoge waarden terug te vinden zijn. De vraag is of deze gemeten windsnelheden effectief overeenkomen met wind in Koksijde, of of het hier gaat over meet- of tikfouten.

**Middelkerke** Ook voor Middelkerke plotten we de gemeten windsnelheden voor de periode 1955 tot 2016. Hier valt op dat er geen enkele waarde verdacht hoog lijkt (zie Figuur 3.7). Het is dus ook niet nodig om zoals bij Koksijde een plot per seizoen te maken.

#### Criteria voor outliers

Op basis van een aantal welgekozen criteria is het te rechtvaardigen om een aantal waarden uit de dataset te verwijderen. De vijf indicatoren die worden gehanteerd, staan hieronder weergegeven: drie van deze vijf zijn terug te vinden in Figuur 3.10.



Figuur 3.5: Gemeten windsnelheden in Koksijde (1952 - 2016)

1. De windsnelheid zelf (de absolute grootte van de geregistreerde waarde: het gemiddelde in de voorbije 6 uur van de 10-minutengemiddelden voor elk uur, zie Hoofdstuk 2). Gemiddelde windsnelheden boven 30 m/s (= 108 km/h) in België zijn erg onwaarschijnlijk; een rukwind van 30 m/s is daarentegen wel mogelijk.
2. Het verschil tussen de gemeten windsnelheden in Koksijde en Middelkerke op dezelfde momenten. Dit verschil in windsnelheid in beide meetstations geeft ons in zekere mate een idee van correctheid van de meting. Het lijkt logisch dat de windsnelheden in Koksijde en Middelkerke op gelijke tijdstippen gelijkwaardig zijn, aangezien beide meetstations op amper 20 km van elkaar gelegen zijn. Om deze verschillen te bepalen, moeten we weten voor welke momenten er zowel in Koksijde als in Middelkerke windgegevens zijn.
3. **Windschil**: hiermee wordt de som van de kwadraten van de verschillen met de meting ervoor en met de meting erna bedoeld:  $(w_i - w_{i-1})^2 + (w_i - w_{i+1})^2$ , met  $w_i$  de wind op tijdstip  $i$  (zie Hoofdstuk 2)
4. **Wind peak speed** (windstoten, indien geregistreerd)
5. Informatie over de weersomstandigheden die dag zoals te vinden op historische weerkaarten of rapportering op de KMI-website als extreem weersevent (windstoten worden gerapporteerd wanneer deze 108 km/h = 30 m/s overschrijden).

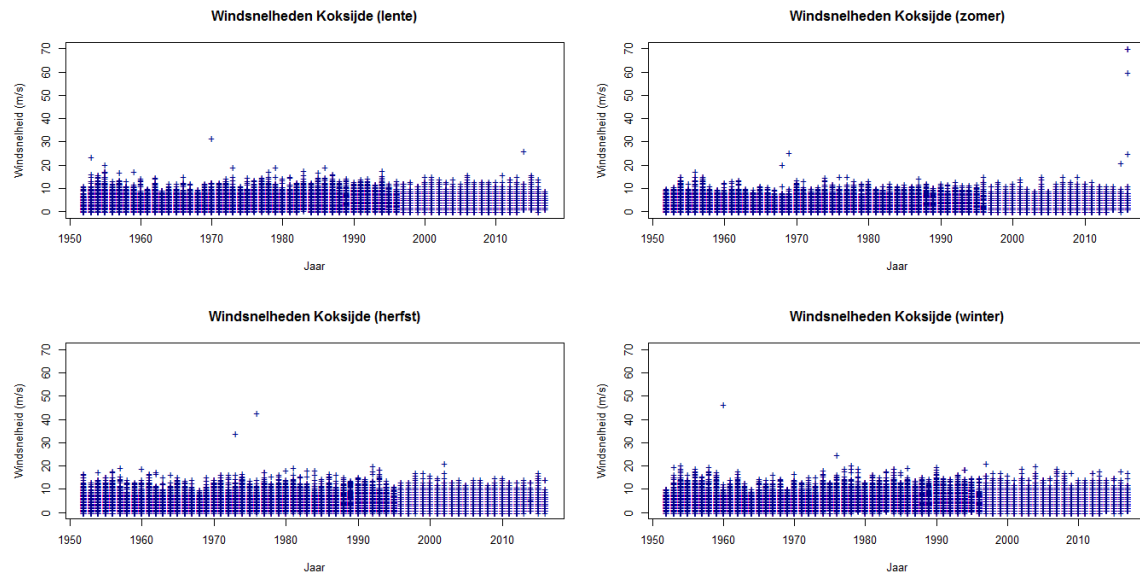
**Visuele inspectie** De scatterplot in Figuur 3.10 toont drie variabelen. Op de  $x$ -as staat het verschil tussen gemeten windsnelheden in Koksijde en Middelkerke (indicator 2). Op de  $y$ -as staat de **windschil** weergegeven (indicator 3) en met een kleurschaal wordt aangegeven wat de gemeten windsnelheden waren in Koksijde (indicator 1). Aan de hand van deze grafiek zijn we dus geneigd om te zeggen dat er zeker 10 outliers zijn waarmee we geen rekening zouden moeten houden omdat deze metingen waarschijnlijk verkeerd zijn. Deze 10 datapunten staan in het rood omcirkeld. Dat dergelijke windsnelheden in België zouden voorkomen, is onwaarschijnlijk, tenzij in een windhoos.

Deze extreme waarden<sup>1</sup> die hieruit blijken zijn te vinden in Tabel 3.8.

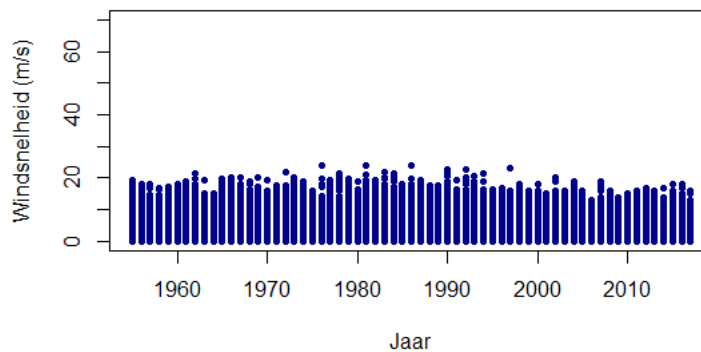
In Figuur 3.10 werden al drie van de vijf indicatoren voor de dubieuze metingen weergegeven.

<sup>1</sup>De metingen worden in Koksijde gedaan op de luchtmachtbasis vlak naast de landingsbaan. Zou het eventueel kunnen dat door een opstijgend, landend of voorbijkomend militair vliegtuig of helikopter deze hoge windsnelheid bekomen werd?





Figuur 3.6: Windsnelheden Koksijde per seizoen in de periode 1952 – 2016



Figuur 3.7: Windsnelheden Middelkerke in de periode 1952 – 2016

Het vierde criterium dat raadzaam zou zijn om te raadplegen, was de **wind peak speed**. Echter, voor geen van de eerste vijf momenten die in de bovenstaande tabel worden weergegeven, is de **wind peak speed** in de dataset aanwezig. Voor de meest recente metingen, i.e. die vanaf 2014, zijn wel waarden voor de **wind peak speed** aanwezig en deze staan ook weergegeven in Tabel 3.8. Op elk van de meetmomenten is de geregistreeerde ogenblikkelijke windsnelheid (veel) hoger dan de zwaarst gemeten windstoot in de voorbije 6 uur, waaruit kan besloten worden dat elk van de windsnelheden verkeerd zijn. We merken op dat in 2016 tweemaal een windsnelheid van 70 m/s (= 252 km/h) en eenmaal één van 60 m/s (= 216 km/h) staat geregistreeerd. Dat dergelijke windsnelheden in België zouden voorkomen als uurlijkse gemiddelden is hoogst onwaarschijnlijk.

Uit de beschrijving van de weersomstandigheden op de dagen van de extreme metingen vóór 1970 kunnen we niet afleiden dat dermate hevige wind aanwezig zou zijn geweest, maar ook niet dat dat niet zo is. De 10 gedetecteerde punten liggen verspreid over de vier seizoenen, komen voor in verschillende jaren op verschillende uren en op verspreide momenten in een maand. Er is dus geen duidelijk patroon te bespeuren in het moment waarop deze outliers worden gedetecteerd.

Deze extreme waarden werden in samenspraak met prof. David Dehenauw, die aan de hand van weerkaarten

Dag	Uur	Seizoen	Wind (Koksijde) (voor na)	Wind (Mid- delkerke) (m/s)	Weersomstandigheden/Windstoot (m/s)
18 januari 1960	18	W	<b>46,4</b> (5 7,1)	5,6	Zwaar bewolkt (7-8/8 bewolkt) met lichte neerslag, 12°C
5 augustus 1969	00	Z	<b>25,4</b> (2,5 1)	0,0	5-6 bewolking, buien overdag, 's nachts brede opklaringen, 15°C
14 mei 1970	06	L	<b>31,6</b> (1 1,5)	2,5	droge dag, 7-8/8 bewolkt, 7°C
27 september 1973	18	H	<b>34,1</b> (3,5 7,6)	1,5	6-8/8 bewolkt, droog overdag, 13°C
17 december 1976	18	H	<b>42,8</b> (6,1 3,5)	3,5	3 mm neerslag tussen 12 en 18 uur, zwaar bewolkt (8/8), 2°C
31 maart 2014	18	L	<b>26</b> (3 1)	2	5
14 juli 2016	12	Z	<b>70</b> (4 2)	6	8
16 juli 2016	0	Z	<b>60</b> (5 2)	3	6
6 augustus 2016	18	Z	<b>70</b> (5 1)	3	6
31 augustus 2016	0	Z	<b>25</b> (3 1)	3	4

Tabel 3.8: Tabel met extreme waarden voor de windsnelheden in m/s in Koksijde.

van op die momenten kon uitmaken dat dergelijke windsnelheden voor de gegeven weersomstandigheden niet mogelijk waren, uit de dataset weggelaten.

### Waarom het aannemelijk is om deze tien waarden als verkeerd te beschouwen

1. In elk van deze punten ligt de gemeten windsnelheid vrij hoog (zeven van de tien snelheden  $> 30$  m/s). Deze windsnelheden zijn eerder zeldzaam aan de Belgische kust. Als deze snelheden al voorkomen, gaat het meestal om *windpieken* van meer dan 30 m/s (zoals bijvoorbeeld in het jaar 1953 en 1954 [33].)
2. Voor elk van deze punten is de som  $(w_i - w_{i-1})^2 + (w_i - w_{i+1})^2$  vrij groot ( $> 1000$  m<sup>2</sup>/s<sup>2</sup>). In theorie is het echter mogelijk dat er op een bepaald tijdstip zeer hevige wind waait, maar dat het zes uur eerder en zes uur later rustig was. Toch lijkt dit al vrij onwaarschijnlijk.
3. Bij elk van deze vijf punten is het verschil in gemeten windsnelheden tussen Koksijde en Middelkerke groter dan 20 m/s ( $\approx 72$  km/h). Het is onwaarschijnlijk dat het in meetstations die maar 19,58 km van elkaar liggen, er zo'n groot verschil is in windsnelheid.
4. Een analyse van de weerkaarten op de betreffende momenten versterkt het vermoeden dat het gaat over foutieve gegevens.

**Nog vier verdachte waarden in Koksijde** Verder zijn er op Figuur 3.10 ook nog vier andere punten te zien (in blauwe wolk) waar het verschil in windmetingen tussen beide stations hoger lijkt te liggen. Hierbij gaat het over de momenten in Tabel 3.9.

Dit zijn toch vrij grote verschillen in windsnelheid tussen beide stations. De vraag stelt zich wat hiervoor de reden zou kunnen zijn. Hebben zich effectief zo'n grote verschillen voorgedaan of gaat het hier over onnauwkeurig genoteerde metingen? De data waarop deze waarden gemeten werden, zijn niet terug te vinden in de lijst met extreme weerfenomenen op de website van het KMI. Mogelijks was er plaatselijk meer wind.

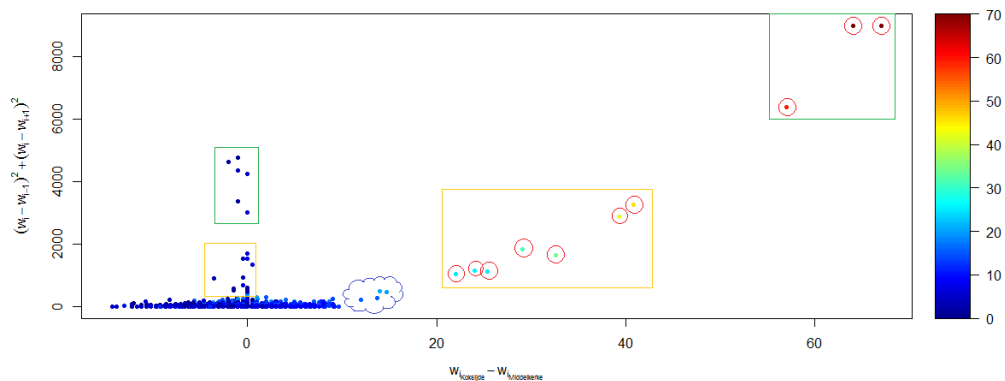
Dag	Uur	Seizoen	Wind (Koksijde)	Wind (Middelkerke)	Weersomstandigheden/Windstoot (m/s)
28 augustus 1968	18	Z	<b>21,3</b> (5 4,5)	5.6	Afwisselend zon en wolken, droog, maximumtemperatuur rond 21°C.
3 augustus 1977	06	Z	<b>15,2</b> (2 4,5)	1.5	Maximumtemperatuur 23°C, toenemende bewolking in de loop van de dag, droog, windstoot van 23 m/s
29 april 2011	18	L	<b>16,0</b> (6 5)	4.0	Wisselend tot zwaarbewolkt, tijdelijk lichte regen of een bui, maximumtemperatuur rond 19°C.
22 juni 2015	6	Z	<b>21</b> (4 6)	7	9

Tabel 3.9: Tabel met verdachte waarden voor de windsnelheid in m/s in Koksijde

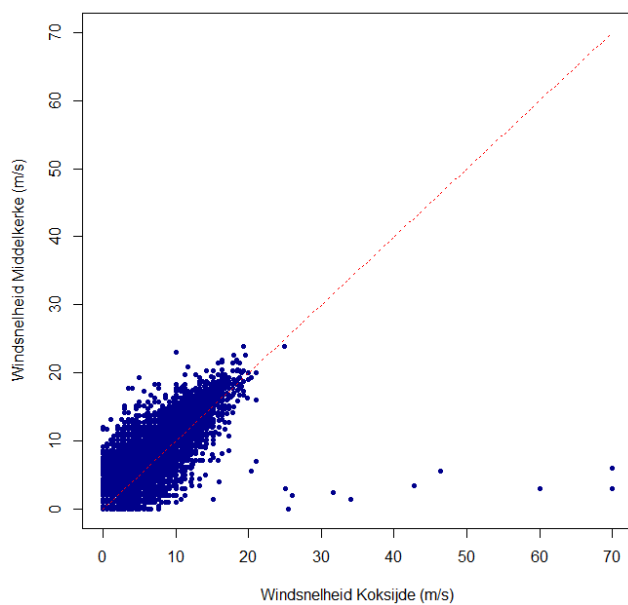
- Op 28 augustus 1968 geven de weersomstandigheden aan dat dit een vrij mooie zomerdag geweest is. Er is voor deze entry in de dataset geen **wind peak speed** beschikbaar, dus kan er ook niet gecontroleerd worden of de waarde van 20,3 m/s ( $\approx 72$  km/h) accuraat was. Op basis van de beschikbare gegevens (weersomstandigheden en wind in Middelkerke) wordt besloten om bij de verdere analyse deze windsnelheid op NA (Not Available) te zetten zodat hier in de verdere analyse geen rekening mee gehouden wordt.
- Op 3 augustus 1977 neemt in de loop van de dag de bewolking toe. Het is niet uitgesloten dat dan een wind waait van 15,2 m/s ( $\approx 54$  km/h). In Middelkerke werd op 3 augustus 1977 als hoogste windsnelheid een snelheid van 23 m/s gemeten, dus deze waarde 15,2 is aannemelijk. Hoewel ook dit een twijfelgeval is, zal deze waarde toch behouden blijven in de dataset.
- De weersomstandigheden leren dat 29 april 2011 geen al te mooie dag was. In dit opzicht is een windsnelheid van 16 m/s ( $\approx 57$  km/h) eventueel mogelijk, dus ook deze waarde blijft behouden in de dataset. Deze keuze kan bovendien gerechtvaardigd worden door het feit dat de meting vrij recent werd gedaan en zich in het gedeelte van de dataset bevindt dat reeds een controle onderging.
- Vervolgens kunnen we voor de meting van 22 juni 2015 vrij zeker zijn dat deze meting verkeerd is aangezien voor dit moment ook de **wind peak speed** gegeven is, die slechts 9 m/s blijkt te zijn. Ook deze windsnelheid van 21 m/s zal in de dataset op NA worden gezet.

Tot slot is in Figuur 3.10 naast de voornoemde waarden nog een andere groep punten met een hoge waarde voor de variabele  $(w_i - w_{i-1})^2 + (w_i - w_{i+1})^2$  opvallend, namelijk deze waar het verschil in windsnelheid tussen beide stations nagenoeg nul is. Dit betekent dus dat twee opeenvolgende metingen relatief veel van elkaar verschillen, m.a.w. dat er een plotse, hevige wind de kop opstak. Hoewel het op het eerste zicht vreemd lijkt om dergelijke waarden terug te vinden, is het vrij evident dat deze er staan; dit zijn precies de metingen net voor en net na de ‘verkeerde’ metingen in Koksijde. Er wordt nu echter wel een goede meting genoteerd want het verschil tussen het station in Koksijde en Middelkerke is nagenoeg nul. De bovenste zes punten zijn in verband te brengen met de drie meest extreme waarden (aangeduid met een groene rechthoek in Figuur 3.10), de andere 14 punten kunnen geassocieerd worden met de zeven minder, maar toch nog extreme waarden (aangeduid met een oranje rechthoek).

Uit Figuur 3.11 kunnen we afleiden dat de windsnelheden in Middelkerke gemiddeld hoger liggen dan de windsnelheden in Koksijde. Er is namelijk te zien dat meer punten van de dichte puntenwolk boven de eerste bissectrice liggen dan eronder.



Figuur 3.10: Visualisatie van windsnelheden voor detectie van foutieve outliers. Op de  $x$ -as staat het verschil in gemeten windsnelheid tussen Koksijde en Middelkerke weergegeven. Op de  $y$ -as staat de variabele `windschil` in Middelkerke. De kleurschaal geeft de windsnelheid in  $m/s$  aan in Middelkerke. De punten in de linkse rechthoeken komen overeen met metingen net vóór en net na metingen in de rechtse rechthoeken.



Figuur 3.11: Windsnelheden in Middelkerke uitgezet ten opzichte van de windsnelheden in Koksijde (1955 tot 2016).

Op het zicht kan men bij deze puntenwolk al zien dat er een sterke positieve correlatie lijkt te zijn tussen de windsnelheid in Middelkerke en de windsnelheid in Koksijde. Dit is in zekere zin ook wat we verwachten: beide kustgemeenten liggen op 19,58 km van elkaar verwijderd, dus de winden zullen in elk van de twee gemeenten ook min of meer dezelfde zijn. De Pearson correlatiecoëfficiënt is 0,87, een waarde die dicht bij één ligt. Het vermoeden dat beide variabelen sterk met elkaar gecorreleerd zijn, wordt hiermee dus bevestigd. Ook op deze plot (Figuur 3.11) zijn opnieuw de tien extreme waarden rechts te zien.

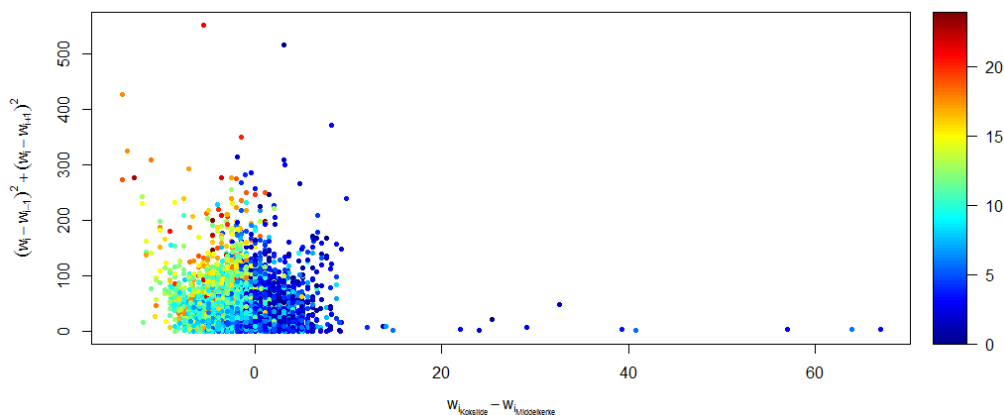
## Conclusie

Op basis van de analyse die werd gemaakt, kunnen we met vrij grote zekerheid stellen dat de extreme meetgegevens effectief foute waarden zijn. Een extra gegeven, dat onze conclusie kracht bijstaat, is het feit dat er zich, ondanks controles en correcties, nog foutieve metingen in de dataset bevinden. We kunnen er hier dus van uitgaan dat de twaalf extreme waarden van windsnelheden in Koksijde niet correct zijn en bijgevolg zetten we ze op NA voor verder onderzoek. Merk op dat de extreme waarden enkel voorkomen vóór 1979 en na 2013. Dit heeft waarschijnlijk te maken met het feit dat de datasets aan een kwaliteitscontrole onderworpen werden, maar dat deze controle tot nog toe slechts is uitgevoerd op de jaren 1980 – 2013.

De methode die hier gehanteerd werd om de dataset op te schonen, zijn ook in de literatuur terug te vinden. Zo wordt bijvoorbeeld in [47] al het volgende geschreven in verband met het identificeren van foute waarden in de dataset: “After a preliminary inspection, data that were obviously wrong (duplications, impossible wind directions) were easily identified and removed from the data set. Identification of incorrect velocity values was more problematic. All data suggesting three-hourly wind velocities greater than 40 knots have been checked in relation to the preceding and following values, to the corresponding daily maximum gust (when this value was available) and to the situation at nearby stations. Isolated, abnormally high velocity values were usually considered as errors, and therefore cancelled and transformed in missing observations.”

## Betrouwbaarheid meetgegevens Middelkerke

De meetgegevens uit Middelkerke lijken betrouwbaarder dan deze uit Koksijde wanneer we eenzelfde soort plot maken als bij Koksijde (zie Figuur 3.12). Er is hier ook te zien dat de variabele  $(w_i - w_{i-1})^2 + (w_i - w_{i+1})^2$  slechts gaat tot 500 terwijl die bij Koksijde loopt tot 3000. We besluiten hieruit dat er geen meetgegevens zijn die in die mate onbetrouwbaar zijn dat ze genegeerd of weggegooid kunnen worden. Op deze plot zijn uiterst rechtsonder ook weer de tien verdachte waarden te zien.

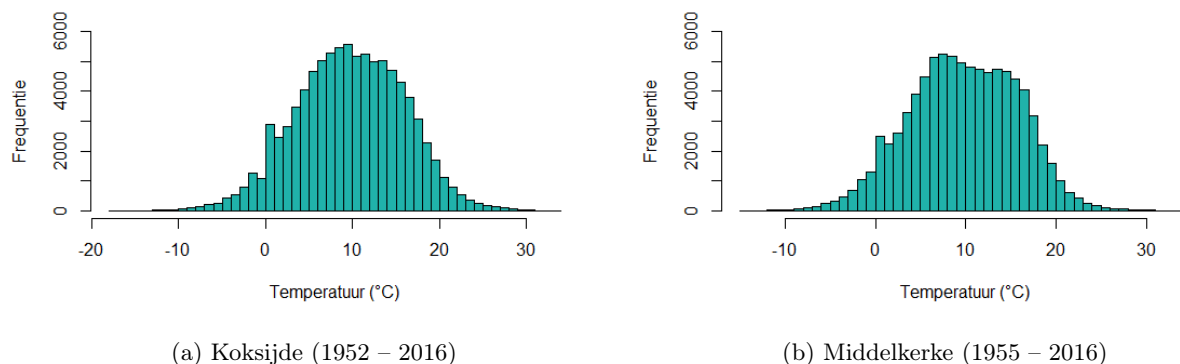


Figuur 3.12: Op de  $x$ -as staat het verschil in gemeten windsnelheid tussen Koksijde en Middelkerke weergegeven, op de  $y$ -as wordt de uitdrukking  $(w_i - w_{i-1})^2 + (w_i - w_{i+1})^2$  weergegeven en met een kleurenschaal wordt de gemeten windsnelheid in Middelkerke aangeduid.

### 3.3 Temperatuur

#### 3.3.1 Spectrum van gemeten waarden

Een eerste globaal beeld van de gemeten temperaturen in Koksijde en Middelkerke wordt gegeven door de histogrammen in Figuur 3.13a en 3.13b. Deze histogrammen vertonen geen overdreven vreemde of gekartelde vorm, maar het histogram van Middelkerke vertoont toch eerder een vreemde deuk. Het enige opvallende gegeven is dat er zowel in Koksijde als in Middelkerke meer metingen van  $0^{\circ}\text{C}$  werden geregistreerd dan men zou verwachten als men de verdeling van de temperatuur bekijkt. Een mogelijkheid zou kunnen zijn dat alleen het deel voor de komma in de databank werd opgeschreven, in plaats van algemeen aan metingen in het interval  $[x, x + 1[$  de waarde  $x$  te geven, ook voor  $x < 0$ . Dit zou betekenen dat temperaturen in  $] -1, 1[$  tot nul werden herleid wat zou verklaren dat er bijna dubbel zoveel metingen van  $0^{\circ}\text{C}$  zijn als verwacht op basis van de rest van het histogram.



Figuur 3.13: Histogram temperaturen

#### 3.3.2 Ontbrekende waarden

Er is geen temperatuur geregistreerd enerzijds op de momenten waarvan er geen enkele meting is geregistreerd in de dataset (Tabellen B.2 en B.1 in Bijlage B), anderzijds op die momenten waarop enkel de temperatuurwaarde ontbreekt, maar niet noodzakelijk de waarden van de andere meteorologische variabelen. In Tabel 3.14 wordt een overzicht gegeven van het aantal niet-geregistreerde waarden in beide datasets.

	Werkelijk aantal meetmomenten	NA's
Koksijde	95 301	47 (0,049%)
Middelkerke	90 918	16 (0,018%)
Totaal	186 219	63 (0,034%)

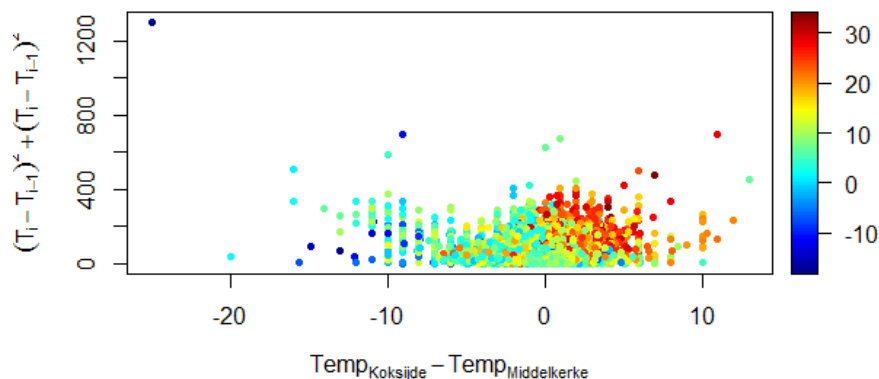
Tabel 3.14: Aantal niet-geregistreerde temperatuurmetingen in de datasets.

#### 3.3.3 Foutieve outliers

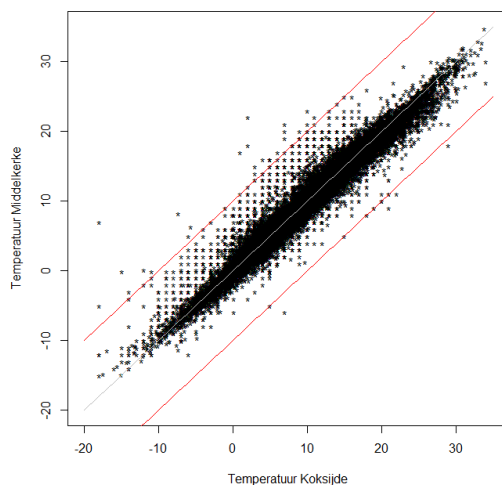
Aan de hand van Figuur 3.15 is slechts één duidelijke outlier te vinden in de dataset van Koksijde. Het gaat hierbij over het datapunt op 20 februari 1958 waarbij de temperatuur in Koksijde geregistreerd werd

als  $-18^{\circ}\text{C}$ , terwijl in Middelkerke de temperatuur  $6^{\circ}\text{C}$  bedroeg. In de database van extreme temperaturen is deze temperatuur niet terug te vinden op de betreffende dag. Wanneer men andere gerapporteerde extreem lage temperaturen gaat opzoeken, vindt men meestal waarden rond  $-16^{\circ}\text{C}$ . Het is dus vrijwel zeker dat deze waarde van  $-18^{\circ}\text{C}$  foutief is, anders was deze wel vermeld geweest in de database. Het vermoeden dat deze waarden foutief zijn, wordt gesterkt door het feit dat op 14 februari 1958 door het KMI het volgende wordt gerapporteerd: “*De minima zijn merkwaardig voor de maand februari: de thermometer zakt niet onder  $12,4^{\circ}\text{C}$  in Zaventem.*” Voor een verdere analyse van de temperatuur kan deze waarde dus op NA worden gezet. Verder zijn geen overduidelijk foutieve waarden op te merken.

Tot slot merken we nog dat in de dataset van zowel Koksijde als Middelkerke zeer hoge temperaturen terug te vinden zijn, met name tot  $34^{\circ}\text{C}$  in Koksijde en tot  $34,8^{\circ}\text{C}$  in Middelkerke. Alle dagen waarop dergelijke hoge temperaturen geregistreerd werden, staan ook in de databank van extreme temperaturen van het KMI. Het gaat hier dus zeker over juiste waarden.



Figuur 3.15: Op de  $x$ -as staat het verschil in temperatuur tussen Koksijde en Middelkerke, op de  $y$ -as staat voor Koksijde de uitdrukking  $(T_i - T_{i-1})^2 + (T_i - T_{i+1})^2$  weergegeven en de kleurschaal geeft de temperatuur in Koksijde weer.



Figuur 3.16: Temperaturen in Middelkerke ten opzichte van de temperaturen in Koksijde (1955 tot 2016).

Om een zicht te krijgen op eventuele onjuiste outliers bij de temperaturen in Koksijde en Middelkerke, zetten we

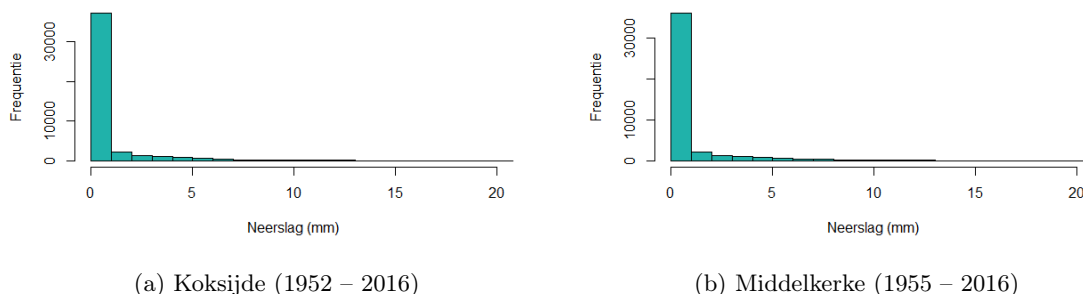
de temperaturen in Middelkerke uit ten opzichte van deze in Koksijde (zie Figuur 3.16). Tussen de rode rechten liggen alle datapunten waar het verschil in temperatuur tussen beide stations kleiner is dan  $10^{\circ}\text{C}$ . Buiten dit gebied is een groot aantal licht verdachte waarden te vinden, in plaats van een klein aantal manifest onjuiste waarden, zoals bij de windsnelheid. Deze grote verschillen in temperatuur zouden te maken kunnen hebben met de zeebries: in het geval dat er een zeebries waait in Middelkerke en niet in Koksijde (of omgekeerd), kan het temperatuurverschil tussen beide gemeenten heel groot zijn.

Aangezien deze thesis een overschouwende analyse is van het Belgische kustklimaat, valt een uitgebreid onderzoek van andere eventueel foutieve waarden buiten het bestek van deze thesis. Voor de analyse van temperatuur zal dan ook gebruik gemaakt worden van de volledige dataset met uitzondering van 20 februari 1958.

## 3.4 Neerslag

### 3.4.1 Spectrum van gemeten waarden

Door de histogrammen in Figuur 3.17 wordt een beeld gegeven van de frequentie van gemeten neerslaghoeveelheden. Hierbij valt op dat droge momenten overheersend zijn en dit zowel in Koksijde als Middelkerke.



Figuur 3.17: Histogrammen neerslaghoeveelheden

### 3.4.2 Ontbrekende waarden

In elk jaar zijn er veel ontbrekende waarden, meestal rond de 730 per jaar. Dit komt natuurlijk omdat neerslaghoeveelheden slechts twee keer per dag gerapporteerd worden, namelijk om 6 UTC en om 18 UTC. Metingen om 00 UTC en om 12 UTC zijn nooit beschikbaar.

Het is opvallend dat sinds het begin van de metingen (tot het jaar 2003) er nauwelijks ontbrekende waarden zijn. Figuur 3.18 geeft het percentage ontbrekende waarden aan per maand in Koksijde respectievelijk Middelkerke. Op beide figuren is duidelijk een overgang te zien. Wanneer we in de dataset gaan kijken, doen deze overgangen zich voor in oktober 2004.

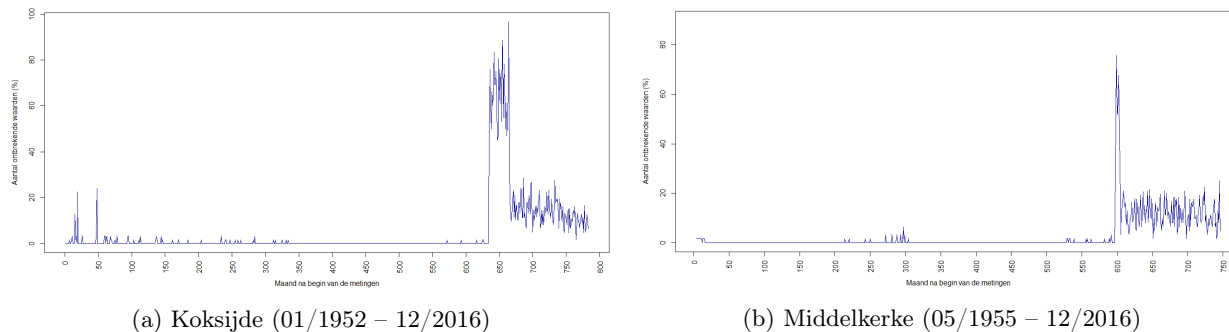
Het is in de periode oktober 2004 t.e.m. mei 2007 dat er opvallend veel waarden ontbreken in Koksijde<sup>2</sup>. Na mei 2007 neemt het aantal niet-geregistreerde waarden opnieuw aanzienlijk af. Toch blijft tot eind 2016 gemiddeld 13% van de data niet beschikbaar.

In Middelkerke is de periode met een groot aantal ontbrekende waarden korter: oktober 2004 t.e.m. mei 2005. Toch is ook hier na mei 2005 nog steeds een aanzienlijk aantal ontbrekende waarden, gemiddeld 11% per maand.

<sup>2</sup>Om dit fenomeen te kunnen verklaren, is een gedetailleerdere historiografie van het meetstation wenselijk.



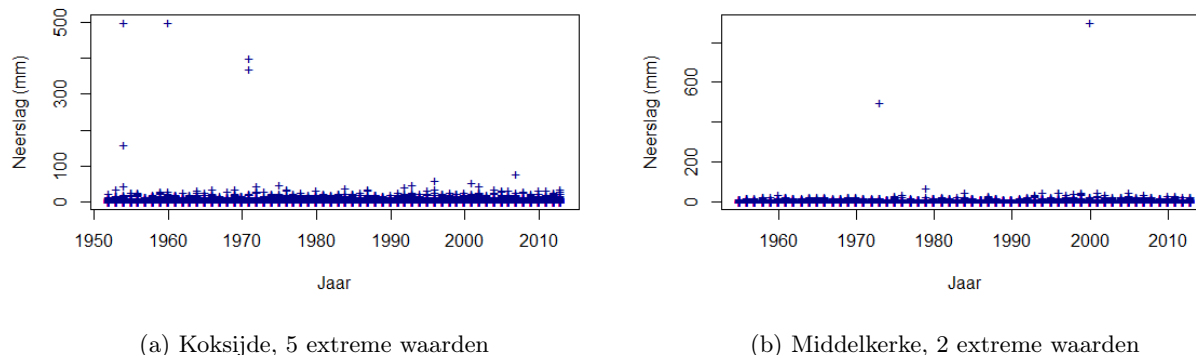
Het is niet volledig duidelijk waarom er vanaf 2004 een pak meer ontbrekende waarden zijn, maar een plausible verklaring kan gevonden worden in het feit dat de meettoestellen in de jaren negentig geautomatiseerd werden (persoonlijke communicatie met David Dehenauw per e-mail op 23 mei 2017 en met Vicky Schroeders, Meteo Wing luchtmacht Belgisch leger, per e-mail op 16 mei 2017). Mogelijks waren deze toestellen op termijn verouderd, want het moment waarop het aantal ontbrekende waarden afnam in Koksijde, valt samen met de installatie van een ‘juister en uitgebreider’ systeem (persoonlijke communicatie met Vicky Schroeders, Meteo Wing luchtmacht Belgisch leger, per e-mail op 16 mei 2017).



Figuur 3.18: Percentage niet-geregistreerde metingen van neerslaghoeveelheden per maand

### 3.4.3 Foutieve outliers

Om te bepalen of er zich eventueel verkeerde waarden voor gemeten neerslag in de datasets bevinden, blijkt het nuttig te zijn om een plot te maken waarbij de hoeveelheid neerslag wordt uitgezet tegenover de tijd (Figuur 3.19).



Figuur 3.19: Extreme waarden neerslag

In Tabel 3.20 staat weergegeven op welke momenten deze extreme neerslaghoeveelheden staan geregistreerd. Bovendien wordt ook weergegeven hoeveel neerslag op hetzelfde moment in de andere kustgemeente werd gemeten.

Er is geen twijfel mogelijk dat dit over foutief geregistreerde neerslaghoeveelheden gaat. Ten eerste wordt geen van deze dagen in de database van het KMI met extreme neerslag<sup>3</sup> vermeld. Bovendien geven statistieken [33] aan dat de gemiddelde *maandelijkse* neerslaghoeveelheid ligt tussen ongeveer 50 en 80 mm. Het feit

<sup>3</sup>Neerslaghoeveelheden die het rapporteren waard blijken, liggen rond 70 mm neerslag. Neerslagmetingen overschrijden niet vaak de 100 mm.

Dag	Uur	Station	Gerapporteerde neerslag (mm)	Neerslag ander station (mm)
23 mei 1954	18	Koksijde	<b>160</b>	NA
18 augustus 1954	06	Koksijde	<b>500</b>	NA
27 augustus 1960	06	Koksijde	<b>500</b>	0,3
4 februari 1971	18	Koksijde	<b>370</b>	0,0
17 november 1971	06	Koksijde	<b>400</b>	0,6
3 december 1973	06	Middelkerke	<b>500</b>	0,6
14 oktober 2000	18	Middelkerke	<b>902</b>	0,0

Tabel 3.20: Foutieve neerslaghoeveelheden in Koksijde en Middelkerke

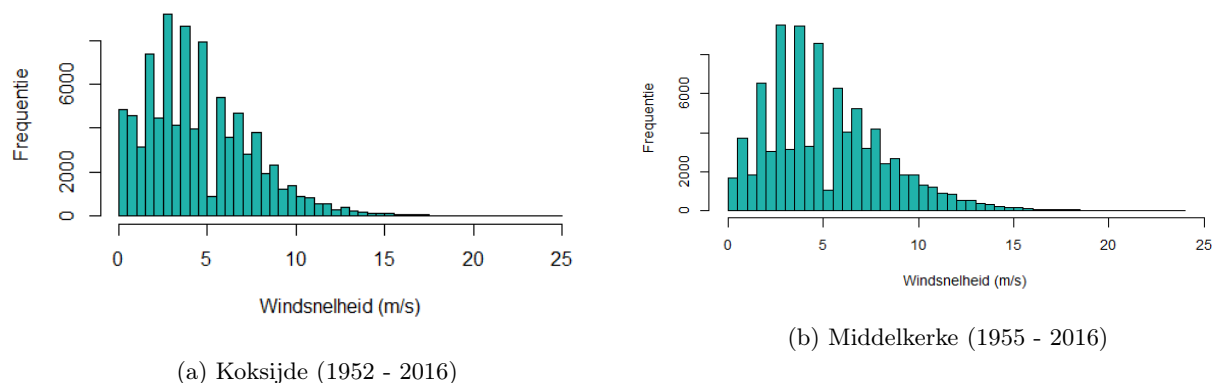
dat de neerslaghoeveelheden hier ruim de 200 mm overschrijden in een tijdspanne van 12 uur, lijkt zeer onwaarschijnlijk. Voor een analyse van de neerslag zetten we bovenstaande zeven waarden op NA.

Een andere twijfelachtige meting in Middelkerke is die van 15 februari 1979. Hier wordt 70 mm neerslag aangegeven. Deze dag staat ook niet vermeld in de KMI-database. In Koksijde wordt op hetzelfde moment 0,0 mm neerslag gemeten. Het zou dus kunnen dat ook dit een verkeerde waarde is in de dataset van Middelkerke. We kunnen dit echter niet met zekerheid zeggen, dus deze waarde wordt in de dataset gehouden.

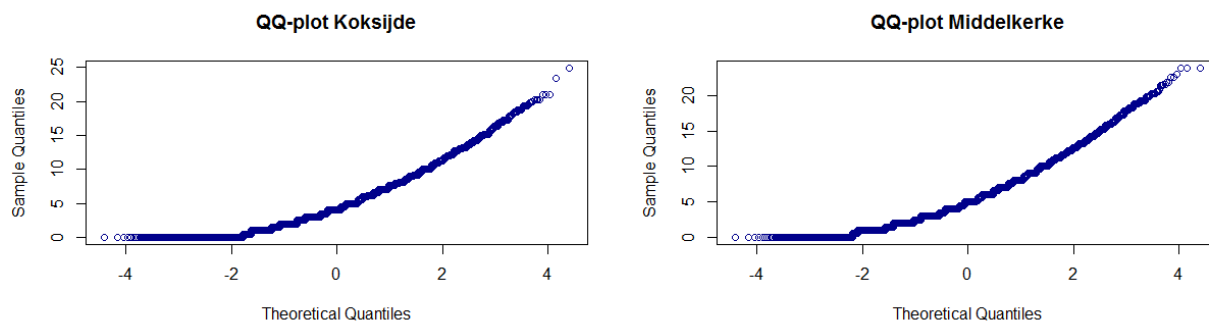
# Hoofdstuk 4

## Verdeling van de windsnelheden

Hoe zijn de windsnelheden verdeeld? Nog voor we dit onderzoeken, kunnen we met gezond verstand al vaststellen dat ze niet normaal verdeeld zullen zijn: het komt namelijk vaak voor dat er geen of weinig wind is, en maar in beperkte mate dat er veel wind is. Om te controleren of deze intuïtieve redenering wel degelijk correct is, maken we gebruik van een histogram van de windsnelheden in Koksijde en Middelkerke (Figuur 4.1). Deze geven een informatief beeld over de vorm en de ligging van de verdeling.



Figuur 4.1: Histogrammen van de windsnelheden



Figuur 4.2: Normale-QQ-plot van de windsnelheden in Koksijde (links) en Middelkerke (rechts).

Op een QQ-plot worden vervolgens ook nog eens de kwantielen van de verdeling van de gemeten windsnelheden uitgezet tegenover de kwantielen uit de theoretische normale verdeling. De QQ-plots die we bekommen voor Koksijde en Middelkerke zijn te zien in Figuur 4.2. Zoals te zien is op de figuren blijkt ons vermoeden te kloppen en zijn de windsnelheden niet normaal verdeeld. De vraag die spontaan rijst, is welke verdeling deze windsnelheden dan wel volgen. Uit de literatuur wordt duidelijk dat de Weibulldistributie hier een optie zou kunnen zijn [21, 22, 54]. Dit is een parametrische verdeling (2 parameters), die een veralgemening is van de Rayleighdistributie (1 parameter).

## 4.1 Rayleighdistributie

### 4.1.1 Definitie

De Rayleighdistributie is genoemd naar de Britse natuurkundige John William Strutt, Lord Rayleigh (1842–1919). Hij is de ontdekker van het element argon en kreeg hiervoor de Nobelprijs in 1904. Verder ontdekte hij het principe van *Rayleighverstrooiing* dat toegepast wordt bij de verklaring waarom de hemel blauw is. De Rayleighdistributie is de verdeling van de Euclidische norm van een tweedimensionale vector met onafhankelijke, identiek en normaal verdeelde coördinaten met gemiddelde nul. Veronderstel dat  $U$  en  $V$  onafhankelijke stochastische variabelen zijn met normale verdeling met gemiddelde nul en gelijke variantie  $\sigma^2$  groter dan nul. Dan heeft de variabele  $W = \sqrt{U^2 + V^2}$  een Rayleighdistributie met parameter  $\sigma$ . Deze distributie wordt soms gebruikt bij de modellering van windsnelheden waarbij de componenten op deze manier verdeeld zijn.



Figuur 4.3: John William Strutt

### Cumulatieve en kansdichtheidsfunctie

**Eigenschap 1.** De cumulatieve distributiefunctie van een Rayleigh verdeelde variabele met schaalparameter  $\sigma$  wordt gegeven door

$$F_\sigma(x) = 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

*Bewijs.* Om de distributiefunctie van  $W = \sqrt{U^2 + V^2}$  te vinden, bepalen we eerst de gezamenlijke kansdichtheidsfunctie van  $(U, V)$ . Hiervoor maken we gebruik van de kennis van de kansdichtheidsfuncties van  $U$  en  $V$ :

$$f_U(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{u^2}{2\sigma^2}\right) \quad \text{en} \quad f_V(v) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{v^2}{2\sigma^2}\right)$$

Doordat  $f_{U,V}(u,v) = f_V(v) \cdot f_{U|V}(u|v) = f_V(v) \cdot f_U(u)$  vinden we de gemeenschappelijke kansdichtheidsfunctie:

$$f_{U,V}(u,v) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{u^2 + v^2}{2\sigma^2}\right).$$

Vandaar dat

$$F_\sigma(x) = P(W \leq x) = \int_{C_x} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{u^2 + v^2}{2\sigma^2}\right) dudv, \quad \text{waarbij } C_x = \{(u,v) \in \mathbb{R}^2 : \sqrt{u^2 + v^2} \leq x\}.$$

Door het invoeren van poolcoördinaten ( $u = r \cos \theta$  en  $v = r \sin \theta$ ) verkrijgen we

$$F_\sigma(x) = P(W \leq x) = \int_0^{2\pi} \int_0^x \frac{1}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) r dr d\theta = \int_0^{2\pi} d\theta \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right]_0^x = 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Uitwerken van deze integraal geeft het gewenste resultaat. Hierbij is  $x$  de windsnelheid en  $\sigma$  de schaalparameter.  $\square$

De kansdichtheidsfunctie van de Rayleighverdeling met parameter  $\sigma$  wordt dan gegeven door de afgeleide

$$F'_\sigma(x) = f_\sigma(x) = \frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right), \text{ voor } x > 0. \quad (4.1)$$

### 4.1.2 Wiskundige analyse van cumulatieve en kansdichtheidsfunctie

**Eigenschap 2.** Zij  $X$  een Rayleigh verdeelde variabele met schaalparameter  $\sigma$ . Dan is

$$E[X] = \sigma \cdot \sqrt{\frac{\pi}{2}} \approx 1,253 \cdot \sigma.$$

*Bewijs.* Daar het gemiddelde van een stochastische variabele  $X$  met verdelingsfunctie  $f$  wordt gegeven door  $E[X] = \int_{-\infty}^{+\infty} xf(x)dx$ , berekenen we door achtereenvolgens substitutie ( $\frac{x}{\sigma} = t$ ) en partiële integratie toe te passen.

$$\begin{aligned} \int_0^{+\infty} x \cdot \frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx &= \sigma \int_0^{+\infty} t^2 \exp\left(-\frac{t^2}{2}\right) dt \\ &= \sigma \left( \left[ -t \exp\left(-\frac{t^2}{2}\right) \right]_0^{+\infty} + \int_0^{+\infty} \exp\left(-\frac{t^2}{2}\right) dt \right) \\ &= \sigma \frac{1}{2} \int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2}\right) dt \\ &= \sigma \frac{1}{2} \sqrt{2\pi} \\ &= \sigma \sqrt{\frac{\pi}{2}} \end{aligned} \quad \square$$

**Eigenschap 3.** De variantie van een Rayleigh verdeelde variabele wordt gegeven door  $\sigma^2 \cdot \left(\frac{4-\pi}{2}\right)$ .

*Bewijs.* We berekenen eerst  $E[X^2]$ .

$$\begin{aligned} E[X^2] &= \int_0^{+\infty} x^2 \cdot \frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx \\ &\stackrel{\text{Subs}}{=} \sigma^2 \int_0^{+\infty} y^3 \exp\left(-\frac{y^2}{2}\right) dy \\ &\stackrel{\text{PI}}{=} \sigma^2 \left\{ \left[ -y^2 \exp\left(-\frac{y^2}{2}\right) \right]_0^{+\infty} + \int_0^{+\infty} 2y \exp\left(-\frac{y^2}{2}\right) dy \right\} \\ &\stackrel{\text{Subs}}{=} 2\sigma^2 \int_0^{+\infty} \exp(-z) dz \\ &= 2\sigma^2 \cdot [-\exp(-z)]_0^{+\infty} \\ &= 2\sigma^2 \end{aligned}$$

Nu is  $\text{Var}[X] = E[X^2] - (E[X])^2 = 2\sigma^2 - (\sigma\sqrt{\frac{\pi}{2}})^2 = \sigma^2 \cdot \left(\frac{4-\pi}{2}\right)$ .  $\square$

**Eigenschap 4.** Zij  $X$  een Rayleigh verdeelde stochastische variabele met schaalparameter  $\sigma$ .

1.  $X$  kan enkel positieve waarden aannemen.
2. Ongeveer 39%  $\left(1 - \frac{1}{\sqrt{e}}\right)$  van de metingen worden verwacht kleiner te zijn dan  $\sigma$ .
3. De modus van  $X$  is  $\sigma$ .

*Bewijs.* 1. Kies  $x = 0$ . Dit betekent dat

$$F(x = 0) = 1 - \exp(0) = 0.$$

Concreet betekent dit dat alleen positieve waarden voor  $x$  mogelijk zijn omdat  $F(x = 0) = P(W \leq 0)$ .

2. Stel  $x$  gelijk aan de waarde van de parameter  $\sigma$ :

$$F(x = \sigma) = 1 - \exp\left(\frac{-\sigma^2}{2\sigma^2}\right) = 1 - \frac{1}{\sqrt{e}} \approx 0.39.$$

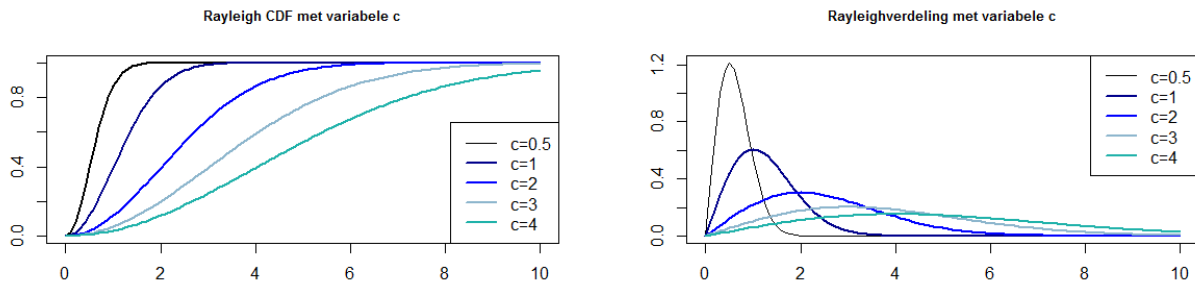
Hier kan men afleiden dat ongeveer 39 % van de gegevens kleiner zijn dan  $\sigma$ .

3. Om te bepalen waar de top van de dichtheidsfunctie zich bevindt, bepalen we waar de afgeleide van de kansdichtheidsfunctie nul wordt:

$$\begin{aligned} f'_\sigma(x) &= 0 \\ \Leftrightarrow \frac{1}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right) + \frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right) \cdot \left(\frac{-2x}{2\sigma^2}\right) &= 0 \\ \Leftrightarrow \frac{1}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right) \left[1 - \frac{x^2}{\sigma^2}\right] &= 0 \\ \Leftrightarrow 1 - \frac{x^2}{\sigma^2} &= 0 \\ \Leftrightarrow x = \sigma \quad (\text{want } \sigma, x > 0) \end{aligned}$$

We vinden dus dat de kansdichtheidsfunctie zijn maximum bereikt voor  $x = \sigma$ . □

**Invloed van de schaalparameter** De Rayleighdistributie heeft slechts één parameter, de schaalparameter  $\sigma$ , die verschillende waarden kan aannemen. Naargelang de waarde van  $\sigma$  gedraagt de distributie zich anders. Om inzicht te krijgen in de invloed van de schaalparameter op de vorm van de kansdichtheidsfunctie en cumulatieve distributiefunctie, werden in Figuur 4.4 de distributies geplott voor verschillende waarden van  $\sigma$ .



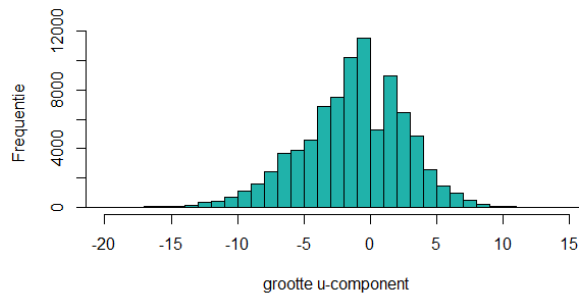
Figuur 4.4: Cumulatieve distributie (links) en Rayleighkansverdeling (rechts) voor verschillende waarden van de schaalparameter  $\sigma$ .

In de figuren laten we de waarde voor  $\sigma$  variëren van 0,5 tot 4. Dit is een arbitraire keuze, met als enige doel de invloed van de parameter  $\sigma$  te zien. In de linkse plot van Figuur 4.4 is te zien dat hoe groter we  $\sigma$  kiezen, hoe minder gepiekt de distributiefunctie wordt en hoe breder de staart wordt uitgesmeerd.

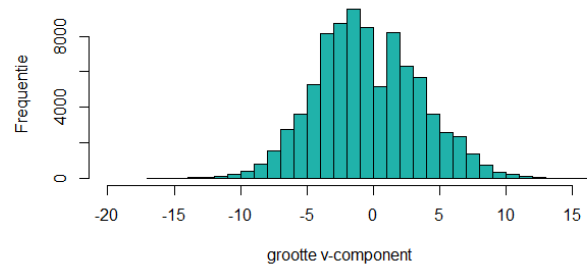
### 4.1.3 Geschikt voor modellering van windsnelheden?

#### Voorwaarden

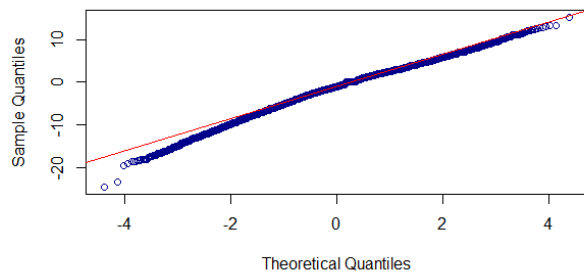
De Rayleighdistributie is theoretisch gebaseerd op twee variabelen die normaal verdeeld zijn rond nul, dezelfde variantie hebben en ongecorrleerd zijn. Om dit na te gaan, kan gebruik gemaakt worden van de orthogonale componenten van de windvector. De  $u$ -component stelt hierbij de windvectorcomponent uit het oosten voor (mogelijks negatief), de  $v$ -component die uit het noorden. In Figuur 4.5 worden een histogram en QQ-plot weergegeven van deze twee componenten. Een QQ-plot van beide variabelen laat zien dat de  $u$ -component niet normaal verdeeld is, terwijl de  $v$ -component vrij goed normaal verdeeld is.



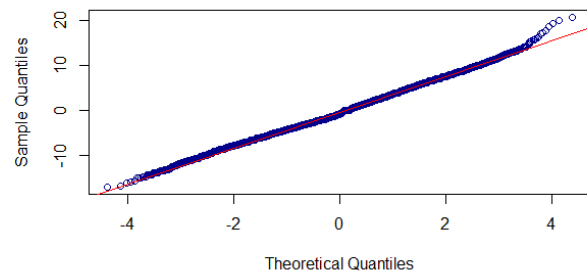
(a) Histogram  $u$ -component in m/s



(b) Histogram  $v$ -component in m/s



(c) Normale-QQ-plot van de  $u$ -componenten



(d) Normale-QQ-plot van de  $v$ -componenten

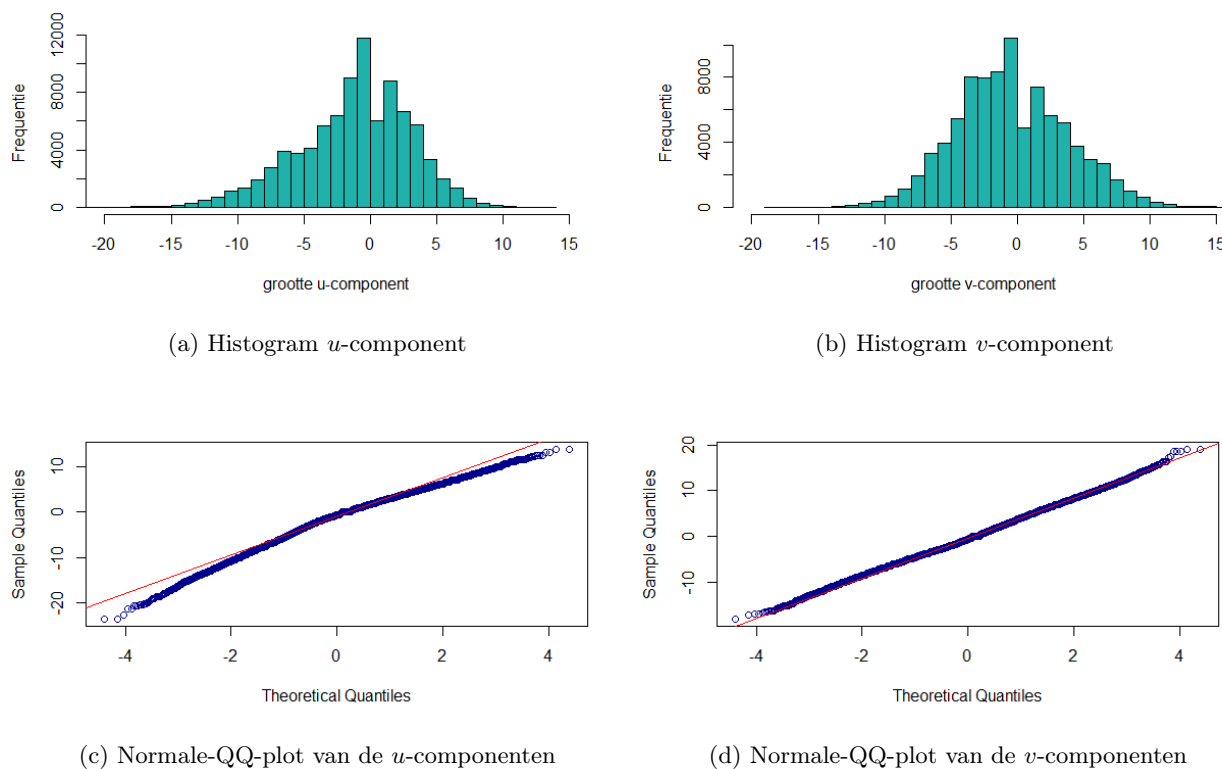
Figuur 4.5: Histogrammen en normale-QQ-plot componenten windvector in Koksijde

Voor Middelkerke vinden we zeer analoge resultaten. Een histogram en bijhorende QQ-plot voor beide componenten van de windvector zijn te vinden in Figuur 4.6.

Het gemiddelde en de standaarddeviatie voor de componenten  $u$  en  $v$  zijn terug te vinden in Tabel 4.7.

In de sectie 4.1.1 werden voorwaarden opgesomd waaraan de componenten van de windvector moesten voldoen om Rayleigh verdeeld te zijn. Een eerste voorwaarde is dat de componenten ongecorrleerd moeten zijn. Als correlatiecoëfficiënt vinden we in Koksijde 0,2247 ( $p < 10^{-15}$ ) en Middelkerke 0,221 ( $p < 10^{-15}$ ). Een correlatieplot wordt getoond in Figuur 4.8. Een tweede voorwaarde legt op dat de componenten normaal verdeeld moeten zijn met gemiddelde nul. De Anderson-Darlingtest voor normaliteit laat echter zien dat deze componenten niet normaal verdeeld zijn ( $p < 10^{-15}$ ). Tot slot moeten beide componenten dezelfde variantie hebben. De Barlett-test voor homogeniteit van de varianties toont echter dat ook hier de nulhypothese kan verworpen worden ( $p = 0,013$  in Koksijde,  $p = 0,024$  in Middelkerke) en de varianties dus niet gelijk zijn.

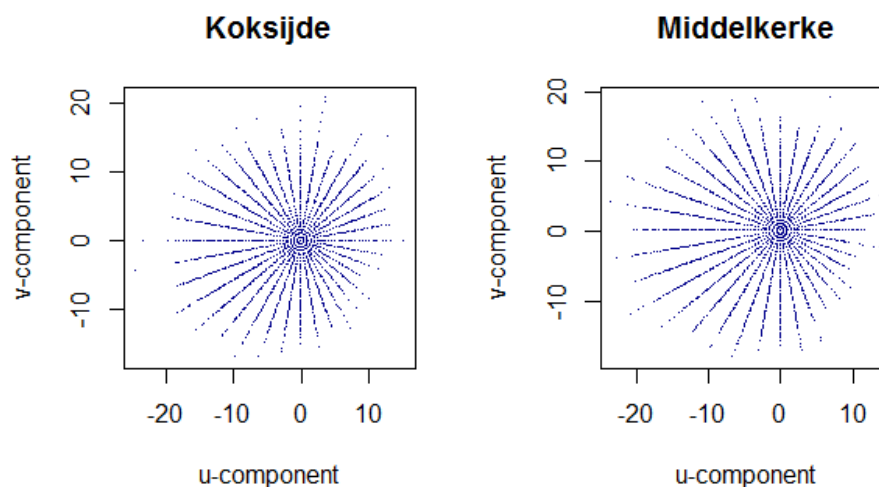
Uit bovenstaande gegevens kan afgeleid worden dat de Rayleighdistributie niet in aanmerking komt om wind in Koksijde en Middelkerke te modelleren, omdat de orthogonale componenten niet voldoen aan de voorwaarden.



Figuur 4.6: Histogram en QQ-plot voor componenten van de windvector in Middelkerke

Koksijde	$\bar{v}$	$s$	Middelkerke	$\bar{v}$	$s$
$u$	-1,22	3,89	$u$	-1,15	4,29
$v$	-0,37	3,93	$v$	-0,45	4,26

Tabel 4.7: Gemiddelde en standaarddeviatie van de componenten  $u$  en  $v$

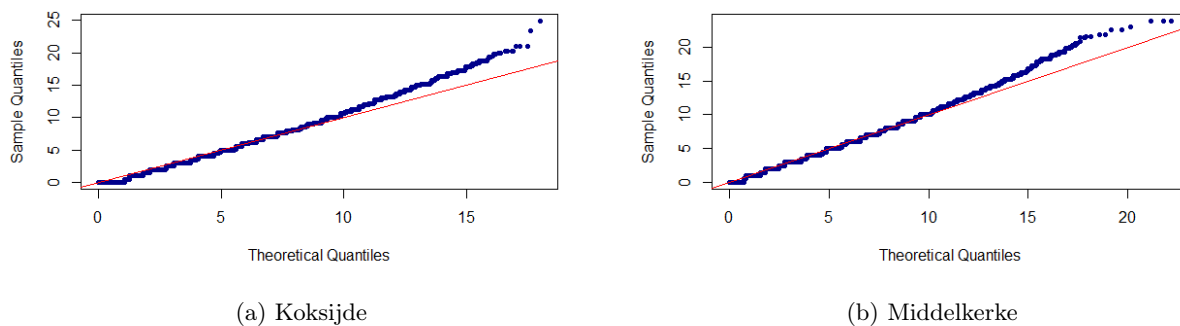


Figuur 4.8: Correlatieplot van de  $u$ - en  $v$ -component in Koksijde (links) en Middelkerke (rechts).



### QQ-plot

Een andere manier om te controleren dat de Rayleighdistributie niet de meest geschikte verdeling is om de windsnelheden te modelleren is aan de hand van een QQ-plot waarbij de kwantilen van de verdeling van de gemeten windsnelheden uitgezet worden tegenover de kwantilen uit de theoretische Rayleighverdeling. Hiervoor dienen we eerst de parameter  $\sigma$  te schatten die de Rayleighdistributie karakteriseert. Deze parameter kan geschat worden via een maximum-likelihoodmethode [64]. De schatting voor de parameter  $\sigma$  wordt gegeven door  $\sqrt{\left(\frac{\sum_i x_i^2}{2N}\right)} = 3,83$  in Koksijde en 4,31 in Middelkerke. In Figuur 4.9 worden de kwantilen van de verdeling van de gemeten windsnelheden uitgezet tegenover de kwantilen uit de theoretische Rayleighverdeling. Uit deze figuren is af te leiden dat de Rayleighverdeling niet de optimale verdeling is voor het modelleren van windsnelheden in Koksijde en Middelkerke en er aldus op zoek moet gegaan worden naar een andere verdeling om windsnelheden te modelleren.



Figuur 4.9: Rayleigh-QQ-plot van de windsnelheden

## 4.2 Weibullverdeling

De Weibullverdeling is een tweeparameterdistributie die vaak gebruikt wordt om windsnelheden te modelleren. In de literatuur zijn al heel wat verschillende verdelingen voorgesteld die gebruikt zouden kunnen worden om de verdeling van de windsnelheden te beschrijven. Er is gebleken dat de Weibullkansverdeling een goed model is om de uurgemiddelde windsnelheden op een bepaalde plaats te beschrijven [54]. We willen nagaan of ook de gemeten windsnelheden aan de kust door een Weibullverdeling kunnen beschreven worden.

De Weibullverdeling is genoemd naar de Zweedse ingenieur, wetenschapper en wiskundige Waloddi Weibull (1887 — 1979). Weibull werd professor in 1924 en behaalde zijn doctoraat aan de universiteit van Uppsala in 1932. In 1939 publiceerde hij zijn eerste werk over de Weibullkansverdelingsfunctie. In 1972 ontving Weibull de gouden medaille van de *American Society of Mechanical Engineers* aan wie hij 20 jaar eerder zijn belangrijkste werk over de Weibullverdeling had voorgelegd [65].



Figuur 4.10: Waloddi Weibull

### 4.2.1 Definitie

De twee parameters die de Weibulldistributie karakteriseren zijn de vormparameter (*shape parameter*)  $k$  en de schaalparameter (*scale parameter*)  $c$ . De schaalparameter  $c$  wordt uitgedrukt in m/s, terwijl de vormparameter  $k$  een dimensieloze grootte is. De cumulatieve verdelingsfunctie wordt gegeven door

$$F_{k,c}(x) = 1 - \exp\left[-\left(\frac{x}{c}\right)^k\right], \text{ voor } x \geq 0. \quad (4.2)$$

De cumulatieve distributiefunctie van de Rayleigh werd gegeven door

$$F_\sigma(x) = 1 - \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right], \text{ voor } x \geq 0$$

dus wanneer we  $k = 2$  en  $c = \sqrt{2}\sigma$  kiezen, krijgen we een bijzonder geval van Weibullverdeling, namelijk de Rayleighdistributie (zie 4.1). Door afleiden van (4.2) vinden we dat de kansdichtheidsfunctie van een Weibulldistributie wordt gegeven door

$$f_{k,c}(x) = \begin{cases} \left(\frac{k}{c}\right) \left(\frac{x}{c}\right)^{k-1} \exp\left[-\left(\frac{x}{c}\right)^k\right] & x \geq 0 \\ 0 & x < 0. \end{cases} \quad (4.3)$$

### 4.2.2 Wiskundige analyse van de cumulatieve en kansdichtheidsfunctie

Voor de analyses zullen we gebruik maken van de gammafunctie  $\Gamma(z) = \int_0^{+\infty} t^{z-1} \exp(-t) dt$ . Dit is een analytische voortzetting van de faculteitsfunctie naar de reële en complexe getallen, d.w.z. dat voor  $n \in \mathbb{N}$ :  $\Gamma(n) = (n-1)!$ .

#### Gemiddelde en variantie

Zij  $X$  een Weibull verdeelde stochastische veranderlijke met vorm- en schaalparameter  $k$  en  $c$ . Dan geldt dat

$$E[X] = c \cdot \Gamma\left(1 + \frac{1}{k}\right).$$

Het gemiddelde vinden we op de volgende manier

$$\begin{aligned} E[X] &= \int_0^{+\infty} x \cdot f_{k,c}(x) dx \\ &= \int_0^{+\infty} x \cdot \left(\frac{k}{c}\right) \left(\frac{x}{c}\right)^{k-1} \exp\left[-\left(\frac{x}{c}\right)^k\right] dx \end{aligned}$$

Aan de hand van de eenvoudige substitutie  $\left(\frac{x}{c}\right)^k = u$  wordt de integraal berekend:

$$\int_0^{+\infty} x \cdot \left(\frac{k}{c}\right) \left(\frac{x}{c}\right)^{k-1} \exp\left[-\left(\frac{x}{c}\right)^k\right] dx = c \cdot \int_0^{+\infty} u^{1/k} \exp(-u) du = c \cdot \Gamma\left(1 + \frac{1}{k}\right).$$

#### Eigenschap 5.

$$\text{Var}[X] = c^2 \left[ \Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2 \right]$$

*Bewijs.* Softwareberekening, aan de hand van  $\text{Var}[X] = E[X^2] - (E[X])^2$ . □

**Eigenschappen 6.** Zij  $X$  een Weibull verdeelde stochastische variabele met vormparameter  $k$  en schaalparameter  $c$ .

1.  $X$  kan enkel positieve waarden aannemen.
2. Ongeveer 63% ( $1 - \exp(-1)$ ) van de metingen wordt verwacht kleiner te zijn dan  $c$ .
3. De modus van  $X$  is gelijk aan  $\sqrt[k]{\frac{k-1}{k}} \cdot c$ .

*Bewijs.* 1. Stel  $v = 0$ , dan vinden we

$$F(x = 0) = 1 - \exp(0) = 0,$$

dus ook hier zijn enkel positieve waarden voor  $x$  mogelijk. Het is inderdaad zo dat windsnelheden enkel positief kunnen zijn.

2. Kiezen we  $x = c$ , dan komt er

$$F(x = c) = 1 - \exp(-1) \approx 0,63,$$

wat betekent dat dat kans ongeveer 63% is dat de variabele  $X$  waarden kleiner dan de parameter  $c$  aanneemt. Deze conclusie is onafhankelijk van  $k$ . Wanneer op eenzelfde figuur een aantal cumulatieve distributiefuncties worden getekend voor verschillende  $k$ , maar met eenzelfde  $c$  betekent dit dat deze grafieken elkaar snijden bij  $x = c$  (zie Figuur 4.12).

3. De modus is de top van de kansdichtheidsfunctie.

$$\begin{aligned} f'(x) &= \left(\frac{k}{c}\right) \left\{ (k-1) \left(\frac{x}{c}\right)^{k-2} \frac{1}{c} \cdot \exp\left[-\left(\frac{x}{c}\right)^k\right] + \left(\frac{x}{c}\right)^{k-1} \cdot \exp\left[-\left(\frac{x}{c}\right)^k\right] \left(-\frac{1}{c^k}\right) kx^{k-1} \right\} \\ &= \frac{k}{c} \exp\left[-\left(\frac{x}{c}\right)^k\right] \cdot \left[ \frac{(k-1)}{c} \left(\frac{x}{c}\right)^{k-1} - x^k \cdot k \cdot \frac{1}{c^k} \right] \end{aligned} \quad (4.4)$$

Uitdrukking (4.4) gelijk stellen aan nul en oplossen naar  $x$  geeft de  $x$ -coördinaat van de top van de dichtheidsfunctie:

$$\begin{aligned} \frac{k}{c} \exp\left[-\left(\frac{x}{c}\right)^k\right] \cdot \left[ (k-1) - x^k \cdot k \cdot \frac{1}{c^k} \right] &= 0 \\ \Leftrightarrow \left[ (k-1) - x^k \cdot k \cdot \frac{1}{c^k} \right] &= 0 \\ \Leftrightarrow x^k &= \frac{c^k(k-1)}{k} \\ \Leftrightarrow x &= \sqrt[k]{\frac{k-1}{k}} \cdot c \end{aligned}$$

□

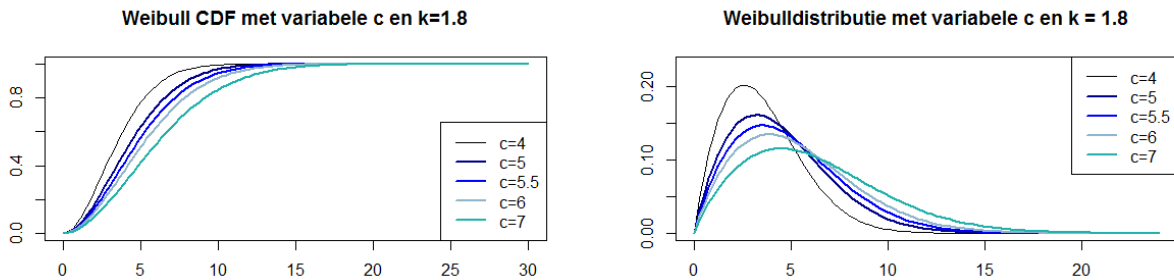
Voor  $k = 2$  komt er precies de uitdrukking  $\frac{c}{\sqrt{2}} = \sigma$  te staan die we als modus vonden bij de Rayleighdistributie.

In theorie kan de parameter  $k$  een willekeurig positief getal aannemen, wat betekent dat als  $k \rightarrow +\infty$ , dan  $\sqrt[k]{\frac{k-1}{k}} \rightarrow 1$ . Dit betekent dat de modus van een Weibullvariabele dichter bij de waarde van de parameter  $c$  komt te liggen naarmate  $k$  groter wordt.

Windsnelheden die een Weibulldistributie volgen, kunnen geen waarde nul aannemen. De kans dat de windsnelheid in een infinitesimale omgeving van nul ligt, is namelijk nul, want voor  $x = 0$  is er een horizontale raaklijn aan de kansdichtheidscurve. Naast  $f(x) = 0$  is immers ook

$$f'(x = 0) = \frac{k}{c} \exp\left[-\left(\frac{0}{c}\right)^k\right] \cdot \left[ \frac{(k-1)}{c} \left(\frac{0}{c}\right)^{k-1} - 0^k \cdot k \cdot \frac{1}{c^k} \right] = 0. \quad (4.5)$$

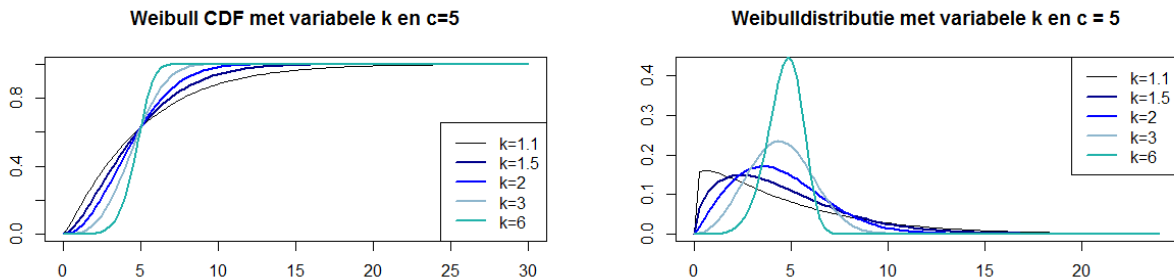
**Invloed van de schaalparameter en vormparameter** We laten de waarde voor de schaalparameter  $c$  variëren van 4 tot 7. In principe kunnen we elke willekeurige waarde voor  $c$  kiezen, zolang deze groter is dan 0. In het geval van de windmodellering zagen we in [54] dat vooral waarden in het interval  $[4,7]$  m/s realistisch blijken te zijn. In Figuur 4.11 staan de cumulatieve distributie en de verdelingsfunctie voor een aantal waarden van  $c$  weergegeven.



Figuur 4.11: Cumulatieve distributie (links) en Weibullkansdichtheid (rechts) voor verschillende waarden van de vormparameter  $c$  bij  $k = 1,8$ .

Figuur 4.12 laat de verschillen zien voor veranderende waarden van de vormparameter  $k$ . Omdat we ons in dit project concentreren op de modellering van de windsnelheid, kiezen we  $1 \leq k \leq 10$  [54]. Op beide figuren is Eigenschap 2 en Eigenschap 3 duidelijk zichtbaar:

- De verschillende distributiefuncties snijden elkaar bij  $x = c$ .
- Op de figuur links zien we ook dat de piek van de grafiek van de functie steeds dichterbij deze waarde  $c$  komt te liggen voor grotere waarden van  $k$ .



Figuur 4.12: Cumulatieve distributie (links) en Weibullkansdichtheid (rechts) voor verschillende waarden van de vormparameter  $k$ .

Uit Figuren 4.11 en 4.12 kunnen een aantal besluiten worden getrokken:

- Laten we  $k$  variëren en houden we  $c$  constant, dan wordt, voor toenemende  $k$ , de piek van de dichtheidsfunctie meer uitgesproken, terwijl de ‘staarten’ minder breed worden uitgesmeerd.
- Laten we  $c$  variëren en houden we  $k$  constant, dan schuift met toenemende waarden voor  $c$  de dichtheidsfunctie geleidelijk op naar rechts zodat de modus rond  $c$  zit.

In wat volgt zullen we vaststellen dat de Weibullverdeling geschikt is om de windsnelheden aan de Belgische kust te modelleren.

### 4.3 Schatten van de vorm- en schaalparameter

In de literatuur worden verschillende methoden gebruikt om de Weibullparameters  $k$  en  $c$  te schatten. De methoden die het vaakst worden gebruikt zijn de momentmethode en de maximum-likelihoodmethode [42, 35, 54]. Voor het schatten van de parameters  $k$  en  $c$  die het best de verdeling van de windsnelheden in Koksijde en Middelkerke beschrijven, zullen we deze beide methoden gebruiken en de resultaten met elkaar vergelijken.

Andere gebruikte methoden zijn de energy pattern factor methode [22, 2] en de least squares methode [42].

#### 4.3.1 De momentmethode

Wanneer het steekproefgemiddelde  $\bar{X}$  (eerste moment) en de steekproefvariantie  $s$  (tweede moment) van de windsnelheden beschikbaar zijn, kan deze methode worden toegepast. Het is net omdat de parameters  $k$  en  $c$  gerelateerd zijn aan het gemiddelde en de variantie, dat deze momenten gebruikt kunnen worden om de parameters te schatten. We weten dat

$$\begin{cases} E[X] = c \cdot \Gamma\left(1 + \frac{1}{k}\right) & (4.6) \\ \text{Var}[X] = c^2 \cdot \left[ \Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2 \right] & (4.7) \end{cases}$$

Eliminatie van  $c$  uit 4.7 geeft

$$\frac{\text{Var}(X)}{E(X)^2} = \frac{\Gamma\left(1 + \frac{2}{k}\right)}{\left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2} - 1.$$

Volgens [54, 24] kan deze uitdrukking benaderd worden door een macht van  $k$ :

$$\left(\frac{\sqrt{\text{Var}(X)}}{E[X]}\right)^2 = \frac{\Gamma\left(1 + \frac{2}{k}\right)}{\left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2} - 1 \approx k^{-2 \cdot 1.086^{-1}}. \quad (4.8)$$

Omdat  $s^2$  een schatter is voor  $\text{Var}[X]$ , en  $\bar{X}$  voor  $E[X]$ , zal  $\frac{s}{\bar{X}}$  een schatter zijn voor  $\frac{\sqrt{\text{Var}[X]}}{E[X]} = \sqrt{\frac{\Gamma\left(1 + \frac{1}{k}\right)}{\Gamma\left(1 + \frac{1}{k}\right)^2}} \approx k^{-1.086^{-1}}$ , waaruit  $k$  kan geschat worden:

$$\hat{k} = \left(\frac{s}{\bar{X}}\right)^{-1.086}.$$

Eenmaal  $\hat{k}$  bepaald is, kan de parameter  $\hat{c}$  eenvoudig bepaald worden uit vergelijking (4.6):

$$\hat{c} = \frac{\bar{X}}{\Gamma\left(1 + \frac{1}{\hat{k}}\right)} = \frac{\bar{X}}{\int_0^{+\infty} t^{\frac{1}{\hat{k}}} \exp(-t) dt}. \quad (4.9)$$

Dit is een empirische methode op basis van de momentmethode [54]. Het voordeel van deze methode is dat de parameters  $k$  en  $c$  zeer snel en eenvoudig geschat kunnen worden. Enkel  $s$  en  $\bar{X}$  zijn nodig.

De geraadpleegde literatuur blijft echter op de vlakte wat betreft de kwaliteiten die deze schatters vertonen, zoals consistent of (asymptotisch) onvertkend.

#### 4.3.2 De maximum-likelihoodmethode

De maximum-likelihoodmethode is de meest algemeen gebruikte methode om parameters te schatten. Deze methode bepaalt de parameters zodanig dat de waarschijnlijkheid om de data die je hebt te observeren maximaal is onder alle mogelijke andere waarden die de parameters kunnen aannemen.

Deze methode kan eveneens gebruikt worden om de parameters  $k$  en  $c$  van de Weibullverdeling te schatten. De parameters kunnen gevonden worden door de (log-)waarschijnlijkheidsfunctie te maximaliseren.

Nemen we de natuurlijke logaritme van de Weibuldichtheidsfunctie (4.3), dan vinden we de volgende uitdrukking, voor  $c \geq 0$ :

$$\ln f_{k,c}(x) = \ln k - \ln c + (k-1)(\ln x - \ln c) - \left(\frac{x}{c}\right)^k \quad (4.10)$$

De log-waarschijnlijkheidsfunctie  $\ell(k,c)$  wordt gegeven door

$$\begin{aligned} \ell(k,c) &= \ln \prod_{i=1}^n f_{k,c}(x_i) \\ &= \sum_{i=1}^n \ln f_{k,c}(x_i) \\ &= \sum_{i=1}^n \ln k - \ln c + (k-1)(\ln x_i - \ln c) - \left(\frac{x_i}{c}\right)^k \\ &= n \ln k - kn \ln c + (k-1) \sum_{i=1}^n \ln(x_i) - \sum_{i=1}^n \left(\frac{x_i}{c}\right)^k. \end{aligned}$$

Deze log-waarschijnlijkheidsfunctie bereikt een maximum voor welbepaalde  $\hat{c}$  en  $\hat{k}$ , die we zoeken. Dit gebeurt wanneer de afgeleide van  $\ell(k,c)$  naar  $k$  en  $c$  nul wordt. We zoeken dus een oplossing van het stelsel

$$\begin{cases} \frac{\partial \ell(k,c)}{\partial k} = 0 \\ \frac{\partial \ell(k,c)}{\partial c} = 0. \end{cases}$$

Met name:

$$\begin{cases} \frac{\partial \ell(k,c)}{\partial k} = \frac{n}{k} - n \ln c + \sum_{i=1}^n \ln(x_i) - \sum_{i=1}^n \left(\frac{x_i}{c}\right)^k \cdot \ln\left(\frac{x_i}{c}\right) = 0 \\ \frac{\partial \ell(k,c)}{\partial c} = \frac{-kn}{c} + k \sum_{i=1}^n x_i^k \frac{1}{c^{k+1}} = 0 \end{cases} \quad (4.11)$$

$$\frac{\partial \ell(k,c)}{\partial c} = \frac{-kn}{c} + k \sum_{i=1}^n x_i^k \frac{1}{c^{k+1}} = 0 \quad (4.12)$$

Uit deze laatste gelijkheid kan de parameter  $c$  bepaald worden:

$$c = \left( \frac{1}{n} \sum_{i=1}^n x_i^k \right)^{\frac{1}{k}}, \quad (4.13)$$

maar deze is afhankelijk van  $k$ . Door  $c$  te elimineren uit (4.11) met uitdrukking (4.13) kan een betrekking voor de parameter  $k$  gevonden worden.

$$k = \left[ \frac{\sum_{i=1}^n x_i^k \ln(x_i)}{\sum_{i=1}^n x_i^k} - \frac{\sum_{i=1}^n \ln(x_i)}{n} \right]^{-1}. \quad (4.14)$$

Dit is een impliciete betrekking voor  $k$ , daar  $k$  ook in het rechterlid voorkomt. In de praktijk wordt deze  $\hat{k}$  bepaald door een iteratief proces. Eens  $\hat{k}$  is gevonden, wordt  $\hat{c}$  bepaald door middel van (4.13).

### 4.3.3 De power-densitymethode

De power-densitymethode is een meer recente methode (2008) om de Weibullparameters te schatten, gebaseerd op de verhouding van het gemiddelde van de derde machten van de geobserveerde windsnelheden en de derde macht van hun gemiddelden. Blijkbaar is deze grootte, die *energy pattern factor* wordt genoemd, een welbekende grootte in de windmolenindustrie.

Doordat

$$\begin{aligned} E[x^3] &= \int_0^{+\infty} x^3 \cdot f_{k,c}(x) dx \\ &= \int_0^{+\infty} x^3 \cdot \left(\frac{k}{c}\right) \left(\frac{x}{c}\right)^{k-1} \exp\left[-\left(\frac{x}{c}\right)^k\right] dx \\ \left(\text{Substitutie } \left(\frac{x}{c}\right)^k = u\right) &= c^3 \cdot \int_0^{+\infty} u^{3/k} \exp(-u) du \\ &= c^3 \cdot \Gamma\left(1 + \frac{3}{k}\right) \end{aligned}$$

is de energy pattern factor

$$\frac{E[x^3]}{(E[x])^3} = \frac{\Gamma\left(1 + \frac{3}{k}\right)}{\Gamma\left(1 + \frac{1}{k}\right)^3} \approx \frac{3,69}{k-1},$$

waarbij deze laatste stap een numerieke benadering is zoals gebruikt in [2]. Als in de praktijk de *energy pattern factor* geschat wordt als  $\frac{\bar{x}^3}{(\bar{x})^3}$ , kunnen we de Weibullparameter  $k$  schatten als

$$\hat{k} = 1 + \frac{3,69}{E_{pf}}.$$

Daarna kan de parameter  $c$  geschat worden zoals met de momentmethode, met name door vergelijking (4.9), zoals Akdağ en Dinler voorstellen, of zoals met de maximum-likelihoodmethode, door betrekking (4.13).

Resultaten van de studie van Akdağ et al. [2] stellen dat deze methode erg geschikt is om de parameters te schatten, volgens hen zelfs beter dan de andere vaak gebruikte methoden zoals de maximum-likelihoodmethode en de momentmethode, voornamelijk door lichtere computationele eisen. De geraadpleegde literatuur is echter zeer onduidelijk wat betreft de kwaliteiten die deze schatters vertonen op vlak van consistentie of (asymptotische) onvertekendheid.

## 4.4 Schatten van de Weibullparameters voor de windsnelheden in Koksijde en Middelkerke

### 4.4.1 Schatting met de momentmethode

In wat volgt zullen we per seizoen gaan kijken naar de verdeling van de gemeten windsnelheden. Er wordt gewerkt met acht subdatasets: één dataset per seizoen, voor Koksijde van 1952 tot 2016, voor Middelkerke van 1955 tot 2016.

Voor de windsnelheid zijn er gegevens beschikbaar om de zes uur. Dit betekent dat voor alle jaren van 1952 tot en met 2016 ongeveer  $n \approx 95\,000$  windsnelheden beschikbaar zijn. Van al deze gegevens berekenen we de

gemiddelde windsnelheid  $\bar{v}$  en de steekproefvariantie  $s^2$ , met  $v_i$  de windsnelheid op tijdstip  $i$ :

$$\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i = 4,605 \text{ m/s}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})^2 = 8,196 \text{ (m/s)}^2.$$

Daarmee vinden we

$$\hat{k} = \left(\frac{s}{\bar{v}}\right)^{-1,086} = 1,676.$$

De voorwaarde dat  $\hat{k}$  gelegen moet zijn tussen 1 en 10 is hier voldaan. Vervolgens kan ook de schaalparameter  $c$  geschat worden. Deze wordt berekend aan de hand van uitdrukking (4.9):

$$\hat{c} = \frac{\bar{v}}{\int_0^{+\infty} t^{1/\hat{k}} \exp(-t) dt} = \frac{4,605}{0,893} = 5,157,$$

waarbij  $\int_0^{+\infty} t^{1,676^{-1}} \exp(-t) dt$  numeriek berekend werd.

Hier vonden we dus dat de verdeling van de windsnelheden in Koksijde van 1952 tot 2016 het best gekarakteriseerd worden door een Weibullverdeling met als parameters

$$\hat{k} = 1,676$$

$$\hat{c} = 5,157.$$

Voor Middelkerke is  $\bar{v} = 5,309 \text{ m/s}$  en  $s^2 = 9,152 \text{ (m/s)}^2$ . Dit betekent dat we als schattingen voor  $k$  en  $c$  de volgende waarden vinden

$$\hat{k} = \left(\frac{3,025}{5,309}\right)^{-1,086} = 1,842$$

$$\hat{c} = \frac{5,309}{\int_0^{+\infty} t^{1,842^{-1}} \exp(-t) dt} = 5,979.$$

Dezelfde procedure kan nu herhaald worden voor de lente, zomer, herfst en winter in Koksijde en Middelkerke. De resultaten hiervan zijn terug te vinden in Tabel 4.13.

In Tabel 4.13 kan gezien worden dat er grotere windsnelheden met meer variabiliteit geschat worden in Middelkerke, en dat voor elk seizoen.

#### 4.4.2 Schatting met de maximum-likelihoodmethode

**Metingen van windstilte** Met de maximum-likelihoodmethode wordt  $k$  geschat door middel van uitdrukking (4.14) waarin  $\ln(v_i)$  voorkomt. Echter, bij het berekenen van  $\sum_{i=1}^n \ln(v_i)$  zou  $-\infty$  bekomen worden wanneer  $v_i = 0$ . In de literatuur [24, 22, 35] staat daarom geschreven dat enkel de niet-nul windsnelheden gebruikt moeten worden om de parameters te schatten en dus dat  $n$  gelijk moet zijn aan het aantal van nul verschillende windsnelheden in de dataset. In de datasets van Koksijde en Middelkerke komen echter wel heel wat nulmetingen voor (zie Tabel 4.14), hoewel de theoretische kans om een windsnelheid van ongeveer 0 m/s te vinden nul is (zie 4.5). Doordat de kleinst geregistreerde windsnelheid groter dan 0 m/s al 0,5 m/s is, laat dit uitschijnen dat het meettoestel een detectielimiet heeft. Een mogelijkheid zou zijn dat een windsnelheid lager dan 0,5 m/s door het toestel niet meer gedetecteerd wordt en dat deze windsnelheid gelijk gesteld wordt aan nul. Een andere mogelijkheid is dat windsnelheden groter of gelijk aan 0,25 m/s gelijkgesteld worden aan 0,5 m/s en dat windsnelheden kleiner dan 0,25 m/s ingegeven worden als 0 m/s.



Koksijde	$\bar{v}$	$s^2$	$\hat{k}$	$\hat{c}$
Lente	4,603	7,542	1,75	5,17
Zomer	4,001	5,861	1,73	4,49
Herfst	4,627	8,770	1,62	5,16
Winter	5,194	9,918	1,72	5,82
Jaar	4,605	8,196	1,676	5,157
Middelkerke	$\bar{v}$	$s^2$	$\hat{k}$	$\hat{c}$
Lente	5,226	7,981	1,95	5,89
Zomer	4,777	6,654	1,95	5,39
Herfst	5,381	10,295	1,75	6,04
Winter	5,861	11,130	1,84	6,60
Jaar	5,309	9,152	1,842	5,979

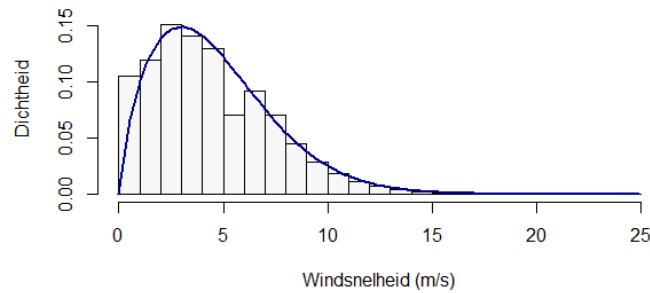
Tabel 4.13: Weibullparameters berekend aan de hand van de momentmethode.  $\bar{v}$  is de gemiddelde windsnelheid in het aangegeven seizoen (over alle jaren),  $s^2$  is de variantie ervan.

	Aantal meetmomenten	Aantal metingen met $v = 0$ m/s
Koksijde	95 301	3586 (3,76%)
Middelkerke	90 918	1301 (1,43%)
Totaal	186 219	4887 (2,62%)

Tabel 4.14: Aantal momenten in de datasets van Koksijde en Middelkerke waarop een windsnelheid van 0 m/s werd geregistreerd

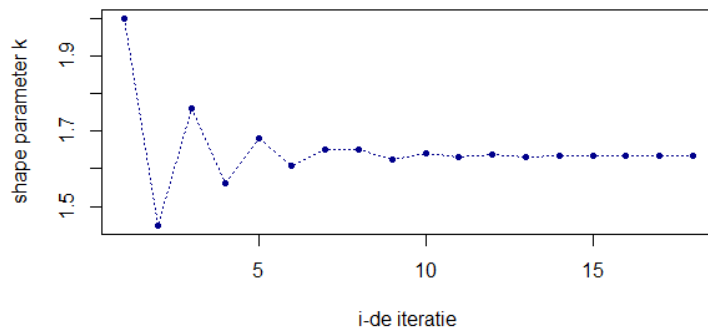
Op het eerste zicht lijkt het vreemd om geen rekening te houden met de windsnelheden die gelijk zijn aan nul. Dit zou betekenen dat in het geval van de dataset Koksijde 3,76% van de gegevens wordt ‘weggegooid’. Het lijkt echter niet te verantwoorden om deze waarden niet in rekening te brengen. We weten namelijk dat al deze nulwaarden in werkelijkheid verschillen van nul. De waarden die de gemeten windsnelheden aannemen, lijken zich te bevinden in een discreet spectrum (zie sectie 3.2.1). 0,5 m/s komt overeen met ongeveer 1 knoop. Mogelijks werden waarden kleiner dan 1 knoop automatisch gelijkgesteld aan nul. Hierdoor wordt de informatie onvolledig en onnauwkeurig. Een oplossing hiervoor (persoonlijke communicatie met Els Goetghebeur op 17 maart 2017) is om de waarden die precies gelijk zijn aan nul in de dataset te vervangen door gesimuleerde waarden tussen 0 en 0,5 die realistisch zijn als de windsnelheden komen uit een Weibullverdeling met parameters  $\hat{k}$  en  $\hat{c}$ . Deze twee parameters werden reeds geschat aan de hand van momentmethoden:  $\hat{k} = 1,676$  en  $\hat{c} = 5,157$  voor Koksijde en  $\hat{k} = 1,842$  en  $\hat{c} = 5,979$  voor Middelkerke. De nulwaarden in de dataset kunnen vervolgens vervangen worden door de realistische random gegenereerde waarden.

Om na te gaan of deze methode te rechtvaardigen is, kan de situatie gesimuleerd worden. Vooreerst wordt de volledige dataset voor Koksijde (analoog voor Middelkerke) op een random manier opgesplitst in twee delen. In het eerste deel van de dataset worden de nulwaarden van de windsnelheid vervangen door de random gegenereerde waarden tussen 0 en 0,5 komende uit een Weibull( $\hat{k}, \hat{c}$ ). Op basis van deze ‘nieuwe’ dataset worden de parameters geschat volgens de methode beschreven in 4.3.2. Deze schatters zijn  $\hat{k} = 1,667$  en  $\hat{c} = 5,185$  voor Koksijde. In een volgende stap wordt een histogram van het andere deel van de dataset gemaakt. Hierboven zal de Weibullkansverdelingsfunctie – met als parameters  $\hat{k}$  en  $\hat{c}$  die op basis van het eerste deel van de dataset werden geschat – geplotted worden. Voor Koksijde is dit resultaat te zien in Figuur 4.15. Uit deze figuur kunnen we opmaken dat de parameters, geschat op basis van een dataset met random gegenereerde waarden tussen 0 en 0,5 m/s, ook de windsnelheden in het tweede deel van de dataset, waarbij de windsnelheden nog steeds de



Figuur 4.15: Gefitte (met gesimuleerde waarden rond 0) Weibullkansverdeling door maximum-likelihoodschatting op basis van het eerste deel, getekend op een histogram van het tweede deel

oorspronkelijke waarden zijn, nog goed fitten. Hieruit kunnen we afleiden dat het inderdaad te rechtvaardigen is om alle nulwaarden te vervangen door random gegenereerde waarden tussen 0 en 0,5 m/s, en dit om de maximum-likelihoodmethode te kunnen toepassen<sup>1</sup>.



Figuur 4.16: Schatting van de vormparameter  $k$  na een aantal iteraties.

minder dan  $10^{-4}$  van elkaar verschillen. De waarde voor de parameters wordt gegeven tot op één honderdste nauwkeurig omdat dit bij de momentmethode ook zo werd gedaan.

In eerste instantie schatten we de parameters  $k$  en  $c$  voor de volledige dataset in Koksijde. Deze bevat  $n = 95\,225$  waarden. Door een iteratief proces vinden we dat  $\hat{k} = 1,63$ . De opeenvolgende waarden die voor  $k$  werden gevonden zijn terug te vinden in Figuur 4.16. Vervolgens vinden we dat  $\hat{c} = 5,15$ .

De waarden voor de schatters werden eveneens berekend met behulp van de package `fitdistr` in R, gebaseerd op maximum-likelihoodfitting van univariate distributies. De maximum-likelihoodmethode en de via software berekende waarden zijn exact dezelfde tot op één honderdste nauwkeurig. Dit komt omdat de methode die door de software gebruikt wordt eveneens de maximum-likelihoodmethode is.

Een overzicht van de gevonden schatters staat in Tabel 4.17.

<sup>1</sup>Na het uitvoeren van de maximum-likelihoodmethode wordt weer verdergewerkt met de oorspronkelijke dataset met veel nulwaarden.

Om de Weibullparameters  $k$  en  $c$  te schatten, zal  $\hat{k}$  eerst bepaald worden uit vergelijking (4.14). Dit gebeurt aan de hand van een iteratie. Volgens [22] is 2 een eerste goede gok voor de parameter  $k$ . Vervolgens laten we  $\hat{k}$  iteratief berekenen tot het verschil tussen twee opeenvolgende waarden van  $\hat{k}$  onder een voorafgelegde threshold duikt. Eens we een goede  $\hat{k}$  gevonden hebben, kunnen we deze gebruiken om de parameter  $\hat{c}$  te berekenen aan de hand van vergelijking (4.13).

In dit geval wordt gekozen de schatter voor de parameter  $k$  als accuraat genoeg te beschouwen wanneer twee opeenvolgende waarden voor  $k$  tijdens de iteratie

	Momentmethode		Maximum likelihoedmethode		Power-densitymethode	
	$\hat{k}$	$\hat{c}$	$\hat{k}$	$\hat{c}$	$\hat{k}$	$\hat{c}$
<b>Koksijde</b>						
<b>Lente</b>	1,75	5,17	1,70	5,15	-	-
<b>Zomer</b>	1,73	4,49	1,68	4,48	-	-
<b>Herfst</b>	1,62	5,16	1,58	5,15	-	-
<b>Winter</b>	1,72	5,82	1,68	5,81	-	-
<b>Jaar</b>	1,68	5,16	1,63	5,15	2,36	5,20
<b>Middelkerke</b>						
<b>Lente</b>	1,95	5,89	1,93	5,90	-	-
<b>Zomer</b>	1,95	5,39	1,93	5,39	-	-
<b>Herfst</b>	1,75	6,04	1,74	6,05	-	-
<b>Winter</b>	1,84	6,60	1,83	6,60	-	-
<b>Jaar</b>	1,84	5,97	1,84	5,98	2,13	5,99

Tabel 4.17: Overzicht van de verschillende schattingen voor de vorm- en schaalparameter in Koksijde en Middelkerke op twee manieren: momentmethode en maximum-likelihoedmethode. De parameters werden eveneens berekend met behulp van software, maar deze resultaten waren tot op twee cijfers na de komma identiek aan de schattingen via de maximum-likelihoedmethode en worden daarom niet in de tabel getoond.

#### 4.4.3 Schatting met de power-densitymethode

In sectie 4.3 werd kort een derde methode voorgesteld die gebruikt kan worden om de Weibullparameters te schatten. Op basis van de volledige dataset voor Koksijde en Middelkerke werden aan de hand van de power-densitymethode de parameters geschat. Voor Koksijde leverde dit op dat  $\hat{k} = 2,36$  en  $\hat{c} = 5,20$  en voor Middelkerke vonden we  $\hat{k} = 2,13$  en  $\hat{c} = 5,99$ . Deze waarden zijn eveneens terug te vinden in Tabel 4.17.

Een histogram met daarboven de kansdichtheidsfunctie (Figuur 4.19) toont dat de schatting voor  $k$  en  $c$  met de maximum-likelihoedmethode heel wat accurater is dan met de power-densitymethode. Mogelijks zijn er bepaalde regio's waar de Weibullparameters accurater geschat kunnen worden met de power-densitymethode dan met bijvoorbeeld de momentmethode, maar aan de Belgische kust is dit blijkbaar niet het geval.

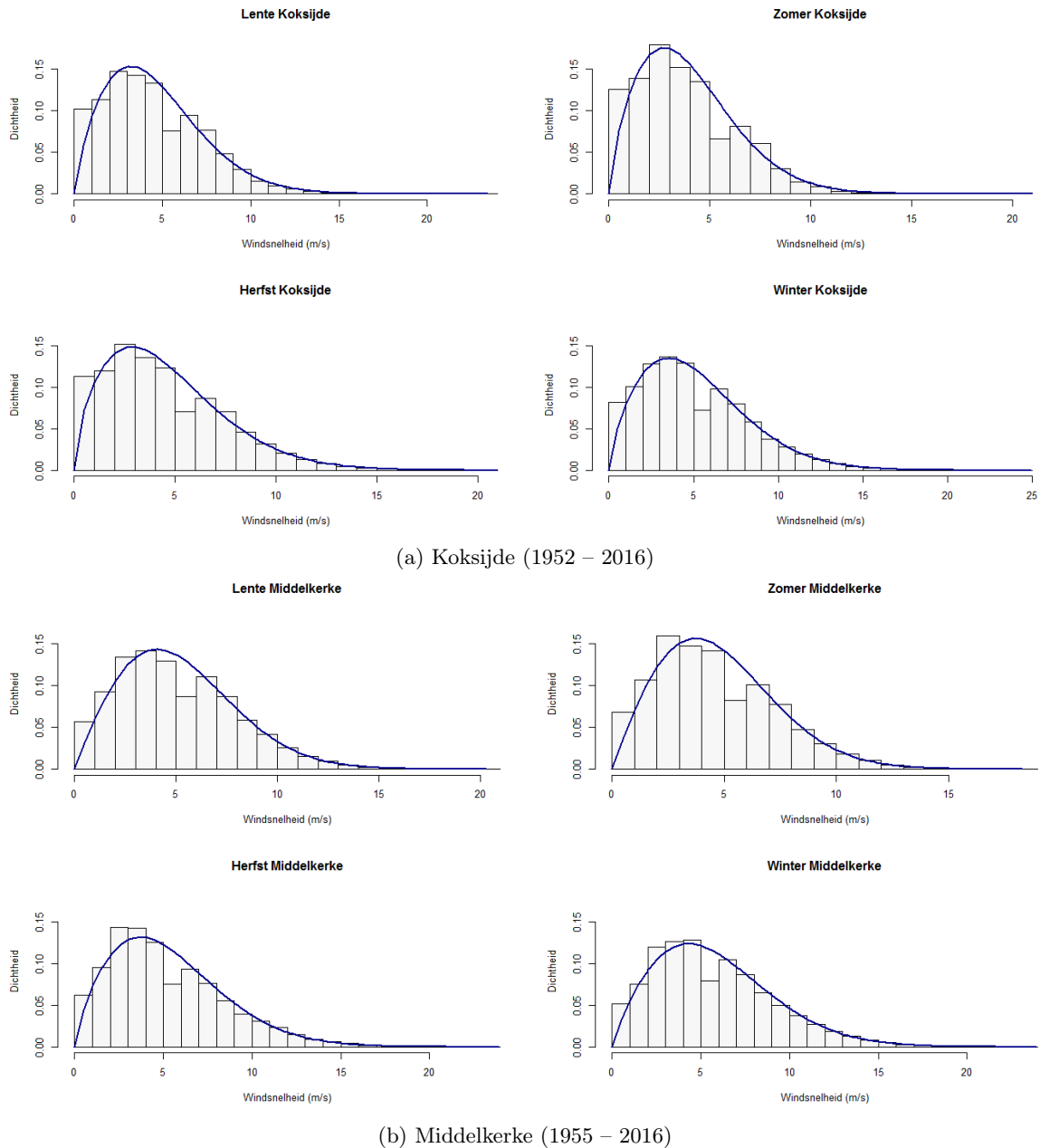
Omdat deze methode voor de kust niet de meest accurate lijkt en bijgevolg minder nuttig is voor de analyse, zullen de Weibullparameters voor de seizoenen niet meer geschat worden.

#### 4.4.4 Overzicht van de resultaten

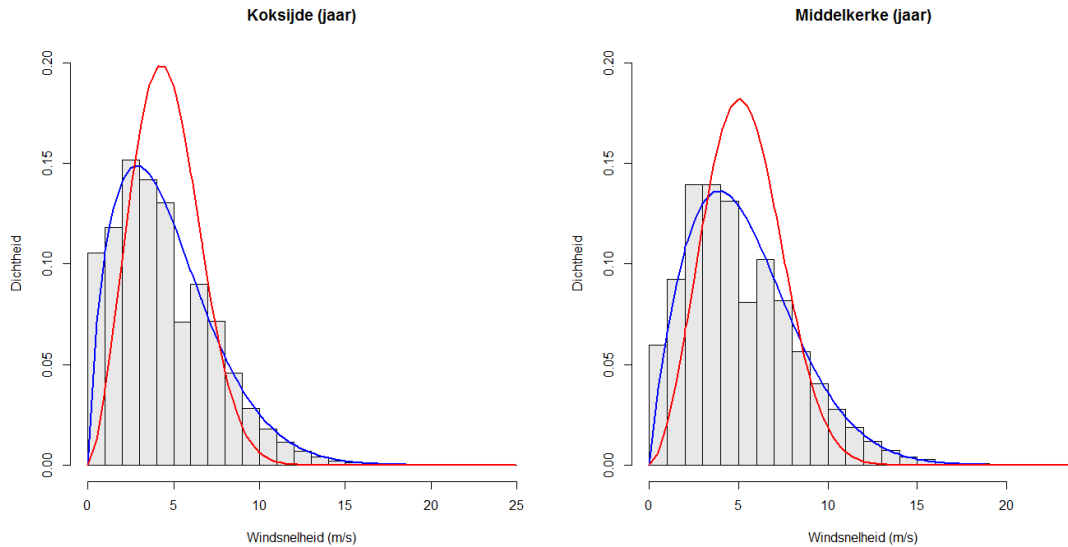
In samenvattende tabel 4.17 staan de schattingen voor de parameters  $k$  en  $c$  weergegeven die gevonden werden met behulp van de moment-, maximum-likelihoed- en power-densitymethode.

### 4.5 Geschikt voor modellering van windsnelheden?

Bij zowel de momentmethode als de maximum-likelihoedmethode vonden we schattingen voor de parameters  $k$  en  $c$  die vrij dicht bij elkaar lagen en de Weibullkansverdelingsfunctie met parameters geschat via de maximum-likelihoedmethode leken de histogrammen van de windsnelheden vrij goed te fitten (Figuur 4.18). Er was echter



Figuur 4.18: Via maximum-likelihoodmethode gefitte Weibulldistributie boven het histogram van de windsnelheden in de vier seizoenen.



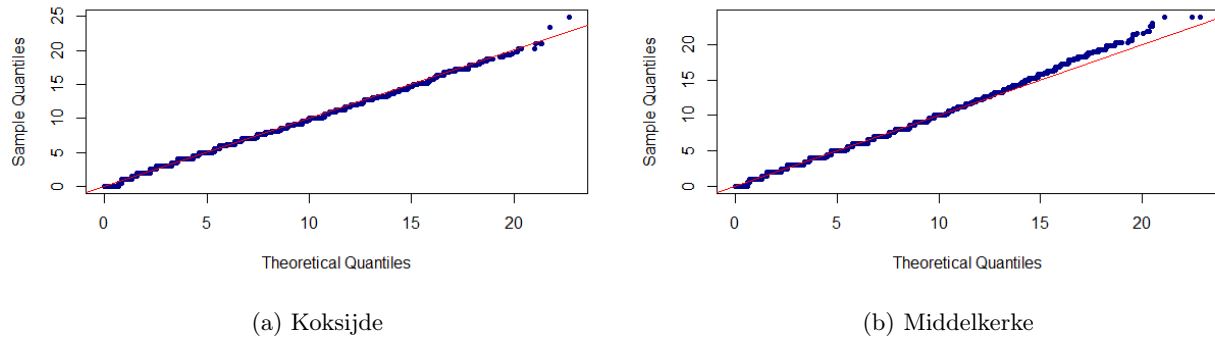
Figuur 4.19: Gefitte Weibulldistributie boven het histogram van de windsnelheden in Koksijde (links) en Middelkerke (rechts). De parameters van de blauwe kansdichtheidsfunctie werden geschat met de maximum-likelihoodmethode, de parameters van de rode kansdichtheidsfunctie met de power-densitymethode.

in de Figuren 4.18a en 4.18b ook te zien dat er te veel windsnelheden zijn tussen 0 en 1 m/s. Er zijn dan weer te weinig windsnelheden tussen 5 en 6 m/s. Dit laat zien dat de windsnelheden geen perfecte Weibullverdeling volgen, maar die wel aanvaardbaar benaderen. Dat er tussen 5 en 6 m/s relatief weinig windsnelheden in de dataset zitten, is vreemd. Het is niet onmiddellijk duidelijk waarom dit zo is.

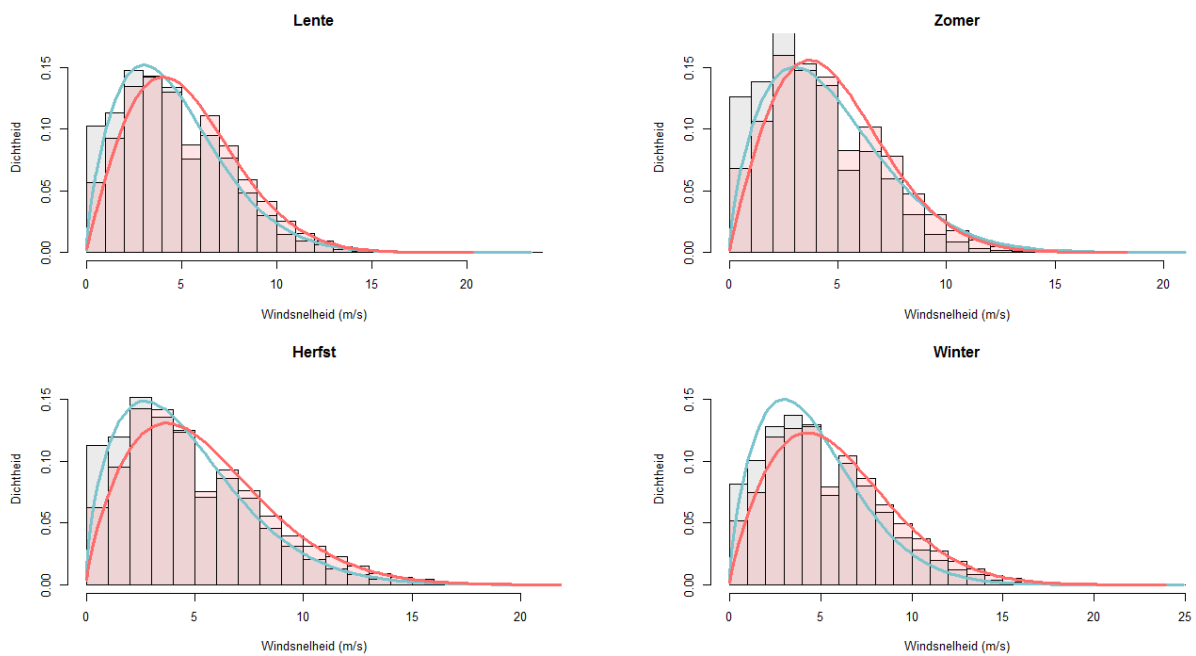
Om te evalueren hoe waarschijnlijk het is dat een bepaalde reeks metingen afkomstig is uit een theoretische stochastische variabele met een zekere kansverdeling, zijn een aantal verschillende tests beschikbaar, de zogenaamde goodness-of-fit-testen. Voorbeelden van zo'n tests zijn de Cramér-von Mises Test, Anderson-Darling Test, Kolmogorov-Smirnov Test. Uit de voorgaande analyse was op het zicht te zien dat de Weibulldistributie de windsnelheden vrij goed leek te fitten. Echter, de  $p$ -waarde bij elk van de testen is zeer klein, namelijk tussen  $10^{-16}$  voor Kolmogorov-Smirnov en  $10^{-8}$  voor Cramér-von Mises, wat betekent dat de nulhypothese, dat de windsnelheden uit een Weibulldistributie getrokken zijn, zou moeten verworpen worden. Mogelijks is de lage  $p$ -waarde te verklaren door de afwijkende frequenties van windsnelheden in  $[0, 1]$  en  $[5, 6]$  m/s of door het onvoorspelbaar discreet karakter van de waarden (zie Figuur 3.1). Bij het vergelijken van tienduizenden waarden uit een eerder discreet spectrum met de theoretische continue Weibullverdeling, merken deze testen de grillige granulariteit mogelijks aan als een systematische afwijking van de theoretische verdeling (persoonlijke communicatie met Els Goetghebeur op 18 april 2017). Een andere mogelijke verklaring is dat de uurgemiddelde windsnelheden al meer neigen naar een normale verdeling door de centrale limietstelling, precies omdat het uurgemiddelden zijn en geen puntmetingen. Mogelijks ligt de ware verdeling van de uurgemiddelde windsnelheden ergens tussen de normale en de Weibullverdeling.

Ondanks deze drastische conclusie van de goodness-of-fit-tests en de drie mogelijke verklaringen, lijkt het resultaat niet te stroken met wat visueel kon worden vastgesteld. Het was de bedoeling om te argumenteren dat de data voldoende goed door een Weibullverdeling benaderd kunnen worden. In dit opzicht volstaat een visueel beeld zoals in Figuur 4.18 en kunnen we er bijgevolg met voldoende accuraatheid van uitgaan dat de uurgemiddelde windsnelheden in Koksijde en Middelkerke Weibull verdeeld zijn.

In Figuur 4.20 wordt nog een Weibull-QQ-plot getoond van de windsnelheden in Koksijde en Middelkerke. Hierop is te zien dat de Weibullverdeling de windsnelheden inderdaad vrij goed kan modelleren, en dit iets beter in Koksijde dan in Middelkerke.



Figuur 4.20: Weibull-QQ-plot van de windsnelheden



Figuur 4.21: Per seizoen wordt een histogram weergegeven van de windsnelheden in Koksijde (blauw) en Middelkerke (rood). Bovenop dit histogram wordt de Weibullkansverdelingsfunctie geplot met de parameters die geschat werden via de maximum-likelihoodmethode.

## 4.6 Vergelijking

### 4.6.1 Koksijde versus Middelkerke

Bij het aanvangen van de deze sectie moet een keuze gemaakt worden voor de parameters  $\hat{k}$  en  $\hat{c}$ . Zal gebruik gemaakt worden van de schattingen gevonden via de momentmethode of via de maximum-likelihoodmethode? Omdat goodness-of-fittestests geen goed resultaat geven, baseren we ons hiervoor op resultaten uit de literatuur. Daar kwam in verschillende studies de maximum-likelihoodmethode naar voren als de betere methode [58, 42].

Aan de hand van de parameters die geschat werden met de maximum-likelihoodmethode kan Figuur 4.21 gemaakt worden. Uit deze grafieken is te zien dat de gefitte kansdichtheidsfunctie van de windsnelheden in Middelkerke voor elk seizoen iets meer naar rechts liggen, wat betekent dat de gemiddelde windsnelheden in Middelkerke iets hoger liggen dan deze in Koksijde.

### 4.6.2 In de winter waait het meer

Zoals reeds een aantal keer werd opgemerkt, ligt de gemiddelde windsnelheid in de winter hoger dan in de andere seizoenen. Dit werd echter enkel op het zicht vastgesteld. Het loont dus zeker de moeite om dit na te gaan met een gepaarde test. De nulhypothese bij deze test is dat het werkelijke verschil in de gemiddelden gelijk is aan nul. Een gepaarde  $t$ -test om de verschillen in gemiddelde windsnelheid per jaar tussen winter en lente te beoordelen geeft een  $p$ -waarde van  $2.11 \cdot 10^{-10}$ . De  $p$ -waarde is zeer klein waardoor de nulhypothese kan verworpen worden en we dus effectief mogen besluiten dat het in de winter gemiddeld meer waait dan in de lente.

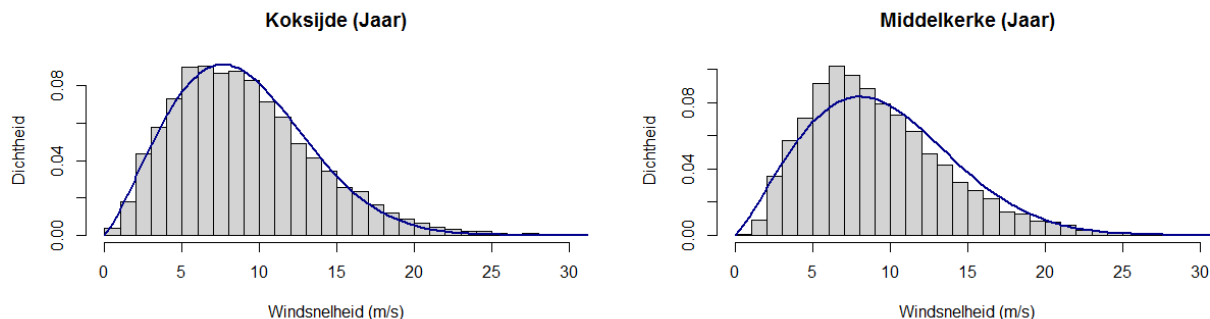
Diezelfde vaststelling werd ook gedaan in een studie van [27]. Deze sterkere windsnelheden komen voor op het moment dat de temperatuurgradiënt tussen de evenaar en de Noordpool het grootst is. Dit kan verklaren waarom in de winter de sterkste windsnelheden te zien zijn: op dit moment van het jaar zijn de temperatuur- en dus drukgradiënt het grootst. De laagste windsnelheden zijn te zien in de zomermaanden. Ook dit zou verklaard kunnen worden door het feit dat op dat moment van het jaar temperatuur- en drukgradiënt tussen evenaar en Noordpool het kleinst zijn.

## 4.7 Verdeling van rukwinden

Naast de windsnelheden zijn ook de rukwinden beschikbaar in de dataset. De vraag stelt zich of ook de rukwinden Weibull verdeeld zijn. Hiervoor bekijken we enkel de gegevens tussen 1994 en 2016 omdat enkel in deze periode voldoende gegevens ter beschikking zijn zowel in Koksijde als in Middelkerke (zie Tabel 3.4). De histogrammen in Figuur 4.23 geven een eerste beeld van de verdeling van rukwinden.

Koksijde	$\hat{k}$	$\hat{c}$	Middelkerke	$\hat{k}$	$\hat{c}$
Lente	2,402	9,879	Lente	2,335	10,320
Zomer	2,383	9,123	Zomer	2,541	9,791
Herfst	2,136	9,979	Herfst	2,113	10,674
Winter	2,120	11,138	Winter	2,124	11,767
Jaar	2,201	10,039	Jaar	2,211	10,649

Tabel 4.22: Geschatte Weibullparameters voor de rukwinden per seizoen in Koksijde en Middelkerke

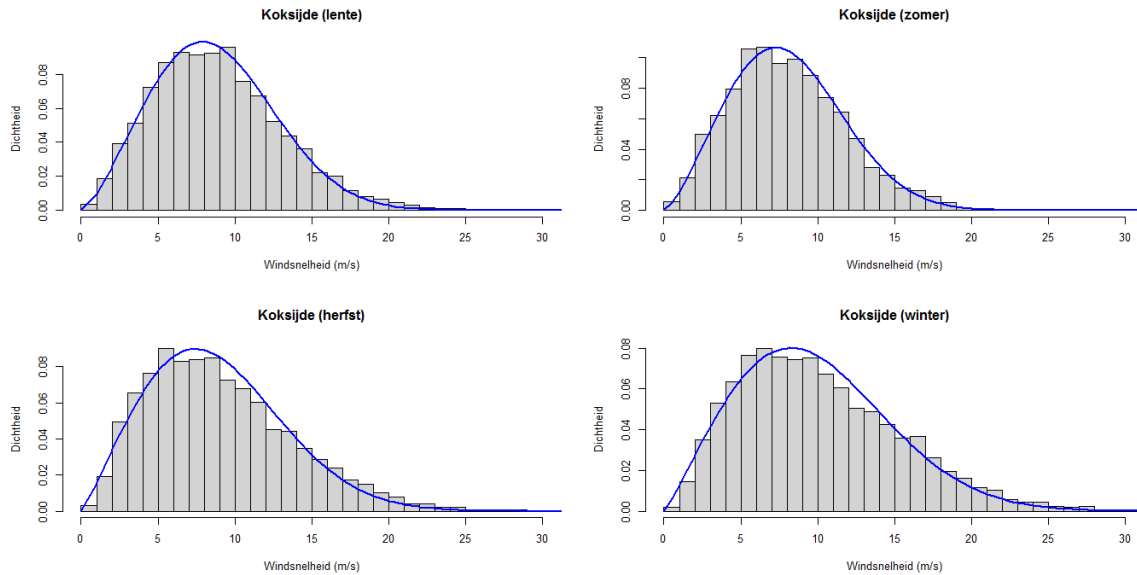


Figuur 4.23: Histogram van de rukwinden van 1994 t.e.m. 2016 in Koksijde (links) en Middelkerke (rechts) met daarboven de Weibullkansverdelingsfunctie geplot met de parameters die geschat werden via de maximum-likelihoodmethode.

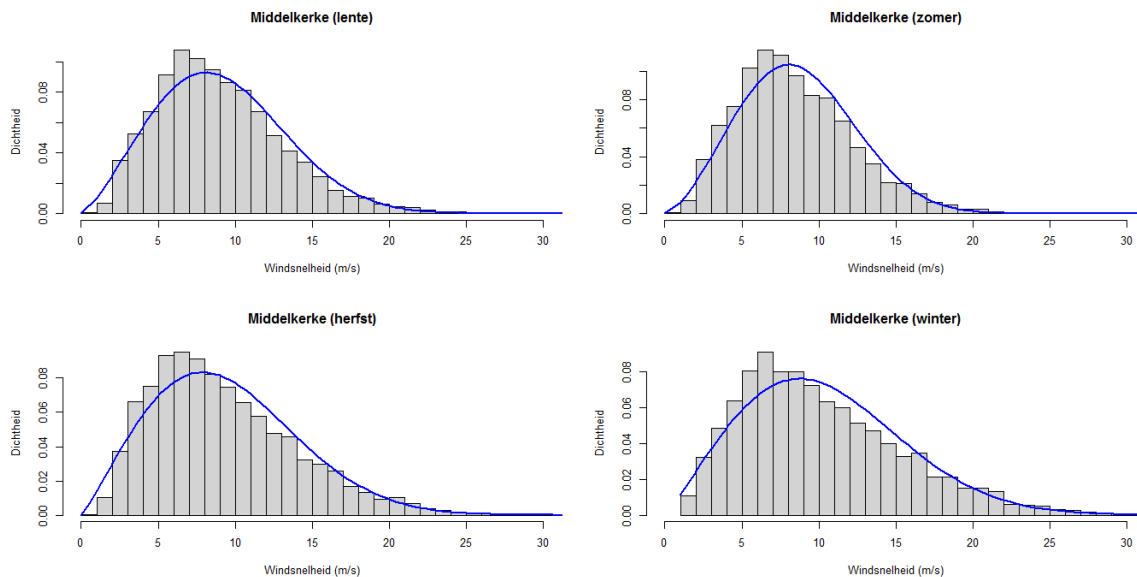
Zoals te verwachten was, omdat ook hier met windsnelheden wordt gewerkt, geeft dit histogram aan dat de rukwinden een Weibullverdeling lijken te volgen. Het is vrij evident dat de rukwinden ‘beter’ Weibull verdeeld zijn dan de windsnelheden: er werden een heleboel minder nulmetingen geregistreerd, waardoor een Weibullverdeling het linkergedeelte van het histogram beter fit.

Tabel 4.22 toont de schattingen van de parameters van de Weibullverdeling aan de hand van maximum likelihoodmethode. Rukwinden hebben per definitie hogere snelheden dan uurgemiddelde windsnelheden en dit laat zich ook merken in de waarden van de schaalparameter  $c$  die veel groter zijn dan bij de windsnelheden. Dit betekent dat de kansdichtheidsfunctie meer naar rechts ligt en dus ook de modus groter is. Verder valt op dat ook de rukwinden in Middelkerke hoger zijn: de parameter  $c$  is systematisch groter dan in Koksijde.





(a) Koksijde (1994 – 2016)



(b) Middelkerke (1994 – 2016)

Figuur 4.24: Per seizoen wordt een histogram weergegeven van de windstoten van 1994 t.e.m. 2016. Bovenop dit histogram wordt de Weibullkansverdelingsfunctie geplot met de parameters die geschat werden via de maximum-likelihoodmethode.

## Hoofdstuk 5

# Testen en methoden voor trendanalyse

Volgen de windsnelheden, de temperatuur, de neerslag ... aan de Belgische kust een systematische trend? Hoe kunnen we een echte trend onderscheiden van toevallige fluctuaties? We duiken in de literatuur om een aantal vaak gebruikte testen of analysemethoden uit te klaren vooraleer we deze toepassen op de weerdata van de Belgische kust. Methoden die geschikt lijken om een beeld te krijgen van de evolutie van jaar-, seizoens- of maandgemiddelden worden hieronder besproken.

### 5.1 LOWESS

LOWESS<sup>1</sup>, of Locally Weighted Scatterplot Smoothing, is een niet-parametrische methode die een gladde kromme genereert door een tijdsreeksgrafiek. Op die manier krijgen we een eerste beeld van de relatie tussen de variabelen, in dit geval de windsnelheden en de jaren. Het voordeel van niet-parametrische methoden is dat zij niet veronderstellen dat de data een bepaalde verdeling volgt.

Om een gladde kromme te fitten, wordt in deze methode een zekere omgeving rond elke  $x$ -waarde gebruikt om een corresponderende  $y$ -waarde te schatten (vandaar *Locally Weighted*). De breedte van de omgeving die rond elke  $x$ -waarde moet gebruikt worden, kan meegegeven worden met behulp van een parameter, de *smoothing parameter*  $\alpha$ . De waarde van  $\alpha$  geeft aan welk percentage van de data wordt gebruikt bij elke fit. Zo kan men ervoor kiezen om een zeer brede omgeving in rekening te brengen. Een te brede omgeving zal voor een regressiekromme zorgen die sterk lijkt op een rechte. Ook smalle omgevingen rond de  $x$ -waarden zijn toegestaan, zolang er genoeg data in de buurt van de  $x$ -waarden aanwezig zijn voor een accurate fit [41]. Deze smalle omgevingen kunnen echter voor het probleem zorgen dat de kromme de data *te goed* zal volgen en dat de onderliggende structuur niet meer zichtbaar is.

Grote waarden voor  $\alpha$  creëren de meest ‘oppervlakkige’ krommen, in die zin dat schommelingen in de data bij deze krommen nauwelijks of niet af te lezen vallen. Hoe kleiner  $\alpha$  wordt gekozen, hoe beter de regressiekromme de individuele data beschrijft.

LOWESS doet ook verfijningen om de ‘gladheid’ te optimaliseren en om de invloed van outliers te beperken, dus om de methode min of meer robuust te maken voor outliers [41].

De voornaamste verdienste van LOWESS is dat deze methode een snel visueel beeld kan geven van trends in een puntenwolk. LOWESS is echter niet in staat om een statistisch gefundeerd antwoord te geven of een bepaalde trend aanwezig is of niet, op een bepaald significantieniveau.

---

<sup>1</sup>Een andere, gerelateerde niet-parametrische regressiemethode is LOESS. In feite verschillen LOWESS en LOESS in de graad die voor de polynomen gebruikt wordt om lokale regressie te doen, maar vaak wordt het onderscheid tussen deze methoden niet gemaakt en worden ze gezien als gelijken.

## 5.2 Lineaire regressie

Een gekende en veelgebruikte methode om een responsvariabele  $Y$  te voorspellen op basis van een predictorvariabele  $X$  is enkelvoudige lineaire regressie. In dit model wordt een lineair verband verondersteld tussen de variabelen  $X$  en  $Y$ . Dit verband wordt gegeven door  $Y \approx \beta_0 + \beta_1 X$ , met  $\beta_0$  en  $\beta_1$  twee parameters die respectievelijk de intercept en de hellingsgraad van de regressierechte voorstellen. Deze parameters zijn in de populatie onbekend en moeten geschat worden op basis van de beschikbare data. Schattingen voor deze parameters worden aangeduid met  $\hat{\beta}_0$  en  $\hat{\beta}_1$ . Aan de hand van deze schatters kan de responsvariabele geschat worden:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \text{ voor een zekere } x.$$

In dit onderdeel van de thesis is de responsvariabele  $Y$  de gemiddelde windsnelheid per jaar of per seizoen en de predictorvariabele  $X$  het jaartal. Dit lineaire model is een benadering van eerste orde voor de trend. Hogereordebenaderingen kunnen bijvoorbeeld via polynomiale regressie (zie sectie 5.3) bekomen worden.

Schattingen voor de parameters bepalen  $\hat{\beta}_0$  en  $\hat{\beta}_1$  zo dat de regressierechte erg dicht bij de datapunten ligt, met name door de residual sum of squares ( $\text{RSS} = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$ ) te minimaliseren – dit is de kleinstekwadratenmethode. Hierbij worden schatters voor  $\beta_0$  en  $\beta_1$  gegeven door

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

en

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

met  $\bar{x}$  en  $\bar{y}$  de steekproefgemiddelden.

## 5.3 Polynomiale regressie

Wanneer het gemiddelde verband tussen een predictor- en responsvariabele niet lineair is, kan gebruik gemaakt worden van een polynomiale functie om de relatie tussen beide variabelen beter uit te drukken. In plaats van een lineair verband te fitten tussen de respons- en predictorvariabele, fitten we een model van de vorm

$$y_i \approx \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \dots + \beta_d x_i^d, \text{ voor een zekere } d \geq 1.$$

Ook hier kunnen schatters voor de  $\beta_j$  met  $j \in \{0, 1, 2, 3, \dots, d\}$  gevonden worden door middel van de kleinstekwadratenmethode.

## 5.4 Niet-parametrische lineaire regressie

Een andere methode om na te gaan of een bepaalde variabele een trend vertoont, is niet-parametrische lineaire regressie, een resampling of permutatiemethode. Dergelijke permutatiemethoden hebben typisch een kleinere power om eenzelfde verschil te detecteren dan parametrische testen [34].

Een permutatietest is een robuuste methode om de significantie van een teststatistiek te schatten. Bij een permutatietest gaat het erom dat de responswaarden van de datapunten  $\{(x_i, y_i) | 1 \leq i \leq n\}$  gepermuteerd worden tot  $\{(x_i, y_{\sigma(i)}) | 1 \leq i \leq n\}$ , voor verschillende permutaties  $\sigma \in \text{Sym}(\{1, 2, \dots, n\})$ . Zijn er  $n$  datapunten in de dataset, dan zijn er  $n!$  permutaties mogelijk. Wanneer er geen trend in de data aanwezig is (i.e. als  $H_0$  geldt), zal de volgorde (met betrekking tot de rangorde van de predictorvariabele  $x_i$ ) van de waarden weinig effect hebben op de regressiehelling. Voor elke permutatie  $\sigma$  die wordt uitgevoerd, kan de schatter voor  $\beta_1$  berekend worden. Deze gesimuleerde regressiecoëfficiënten hebben gemiddelde nul en een symmetrische verdeling, want voor elke permutatie die een bepaalde  $\hat{\beta}_1$  oplevert, heeft de omgekeerde permutatie die  $-\hat{\beta}_1$

zou opleveren, dezelfde kans om getrokken te worden. Nadien kan de oorspronkelijke  $\hat{\beta}_1$  vergeleken worden met de gesimuleerde regressiecoëfficiënten die gevonden werden bij de permutaties. Wanneer deze originele  $\hat{\beta}_1$  opvallend verschilt van de meeste van de gesimuleerde regressiecoëfficiënten, wijst dit erop dat de volgorde van de waarden de helling beïnvloedt en dus dat de nulhypothese allicht verworpen kan worden. Eenmaal we het percentiel  $p$  hebben bepaald van de originele  $\hat{\beta}_1$  in de populatie gesimuleerde  $\hat{\beta}_1$  kunnen we bepalen hoeveel er extremer gelegen zijn, met name  $2 \cdot \min(p, 1 - p)$ .

Een manier om een mate van significantie te bepalen voor de permutatietest, wordt gegeven door [34]. Alle gesimuleerde regressiecoëfficiënten kunnen geordend worden van klein naar groot. Dan wordt bepaald waar de originele regressiecoëfficiënt hoort te staan. Vervolgens bepalen we het aantal regressiecoëfficiënten dat kleiner was dan deze oorspronkelijke en we duiden dit aantal aan met  $k$ . De kans  $q$  dat een regressiecoëfficiënt onder  $H_0$  kleiner (extremer) is dan de originele  $\hat{\beta}_1$  wordt gegeven door  $q = \frac{k}{s}$ , met  $s$  het aantal uitgevoerde permutaties. De significantie van de teststatistiek  $\hat{\beta}_1$  wordt dan gegeven door  $p = 2 \cdot \min(q, 1 - q)$ , een waarde die mogelijks kan dienst doen als een soort van  $p$ -waarde en kan vergeleken worden met het vooropgestelde significantieniveau  $\alpha$ , typisch met  $\alpha = 0.05$ . Dat er evenwel redenen zijn om daaraan te twijfelen, betogen we in sectie 5.7.1.

Hoe groter het aantal permutaties dat wordt uitgevoerd, hoe accurater de significantie wordt. Typisch zijn 100 tot 2 000 permutaties voldoende om een representatief resultaat te krijgen [34].

**Assumpties** Volgens Good [16, 15] moet aan de voorwaarde van *exchangeability* voldaan zijn om de permutatietest goed te kunnen toepassen. Dat betekent dat, mochten de datapunten oorspronkelijk in een andere volgorde staan, dit geen verschil mag uitmaken voor uitkomsten van de test. Er is niet onmiddellijk een reden waarom deze voorwaarde niet voldaan zou zijn in het geval van bijvoorbeeld jaargemiddelde windsnelheden.

## 5.5 Mann-Kendalltest en Sen's slope

De Mann-Kendalltest voor trenddetectie [36, 25] en *Sen's slope* [52] zijn twee methoden die in klimatologische studies reeds vaak gebruikt werden voor het detecteren (Mann-Kendalltest) en kwantificeren (Sen's slope) van trends in tijdsreeksen van meteorologische data [3, 4, 11, 17, 19, 23, 67].

### Mann-Kendalltest

**Doel** De Mann-Kendalltest is een niet-parametrische statistische test. Deze test wordt vaak gebruikt om de significantie van trends in meteorologische/klimatologische tijdsreeksen zoals temperatuur en neerslag te beoordelen [67, 11]. Een belangrijke reden om een niet-parametrische test te gebruiken is omdat deze meer geschikt geacht wordt voor data die niet normaal verdeeld zijn, wat bij meteorologische tijdsreeksen vaak het geval is [67]. De trend hoeft ook niet strikt lineair te zijn. In deze testen wordt echter wel een 'serial independence' geëist. De Mann-Kendalltest vraagt niet dat de onderzochte variabelen normaal verdeeld zijn, in tegenstelling tot parametrische lineaire regressie. Deze test doet uitspraken over het al dan niet aanwezig zijn van een monotone trend die niet noodzakelijk lineair hoeft te zijn [44].

De tweezijdige Mann-Kendalltest gaat na of de nulhypothese  $H_0$  verworpen kan worden ten voordele van de alternatieve hypothese  $H_a$ , waarbij

$$H_0 : \text{geen trend}$$

$$H_a : \text{monotone trend aanwezig (stijgend of dalend over de tijd)}$$

**Assumpties** Om de Mann-Kendalltest met recht te mogen toepassen, moet aan de volgende voorwaarden voldaan zijn:<sup>2</sup>

---

<sup>2</sup>Dit onderdeel is vertaald uit [44].

- Wanneer er geen stijgende of dalende trend is, zijn de metingen die verkregen werden in de onderzoeksperiode onafhankelijk en identiek verdeeld.
- De verkregen waarnemingen over de tijd zijn representatief voor de echte omstandigheden bij steekproeftijden.
- De steekproefverzamelings-, verwerkings- en meetmethoden bieden over de tijd onbevooroordeelde en representatieve observaties van de onderliggende populaties.

De assumpties uit [44] spreken over steekproeven en representatieve observaties en hun voldoende spreiding in de tijd om seriële onafhankelijkheid te vermijden. Bovendien gebruiken alle bronnen de Mann-Kendalltest omdat metingen van meteorologische tijdsreeksen vaak niet normaal verdeeld zijn. Deze twee observaties suggereren sterk dat de Mann-Kendalltest best gebruikt wordt op rechtstreekse metingen, en niet op verwerkingen daarvan zoals jaargemiddelden, immers door de centrale limietstelling kunnen we verwachten dat jaar- of seizoensgemiddelden over verschillende jaren heen eerder normaal zullen verdeeld zijn zodat de voorwaarden voor lineaire regressie zijn voldaan. Niettemin gebruiken [8, 3, 4, 11, 17, 19, 23, 67] allen de Mann-Kendalltest op langetermijngemiddelden en niet op individuele metingen. Naar analogie met deze bronnen zal ook hier de Mann-Kendalltest worden toegepast.

**Procedure** De procedure van de Mann-Kendalltest die gevolgd moet worden om een antwoord te krijgen op de vraag of een stijgende of dalende trend aanwezig is in een tijdsreeks, staat hieronder beschreven in een aantal stappen, gebaseerd op [14]. Een reeks gegeven datapunten  $\{(1, v_1), (2, v_2), \dots, (n, v_n)\}$  wordt gepermuterd. Vervolgens moet het teken bepaald worden van alle mogelijke verschillen  $v_j - v_i$ , waarbij  $j > i$ . Het totaal aantal verschillen waarvan het teken moet bepaald worden is  $\binom{n}{2} = \frac{n(n-1)}{2}$ . Hierbij gaat het over de volgende verschillen:

$$v_2 - v_1, v_3 - v_1, v_4 - v_1, \dots, v_3 - v_2, v_4 - v_2, \dots, v_5 - v_2, \dots, v_n - v_{n-1}.$$

Met deze informatie kan de Mann-Kendall teststatistiek  $S$  berekend worden als

$$S = \sum_{j>i} \text{sgn}(v_j - v_i) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(v_j - v_i),$$

waarbij  $v_i$  en  $v_j$  respectievelijk de  $i$ -de en  $j$ -de observatie zijn,  $n$  het totaal aantal waarden in de dataset weergeeft en  $\text{sgn}$  de signumfunctie<sup>3</sup> is. De teststatistiek  $S$  hangt dus alleen af van de rang (*rank*) van de waarden en niet van hun precieze grootte. Dit maakt dat  $S$  een verdelingsvrije teststatistiek is, m.a.w. een statistiek die geen veronderstellingen maakt over de verdeling van de variabelen [18]. Een voordeel van zo'n statistiek is dat de power en de significantie van dergelijke test niet beïnvloed wordt door de exacte verdeling van de data, dit in tegenstelling tot bijvoorbeeld lineaire regressie, waarbij normaliteit van de data wordt verondersteld en waarbij de power van de test vrij veel kan afnemen bij scheef verdeelde data [18]. Een nadeel is echter dat deze testen een lagere power hebben, net omdat ze geen gebruik maken van de onderliggende verdeling [63].

Men kan inzien dat wanneer  $S$  aanzienlijk groter is dan nul er aan  $S$  veel positieve bijdragen vanwege  $\text{sgn}(v_j - v_i)$  zijn. Dat betekent dat veel verschillen  $v_j - v_i$  groter zijn dan nul, dus waarden die later in de tijd gemeten zijn, zijn vaker groter dan vroegere waarden. Er zal in dat geval een positieve trend ontdekt worden. Een analoge redenering geldt wanneer  $S$  aanzienlijk kleiner is dan 0. Mann en Kendall hebben aangetoond dat voor  $n \geq 8$  de teststatistiek  $S$  onder de nulhypothese bij benadering normaal verdeeld is met gemiddelde nul

<sup>3</sup>Deze is op de volgende manier gedefinieerd

$$\text{sgn}(x) = \begin{cases} 1, & \text{als } x > 0 \\ 0, & \text{als } x = 0 \\ -1, & \text{als } x < 0. \end{cases}$$

( $E(S) = 0$ ) en variantie<sup>4</sup>  $\text{Var } S = \frac{1}{18} (n(n-1)(2n+5))$  [43, 11, 18]. Eens  $S$  en de variantie van  $S$  gekend zijn, kan de gestandaardiseerde Mann-Kendall teststatistiek  $z_{MK}$  bepaald worden:

$$z_{MK} = \begin{cases} \frac{S-1}{\sqrt{\text{Var } S}}, & \text{als } S > 0 \\ 0, & \text{als } S = 0 \\ \frac{S+1}{\sqrt{\text{Var } S}}, & \text{als } S < 0. \end{cases}$$

De eenheid die hierbij wordt opgeteld en afgetrokken is een correctie die door Kendall [25] werd ingevoerd om continuïteit te garanderen. Deze teststatistiek volgt de standaardnormale verdeling met gemiddelde nul en variantie 1 [43]. Indien  $S$  een significant positieve waarde aanneemt, duidt dit op een stijgende trend; is  $S$  negatief, dan is er een dalende trend. Wanneer de nulhypothese waar is, volgt de teststatistiek  $z_{MK}$  bij benadering een standaardnormale verdeling. Om na te gaan of de trend significant is op het significantieniveau  $\alpha$ , wordt de waarde van de teststatistiek  $z_{MK}$  vergeleken met de standaardnormale waarde  $z_{1-\alpha/2}$ . De trend is significant wanneer  $|z_{MK}| > z_{1-\alpha/2}$  [43, 14, 48].

In alle trendanalyses die gedaan zullen worden, zal een onderscheid worden gemaakt tussen 4 significantieniveaus waarop de nulhypothese verworpen<sup>5</sup> kan worden:

Significantieniveau $\alpha$	$H_0$ verwerpen als	
0.1	$z_{MK} < -1,645$	of $z_{MK} > 1,645$
0.05	$z_{MK} < -1,96$	of $z_{MK} > 1,96$
0.01	$z_{MK} < -2,33$	of $z_{MK} > 2,33$
0.001	$z_{MK} < -3,08$	of $z_{MK} > 3,08$

Tabel 5.1: De waarden die de gestandaardiseerde teststatistiek  $z_{MK}$  moet aannemen om de nulhypothese te kunnen verwerpen op een zeker significantieniveau.

## Sen's slope

Sen's slope is een procedure om te bepalen hoe groot de trend is. Daar waar de Mann-Kendallteststatistiek aangeeft of een (stijgende of dalende) trend aanwezig is, en zo ja, wordt de grootte van de trend bepaald door de helling van de rechte te schatten via een niet-parametrische methode, geïntroduceerd door Pranab Kumar Sen, een Indische statisticus (°1937). Deze robuuste methode is een uitbreiding van de test van Theil [14]. Een groot voordeel is dat deze methode niet al te veel beïnvloed wordt door missing data of outliers, dit in tegenstelling tot wanneer de hellingsgraad zou geschat worden met behulp van de kleinstekwadratenmethode bij lineaire regressie [14].

Om Sen's slope te bepalen, worden alle  $\binom{n}{2} = \frac{n(n-1)}{2}$  differentiequotienten  $Q = \frac{v_j - v_i}{j - i}$  berekend, waarbij  $j > i$  is en  $n$  het aantal waarden. Sen's slope  $m$  is tot slot de mediaan van alle  $Q$ 's.

$$m = \text{mediaan} \left\{ Q_1, \dots, Q_{\binom{n}{2}} \right\}$$

$$= \begin{cases} Q_{\frac{\binom{n}{2}+1}{2}} & \text{als } \binom{n}{2} \text{ oneven is} \\ \frac{1}{2} \left( Q_{\frac{\binom{n}{2}}{2}} + Q_{\frac{\binom{n}{2}+2}{2}} \right) & \text{als } \binom{n}{2} \text{ even is} \end{cases}$$

<sup>4</sup>De variantie van de teststatistiek  $S$  wordt kleiner wanneer er gelijke waarden (Engels: *tied ranks*) in de data voorkomen. In dat geval wordt de variantie gegeven door  $\frac{1}{18} \left[ n(n-1)(2n+5) - \sum_{j=1}^m t_j(t_j-1)(2t_j+5) \right]$ , waarbij  $m$  het aantal groepen van gelijke waarden is en  $t_j$  het aantal van deze gelijke waarden in groep  $j$  [18]. In de analyse van de windsnelheden in paragraaf 6.2.5 komen echter geen gelijke jaar- of seizoensgemiddelden voor omdat deze bepaald zijn tot op 6 cijfers na de komma. Hierdoor kan gewerkt worden met de eenvoudigere vorm van de variantie.

<sup>5</sup>Aangezien de gestandaardiseerde teststatistiek groter moet zijn dan de  $z$ -score van de standaardnormale verdeling, maken we gebruik van het feit dat  $z_{0,95} = 1,645$ ,  $z_{0,975} = 1,69$ ,  $z_{0,99} = 2,33$  en  $z_{0,999} = 3,08$

Wanneer de helling  $m$  voor de trendlijn bepaald is, kan ook de intercept  $b$  berekend worden. Deze intercept wordt gegeven door  $\text{mediaan}(v_i - m \cdot i)$  [51]. In het onderzoek naar een trend in jaarlijkse gemiddelde windsnelheid zal de rechte die de trend aangeeft, voorgesteld worden door

$$y = m(\text{jaar} - 195X) + b.$$

Hierbij is  $m$  Sen's slope,  $b$  de intercept,  $195X$  is gelijk aan 1951 voor Koksijde en 1954 voor Middelkerke en 'jaar' staat voor elk jaar in de periode  $195X - 2016$ , zodat  $i = (\text{jaar} - 195X)$  de waarden  $i = 1, \dots, n$  aanneemt, met  $n = 65$  of  $n = 62$ .

## 5.6 Bonferronicorrectie

De xkcd-comic in Figuur 5.2 laat helder zien wat het doel is van de Bonferroni-correctie. Uit onderzoek blijkt volgens de comic dat gelboonsnoepjes geen verband tonen met de ontwikkeling van acne. Wanneer men onderzoekt of er een bepaald kleur zou zijn dat acne veroorzaakt, vindt men plots wel dat groene gelboonsnoepjes acne veroorzaken. In het aannemelijke geval dat gelboonsnoepjes nooit acne veroorzaken, kunnen we toch verwachten dat er één van de 20 kleuren een significant verband zal tonen, immers als het significantieniveau wordt vastgelegd op 5% is de kans op een type-I-fout (het onterecht verwerpen van de nulhypothese) precies 1 op 20, m.a.w. de verwachting is dat voor één kleur een significant resultaat gevonden wordt. De onderzoeksresultaten in de comic zijn dus helemaal te verwachten, maar de conclusie dat groene gelboonsnoepjes acne veroorzaken, wordt onterecht getrokken. Men voert echter verschillende testen uit (multiple testing) om één vraag te beantwoorden, met name of er *een* kleur van gelboonsnoepjes acne veroorzaakt. Dat  $p < 0.05$  voor één van die verschillende testen is dus niet voldoende om de globale vraag te beantwoorden. Een oplossing voor multiple testing is de Bonferronicorrectie voor het interpreteren van significantieniveaus.

**Bonferronicorrectie op seizoensgemiddelden** Bij het zoeken naar een trend in seizoensgemiddelden, hebben we een verzameling van 4 nulhypotheseën:

$$\hat{\beta}_{1l} = 0 \qquad \hat{\beta}_{1z} = 0 \qquad \hat{\beta}_{1h} = 0 \qquad \hat{\beta}_{1w} = 0.$$

Elk van de vier nulhypotheseën wordt afzonderlijk getest. Indien geen van deze nulhypotheseën kon verworpen worden op het 5% significantieniveau, is Bonferroni-correctie niet nodig. Zijn er echter één of meerdere nulhypotheseën die verworpen kunnen worden op dit significantieniveau, dan is een Bonferroni-correctie aangewezen.

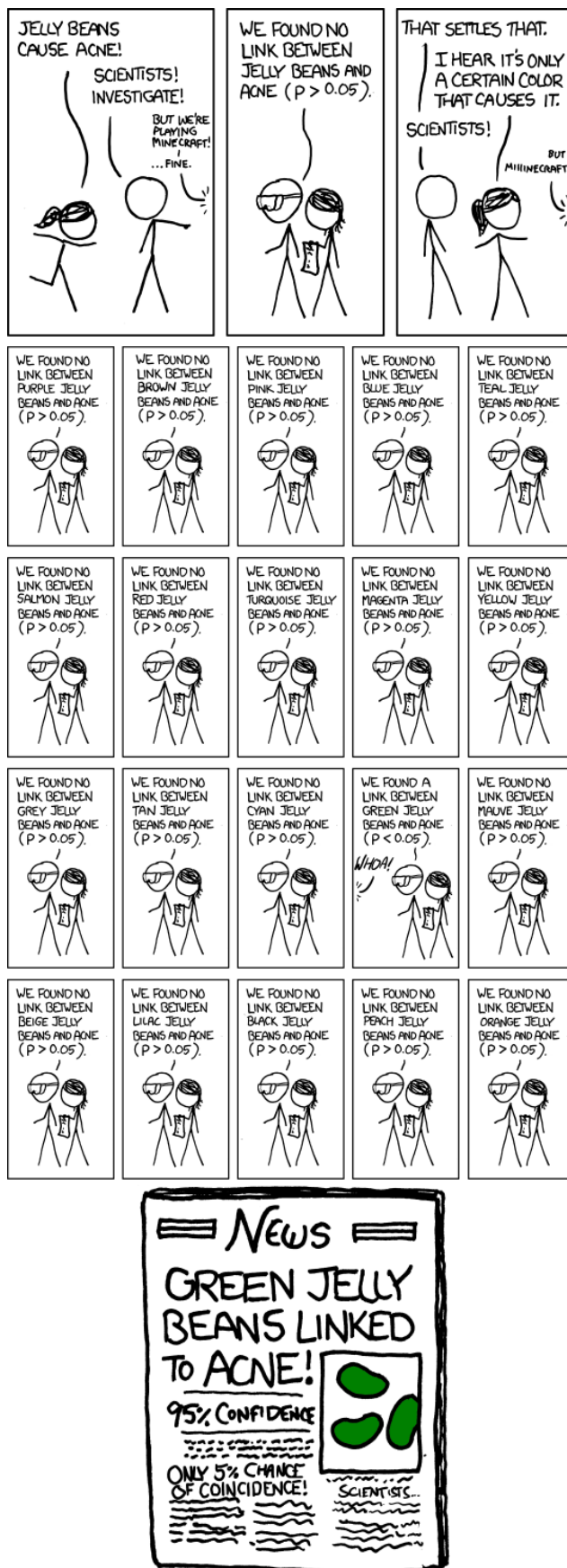
Stel dat we de vraag 'is er een seizoen waarbij de gemiddelde windsnelheden significant stijgen of dalen' willen beantwoorden, dat er  $\tau$  seizoenen zijn en dat we significantieniveau  $\alpha_1$  aanhouden voor één individuele test.

Onder de nulhypothese dat er geen enkel seizoen is met een (dalende of stijgende) trend in gemiddelde windsnelheid, is de kans dat de test voor één seizoen 'per toeval' een significant resultaat geeft – dit is de kans op een type-I-fout –, precies het significantieniveau  $\alpha_1$ . Bij het beantwoorden van de vraag of er *een seizoen bestaat* (onder de  $\tau$  seizoenen) met een dalende of stijgende trend in de gemiddelde windsnelheden is de kans op een type-I-fout

$$\begin{aligned} \alpha_\tau &= P(\text{ten minste één significant resultaat} | H_0) = 1 - P(\text{geen enkel significant resultaat} | H_0) \\ &= P(\text{individueel seizoen geeft geen significant resultaat} | H_0)^\tau \\ &= 1 - (1 - \alpha_1)^\tau \\ &= \tau \cdot \alpha_1 + \mathcal{O}(\alpha_1^2) \end{aligned}$$

waaruit we dus vinden dat

$$\alpha_\tau \approx \tau \cdot \alpha_1.$$



Figuur 5.2: Xkcd-comic "Significant" over multiple testing



Als men dus de vraag of er *een* seizoen met een trend bestaat, wil beantwoorden op een zodanig significantieniveau dat de kans op een type-I-fout  $\alpha_\tau = 5\%$  is, moeten de  $p$ -waarden van de  $\tau$  deeltesten geëvalueerd worden op een individueel significantieniveau

$$\alpha_1 = \frac{\alpha_\tau}{\tau} = \frac{0.05}{\tau}.$$

Bij het geval van testen op de vier seizoenen betekent dit dat het significantieniveau wordt aangepast naar  $0.05/4 = 0.0125$ . Door deze keuze te maken zijn we ervan verzekerd dat er slechts 5% kans is op vals positieven:

$$\begin{aligned} P(\text{ten minste één significant resultaat}|H_0) &= 1 - P(\text{geen enkel significant resultaat}|H_0) \\ &= 1 - (1 - 0.0125)^4 \\ &\approx 0.049. \end{aligned}$$

## 5.7 Vergelijking van de power

Er zullen drie testen gebruikt worden om een trend te bepalen in de jaar- en seizoensgemiddelde windsnelheden en temperaturen, namelijk parametrische lineaire regressie (zie sectie 5.2), niet-parametrische lineaire regressie (permutatietest, zie sectie 5.4) en de Mann-Kendalltest (zie sectie 5.5). Hierbij houdt de nulhypothese in dat er geen trend is, of dus dat  $\beta_1 = 0$ . Interessant hierbij is om te weten welk van deze drie testen de grootste power heeft.

Concreet zal gewerkt worden met de jaarlijkse gemiddelde windsnelheden in Koksijde. Om de vraag te beantwoorden, zullen we bepalen wat de power van de drie testen is om een trend te detecteren van gemiddeld 0,01 m/s per jaar op het 5% significantieniveau voor  $n = 65$ , dit is het aantal jaargemiddelden in Koksijde van 1955 tot 2016. De waarde van 0,01 m/s per jaar werd gekozen omdat vanaf deze waarde gesproken kan worden van een meteorologisch relevante trend (persoonlijke communicatie met David Dehenaauw per e-mail op 18 mei 2017). De power van de drie testen zal via simulatie geschat worden.

### 5.7.1 Benchmarking: schatting kans op type-I-fout

Bij een eerste simulatie worden datasets gegenereerd die de nulhypothese volgen. Er zullen random 10 000 datasets van  $n = 65$  waarden gegenereerd worden die normaal verdeeld zijn met gemiddelde  $\mu = 4,605$  en standaarddeviatie  $\sigma = 2,863$ . Op elk van deze 10 000 datasets worden de drie testen uitgevoerd op het 5% significantieniveau en wordt de  $p$ -waarde keer op keer bepaald en bijgehouden. De verwachting is dat de nulhypothese in 5% van de gevallen verworpen zal worden.

Om de data te simuleren, gebruiken we het lineair model

$$v_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{met } \varepsilon_i \sim N(0, \sigma^2), \quad \text{voor } i = 1, \dots, n \quad (5.1)$$

waarbij we de parameters kiezen als

$$n = 65 \qquad \sigma = 2,86 \qquad \beta_1 = 0 \qquad \beta_0 = 4,605.$$

**Parametrische lineaire regressie** Hierbij vinden we dat in 5013 van de 100 000 gevallen de nulhypothese verworpen wordt op het 5% significantieniveau. Dit komt overeen met 5,013%, wat inderdaad de verwachting was.

**Permutatietest** Wordt de trendanalyse gedaan met de permutatietest, dan wordt de nulhypothese verworpen in 7,2% van de gevallen. Voor deze test werd slechts gewerkt met 1000 random gegenereerde datasets om de rekentijd te beperken. Bovendien werden per test 200 permutaties uitgevoerd. De 7,2% die we hier vinden, wijst op een grote type-I-fout en dus op een bias van de permutatietest richting te veel verwerpen. Dit betekent dat we het resultaat van de powertest met aandacht zullen moeten interpreteren. Nu uit deze simulatie blijkt dat de permutatietest een kans op een type-I-fout heeft die het vooropgestelde significantieniveau ernstig overschrijdt, is een oplossing om onder de  $p$ -waarden te bepalen wat het 95%-kwantiel is, dat blijkbaar kleiner zal zijn dan 0.05, en over verwerping van de nulhypothese te oordelen op basis van deze threshold.

**Mann-Kendalltest** Bij deze test vinden we dat in 4,98% van de gevallen de  $p$ -waarde onder 0.05 duikt, opnieuw een resultaat dat te verwachten was.

Bij lineaire regressie en de Mann-Kendalltest kunnen we besluiten dat het percentage dat verwerpt dicht bij het  $\alpha$ -niveau ligt. Er is geen voor de hand liggende verklaring waarom de permutatietest een afwijkende kans heeft op een type-I-fout.

### 5.7.2 Testen van de power voor een verschil van 0,01 m/s per jaar

Bij de tweede simulatie worden datasets gegenereerd die de alternatieve hypothese volgen. Hierbij wordt het meteorologisch significant verschil  $\beta_1$  gelijk gesteld aan 0,01 m/s per jaar, wat neerkomt op 1 m/s over 100 jaar. Data worden gegenereerd uit het model (5.1) waarbij we de parameters kiezen als

$$n = 65 \qquad \sigma = 0,4452 \qquad \beta_1 = 0,01 \qquad \beta_0 = 4,2757.$$

Hierbij is  $\beta_1$  het vooropgestelde meteorologisch significant verschil dat de test zou moeten detecteren en  $\beta_0$  en  $\sigma$  realistische waarden voor Belgische kustwindsnelheden<sup>6</sup>.

We zullen nu random datasets van  $n = 65$  windsnelheden genereren onder de alternatieve hypothese. Hierbij kunnen de windsnelheden beschreven worden door  $v_i = \beta_0 + \beta_1 \cdot \text{jaar} + \varepsilon_i$ , met  $\varepsilon_i \sim N(0, \sigma^2)$ . Voor elk van de drie methoden doen we de test en gaan we na in hoeveel gevallen de testen erin slagen het aanwezige verschil van 0,01 m/s per jaar te detecteren.

Voor parametrische lineaire regressie vinden we een power van 91,6%, voor de Mann-Kendalltest is de power 90,6% en voor de permutatietest verkrijgen we 92,2%. Mochten we deze procedure meermaals herhalen, dan zouden lichtelijk andere waarden verkregen worden. We kunnen dus stellen dat de drie testen er zeer goed in slagen een verschil te detecteren als dat ook effectief aanwezig is. De verschillen zijn aanwezig, maar niet dermate groot dat bepaalde tests beter niet gebruikt worden. De permutatietest detecteert het vaakst het aanwezige verschil, maar detecteerde ook systematisch vaker een afwezig verschil.

## 5.8 Meer robuuste methoden

Bepaalde trendanalysemethoden zoals gewone lineaire regressie (zie sectie 5.2) zijn gevoelig aan outliers. Door gebruik te maken van robuuste methoden is het mogelijk om de invloed van deze outliers in te perken, wat leidt tot een betere algemene ‘fit’ in het merendeel van de gevallen. Voor situaties met uitschieters bestaan er aangepaste vormen van ‘robuuste lineaire regressie’ die de uitschieters al dan niet betrekken. Voorbeelden hiervan zijn LARS (Least Absolute Residuals), waarbij de som van de absolute afwijkingen van de observaties van hun verwachtingswaarde wordt geminimaliseerd en LMS Regression (Least Median of Squares Regression) waarbij de mediaan van kwadratische afstand tussen uitkomst en predictie wordt geminimaliseerd [41].

<sup>6</sup> $\beta_0$  wordt geschat als  $\frac{1}{n} \sum_{i=1}^n (v_i - \beta_1 \cdot \text{jaar})$  en  $\sigma$  wordt geschat als  $\sqrt{\text{Var}(v_i - \beta_1 \cdot \text{jaar} - \beta_0)}$ . Immers, als de gesimuleerde windsnelheden moeten voldoen aan (5.1), dan is  $\beta_0 = E[v_j - \beta_1 \cdot \text{jaar}]$ . Bovendien is dan  $v - \beta_0 - \beta_1 \cdot \text{jaar}$  normaal verdeeld met variantie  $\sigma^2$ . We kunnen dus een waarde voor  $\sigma^2$  schatten omdat  $\text{Var}(W - \beta_0 - \beta_1 \cdot \text{jaar}) = (0,4452)^2$  is.

De foutieve outliers die zich in de data bevonden, werden verwijderd vooraleer met de regressieanalyse gestart werd en er werden geen andere outliers gesignaleerd. In deze dataset moet dus bij de trendanalyse geen bijzondere rekening gehouden worden met outliers. Aangezien deze robuuste regressiemethoden zoals LARS en LMS geen toegevoegde waarde zouden geven bovenop de gewone lineaire regressie, zullen we ze niet aanwenden om uitspraken te doen over het Belgische kustklimaat.

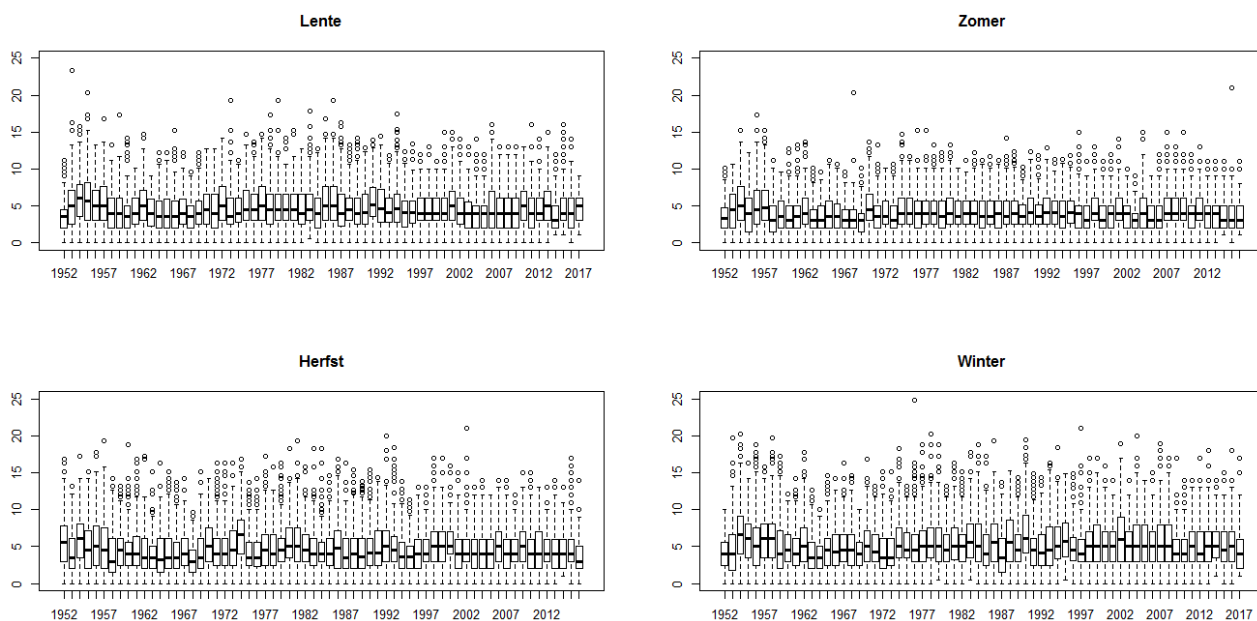
# Hoofdstuk 6

## Trends in de windsnelheden

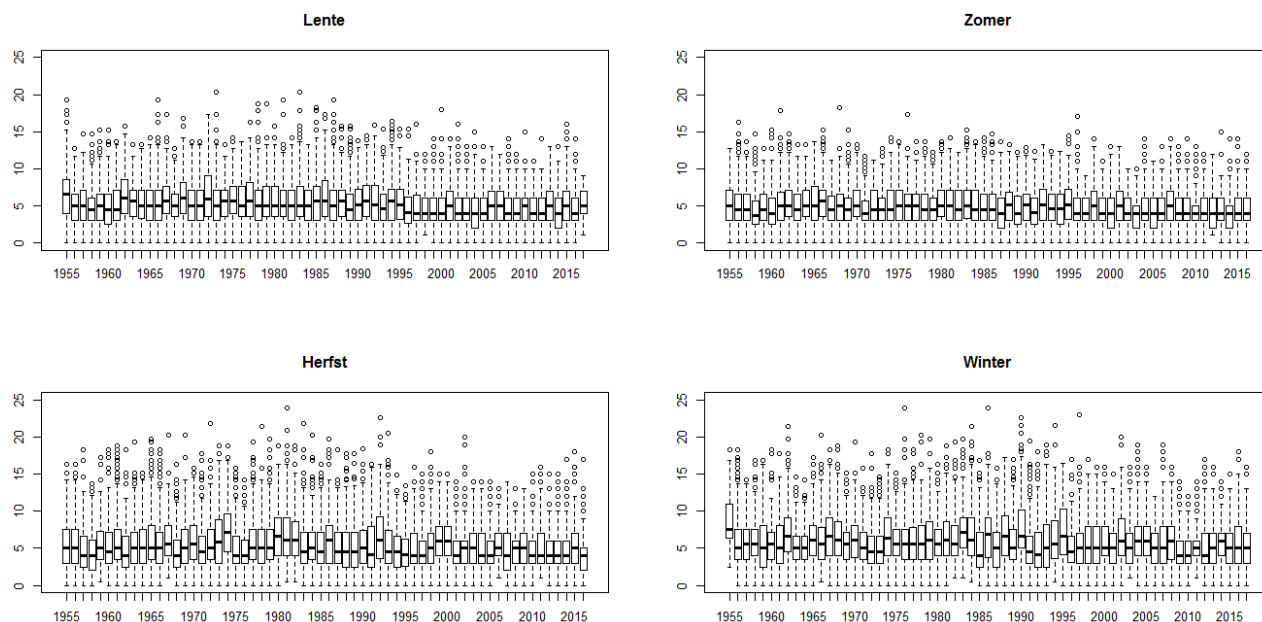
Om na te gaan hoe de jaarlijkse gemiddelde windsnelheid evolueert in de tijd is trendanalyse aangewezen – we hebben hier immers te maken met een meteorologische tijdsreeks. In dit hoofdstuk zal de trend van de jaar- en seizoensgemiddelde windsnelheden bepaald worden. Verderop zal ook gekeken worden naar het aantal dagen waarop het krachtig waaide, maar ook naar het aantal dagen waarop het zwak tot matig waaide. Tot slot zal nog kort iets over de windstoten en de variantie van de windsnelheid per jaar gezegd worden.

### 6.1 Boxplots voor de windsnelheden in de verschillende seizoenen

Figuren 6.1 en 6.2 geven boxplots weer van de windsnelheden voor de 4 seizoenen in Koksijde en Middelkerke. Per seizoen staat voor elk jaar, van 1952 tot 2016, een boxplot getekend met betrekking tot de gemeten windsnelheden in dat jaar. In de boxplots wordt per jaar naast kwartielen en extreme waarden (waarden die verder dan 1,5 keer van de interkwartielafstand liggen van het dichtste kwartiel), de mediaan van de windsnelheden weergegeven. Door boxplots krijgen we in één beeld veel informatie over de verschillende percentielen en de volledige verdeling.



Figuur 6.1: Boxplots van de windsnelheden voor de verschillende seizoenen in Koksijde



Figuur 6.2: Boxplots van de windsnelheden voor de verschillende seizoenen in Middelkerke

## 6.2 Jaar- en seizoensgemiddelden

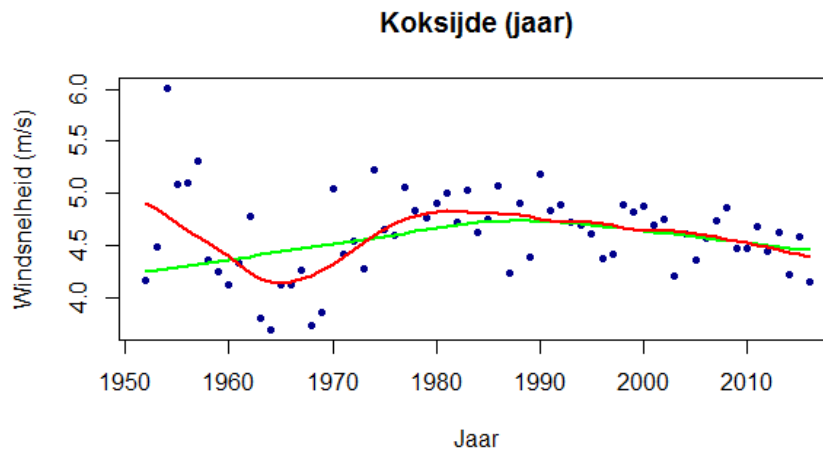
In deze sectie zal aan de hand van een vijftal statistische methoden gekeken worden of een trend gedetecteerd kan worden in de gemiddelde windsnelheid per jaar en per seizoen. Eerst had ik de vraag of de windsnelheden veranderden in de tijd beantwoord door middel van een lineaire regressie (met plots van de jaarlijkse gemiddelde windsnelheden en hun regressierechte voor alle seizoenen en meetstations). Naarmate ik meer literatuur ging bestuderen, leerde die mij dat er ook andere legitieme methoden zijn om meteorologische tijdsreeksen te analyseren. Deze werden bestudeerd en besproken in Hoofdstuk 5.

Om de evolutie van de wind te bestuderen, zal gebruik gemaakt worden van de jaarlijkse gemiddelde windsnelheden. Dit is een gebruikelijke manier om trends in de windsnelheid te bestuderen ([4, 11, 17, 47, 49, 50, 34]). Ook zal een trendanalyse gedaan worden van de seizoensgemiddelde windsnelheden. Deze manier van werken is eveneens terug te vinden in de literatuur ([17, 11, 49]). De gemiddelden werden berekend als het rekenkundig gemiddelde van alle 4 de metingen per dag (gemiddelden voor het uur van rapportering, zie paragraaf 2.2.2). Voor een seizoen worden gemiddelden berekend op basis van ongeveer 368 metingen, voor een jaar gaat dit over 1460 metingen in een gewoon jaar en over 1464 metingen in een schrikkeljaar.

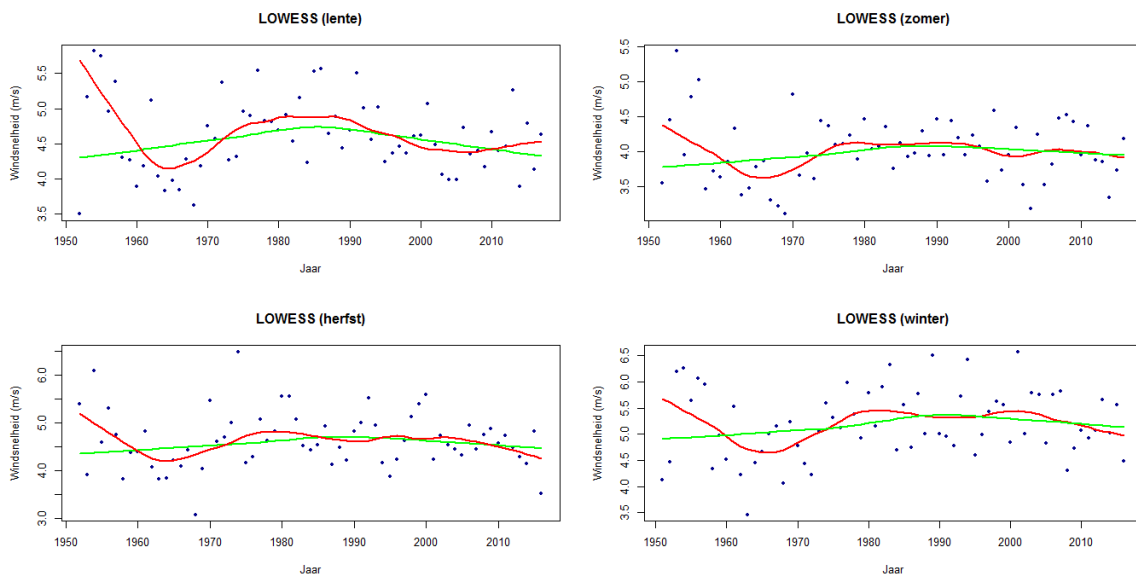
**Gedeeltelijke seizoenen** Gebruik makend van de hulpvariabele `SEIZOENJAAR` werd de gemiddelde windsnelheid per jaar per seizoen berekend. Voor de zomer en de herfst verkrijgen we 65 waarden in Koksijde en 62 in Middelkerke. De lente van 1955 in Middelkerke start volgens de beschikbare metingen echter pas op 1 mei 1955. Hier ontbreken toch heel wat waarden voor de lente van 1955, dus we kiezen ervoor om de gemiddelde windsnelheid in de lente van 1955 niet te laten meetellen. Voor de winter doet het probleem van ‘gedeeltelijke seizoenen’ zich ook eenmaal voor: voor de winter van 1951 in Koksijde ontbreken de gegevens van 21 t.e.m. 31 december 1951. Hier gaat het echter slechts over 11 dagen die ontbreken, dus de winter van 1955 zal meegenomen worden en gegevens van deze winter zullen berekend worden op basis van 1 januari 1955 t.e.m. 20 maart 1955. De winters van 2016 (t.e.m. 20 maart 2017) kunnen voor zowel Koksijde als Middelkerke in hun geheel meegenomen worden omdat data beschikbaar zijn tot na die datum.

### 6.2.1 LOWESS

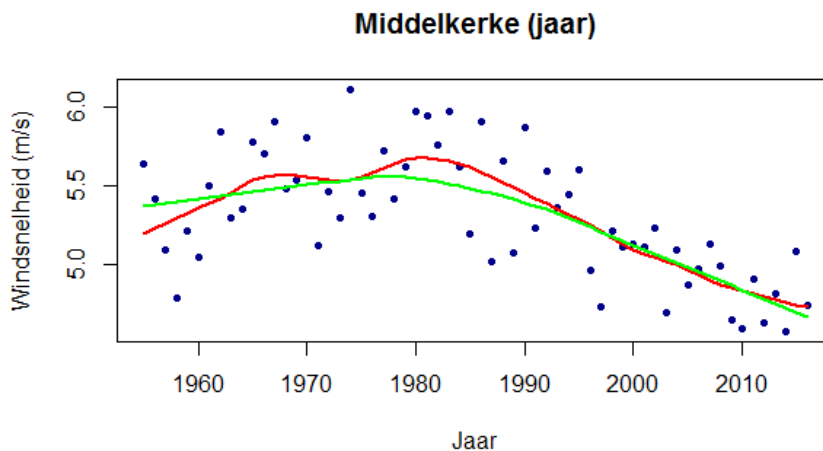
Op de plots in Figuur 6.3 en Figuur 6.5 staan de jaarlijkse gemiddelde windsnelheden in Koksijde respectievelijk Middelkerke weergegeven. In beide figuren worden gefitte krommen getoond voor twee verschillende waarden van  $\alpha$ . Enerzijds werd voor een vrij grote waarde van  $\alpha$  gekozen met name  $\alpha = 0,75$ . In dat geval is te zien dat de kromme zeer glad en breed uitgesmeerd is, zonder pieken of dalen. We zouden bij kleine waarden voor  $\alpha$  kunnen vermoeden dat informatie verloren gaat: ‘met het blote oog’ is vast te stellen dat in Koksijde in de jaren '60 de jaarlijkse gemiddelde windsnelheid opmerkelijk lager lag dan de jaren daarrond. In dat opzicht verwachten we een dal in de gladde kromme die gefit wordt. Wordt voor de smoothing parameter  $\alpha$  een kleinere waarde genomen ( $\alpha = 0,3$  in de figuren), dan is dit dal wel te zien. Deze LOWESS werd ook toegepast op de seizoensgemiddelde windsnelheden. De resultaten daarvan zijn te zien in Figuren 6.4 en 6.6.



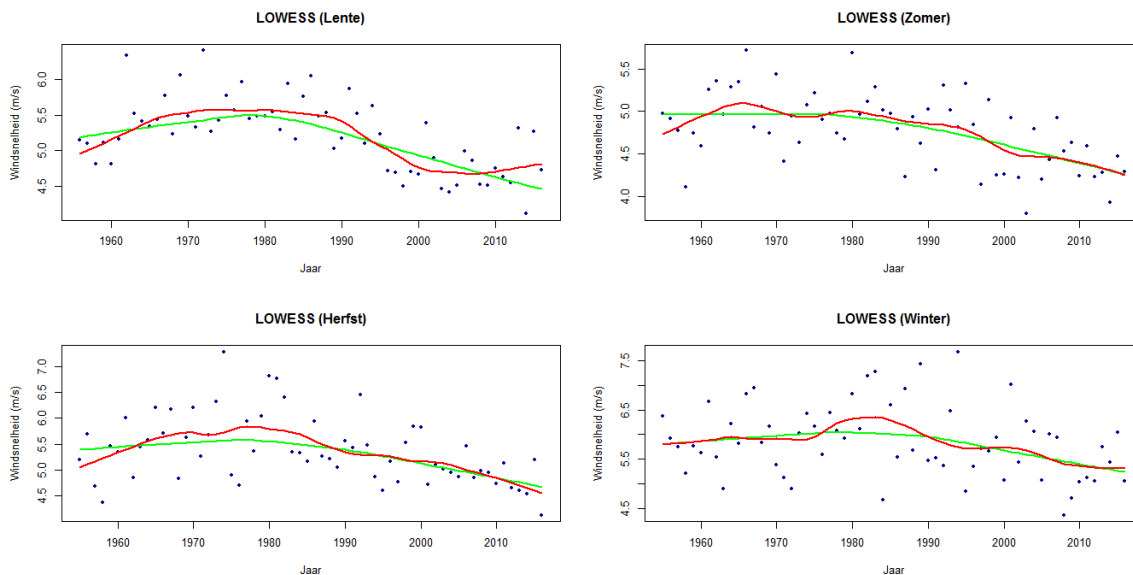
Figuur 6.3: Locally weighted scatterplot smoothing op de jaargemiddelde windsnelheden in Koksijde voor 2 verschillende waarden van de smoothing parameter  $\alpha$ :  $\alpha = 0,75$  (groen),  $\alpha = 0,3$  (rood).



Figuur 6.4: Locally weighted scatterplot smoothing op de seizoensgemiddelde windsnelheden in Koksijde voor 2 verschillende waarden van de smoothing parameter  $\alpha$ :  $\alpha = 0,75$  (groen),  $\alpha = 0,3$  (rood).



Figuur 6.5: Locally weighted scatterplot smoothing op de jaargemiddelde windsnelheden in Middelkerke voor 2 verschillende waarden van de smoothing parameter:  $\alpha = 0,75$  (groen) en  $\alpha = 0,3$  (rood).

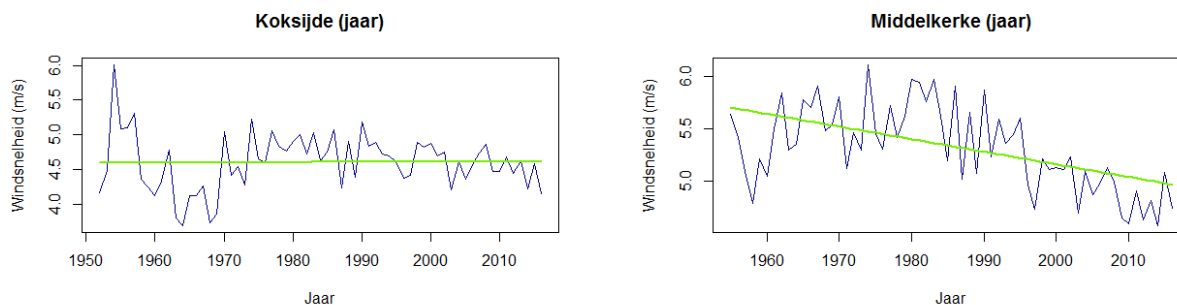


Figuur 6.6: Locally weighted scatterplot smoothing op de seizoensgemiddelde windsnelheden in Middelkerke voor verschillende waarden van de smoothing parameter  $\alpha$ :  $\alpha = 0,75$  (groen),  $\alpha = 0,3$  (rood).

In de Figuren 6.5 en 6.6 waarbij de jaar- en seizoensgemiddelde windsnelheden uit Middelkerke staan weergegeven, valt op dat vanaf het begin van de jaren '80 de gemiddelde windsnelheden een duidelijk dalende evolutie vertonen. In Koksijde is dit minder duidelijk zichtbaar. Aangezien LOWESS een meer visuele methode is, kan hier nog niets gezegd worden in verband met de significantie van de dalende trend. Hiervoor moeten andere methoden voor trendanalyse gebruikt worden.

## 6.2.2 Lineaire regressie

Schattingen van de parameters bekomen via lineaire regressie staan in Tabel 6.8. Uit deze resultaten kan afgeleid worden dat de jaargemiddelde windsnelheden in Middelkerke een significant dalende evolutie doormaken, terwijl in Koksijde geen sprake is van een significante trend. Figuur 6.7 toont de trendlijn  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$  bovenop de jaargemiddelde windsnelheden.



Figuur 6.7: Evolutie jaargemiddelde windsnelheden in Koksijde (links) en Middelkerke (rechts)

## Seizoensgemiddelden

Aan de hand van de  $p$ -waarde voor de geschatte  $\beta_1$  zal een uitspraak kunnen gedaan worden over de trend van de seizoensgemiddelde windsnelheden.

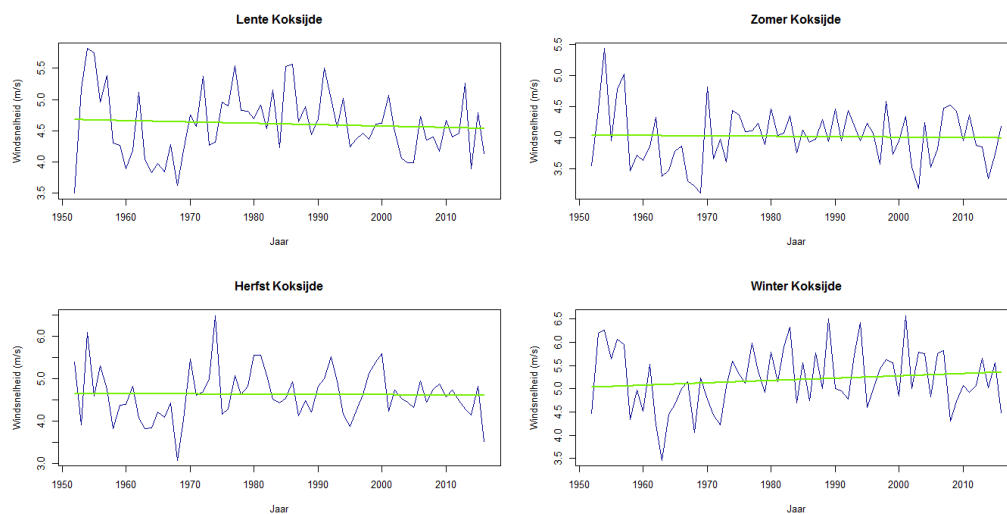
In Tabel 6.8 staan voor zowel Koksijde als Middelkerke schattingen van parameters weergegeven, bekomen via lineaire regressie. In de eerste kolom staat de schatting voor  $\beta_0$  weergegeven samen met de corresponderende standaarderror. Deze waarde voor  $\hat{\beta}_0$  geeft aan welke windsnelheid verwacht wordt in het jaar 1952 voor Koksijde (1951 voor de winter in Koksijde) en 1955 voor Middelkerke (1956 voor de lente in Middelkerke). Verder wordt ook een schatting voor  $\beta_1$  weergegeven met bijhorende standaarderror. Bovendien wordt voor deze  $\hat{\beta}_1$  ook een  $p$ -waarde meegegeven. Deze vertelt hoe waarschijnlijk het is om onder de nulhypothese dat  $\beta_1 = 0$  dergelijke waarde voor  $\hat{\beta}_1$  te vinden. Duikt deze waarde onder 0.05, dan kunnen we bij één individuele test de nulhypothese verwerpen op het 5% significantieniveau ten voordele van de alternatieve hypothese dat  $\beta_1 \neq 0$ .

Uit de gegevens in Tabel 6.8 kunnen we afleiden dat in Koksijde de  $p$ -waarde nooit onder 0.05 duikt. Dit betekent dat de eventueel aanwezige monotone lineaire trend niet statistisch significant is met of zonder correctie voor multiple testing. We vinden in de data dus geen evidentie om te zeggen dat de seizoensgemiddelde windsnelheid significant daalt of stijgt in Koksijde. Een heel ander verhaal doet zich voor in Middelkerke: daar daalt de  $p$ -waarde driemaal onder 0.05. We kunnen besluiten dat er seizoenen zijn waarin de gemiddelde windsnelheid daalt, met name in de lente, zomer en herfst (Bonferronicorrectie voor multiple testing). Omdat in elk van deze drie seizoenen de schatting voor  $\beta_1$  negatief is, betekent dat dat de jaarlijkse gemiddelde windsnelheid in de lente, zomer en herfst een significant dalende evolutie vertoont. In de winter daarentegen vinden we een  $p$ -waarde van 0.079. In de winter vertoont de gemiddelde windsnelheid dus geen significante trend.

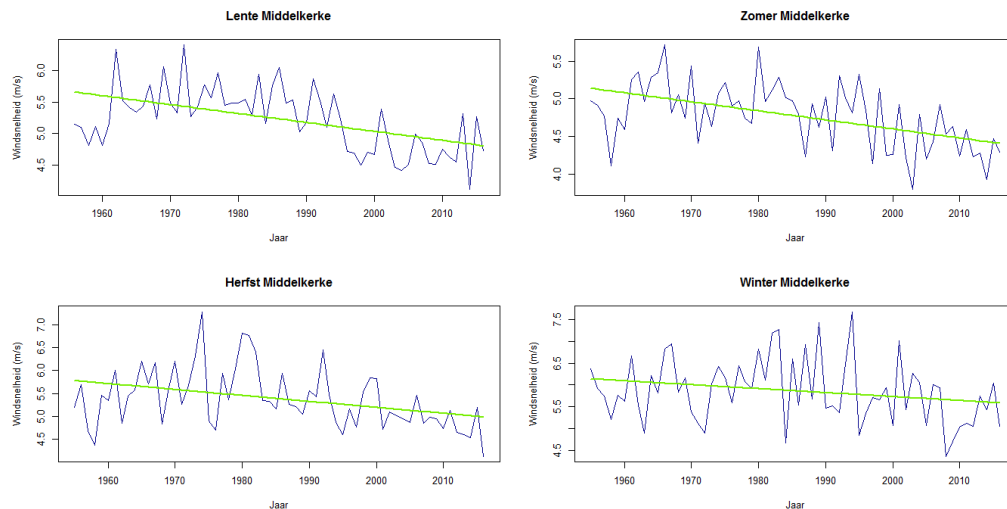


	$\hat{\beta}_0 \pm SE(\hat{\beta}_0)$	$\hat{\beta}_1 \pm SE(\hat{\beta}_1)$	$p$ -waarde t.o.v. $H_0 : \beta_1 = 0$ , $H_a : \beta_1 \neq 0$ (tweezijdig)
<b>Koksijde</b>			
<b>Lente</b>	4,677 ± 0,129	-0,0022 ± 0,0034	0.521
<b>Zomer</b>	4,034 ± 0,111	-0,0007 ± 0,003	0.807
<b>Herfst</b>	4,649 ± 0,1473	-0,00058 ± 0,003	0.884
<b>Winter</b>	5,032 ± 0,157	0,005 ± 0,004	0.238
<b>Jaar</b>	4,592 ± 0,1035	0,0004 ± 0,00271	0.881
<b>Middelkerke</b>			
<b>Lente</b>	5,670 ± 0,111	-0,0142 ± 0,003	$1.74 \cdot 10^{-5}$
<b>Zomer</b>	5,138 ± 0,093	-0,012 ± 0,003	$3.22 \cdot 10^{-5}$
<b>Herfst</b>	5,777 ± 0,150	-0,013 ± 0,004	0.003
<b>Winter</b>	6,136 ± 0,180	-0,009 ± 0.005	0.079
<b>Jaar</b>	5,698 ± 0,086	-0,012 ± 0,00238	$2.91 \cdot 10^{-6}$

Tabel 6.8: Resultaten van lineaire regressie op de jaar- en seizoensgemiddelde windsnelheden voor Koksijde en Middelkerke



Figuur 6.9: Seizoensgemiddelde windsnelheden in Koksijde met daarop de trendlijn

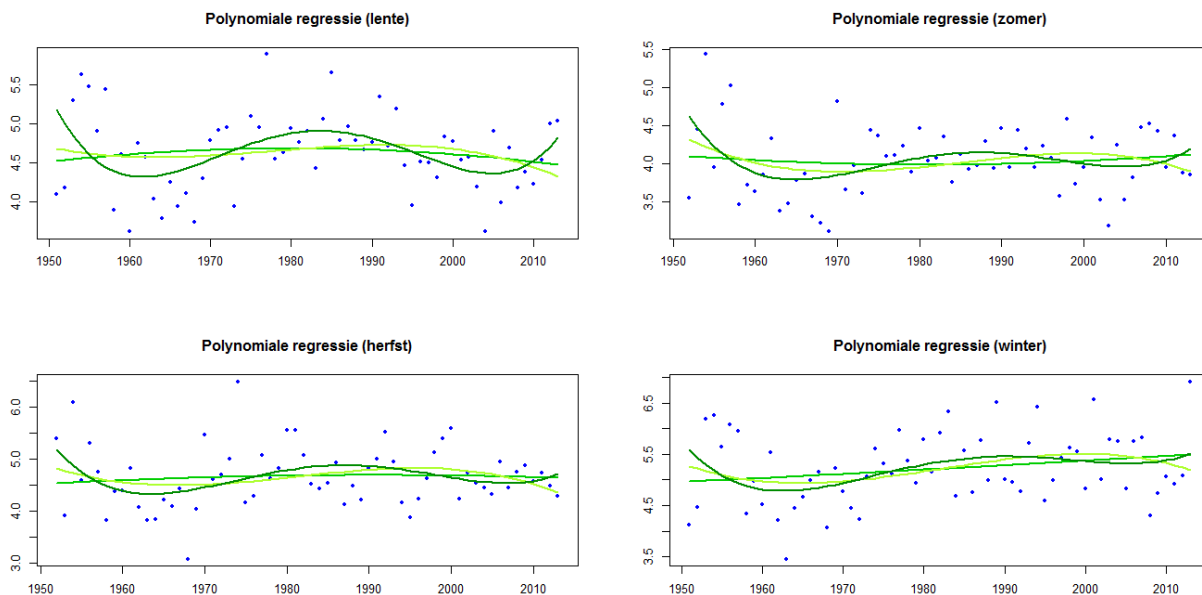


Figuur 6.10: Seizoensgemiddelde windsnelheden in Middelkerke met daarop de trendlijn

In Figuur 6.9 en Figuur 6.10 worden de seizoensgemiddelde windsnelheden over de jaren heen getoond met daarbovenop de rechte met als vergelijking  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ , met  $\hat{\beta}_0$  en  $\hat{\beta}_1$  de schattingen uit Tabel 6.8.

### 6.2.3 Polynomiale regressie

Figuur 6.11 toont het resultaat van polynomiale regressie toegepast op jaar- en seizoensgemiddelde windsnelheden voor zowel een polynoom van de tweede, derde als vierde graad.



Figuur 6.11: Polynomiale regressie met polynomen van graad 2 (grasgroen), graad 3 (lichtgroen) en graad 4 (donkergroen).

Uit Figuur 6.11 kan opgemerkt worden dat het vierdegraadspolynoom sterk begint te lijken op de kromme die via LOWESS in sectie 6.2.1 gevonden werd bij een niet al te grote waarde voor de parameter  $\alpha$ .

### 6.2.4 Niet-parametrische lineaire regressie

De jaarlijkse gemiddelde windsnelheden worden random gepermuteerd. De resultaten van de permutatietest worden in de Tabel 6.12 weergegeven. Hierbij wordt onder andere getoond hoeveel van de gegenereerde schatters voor  $\beta_1$  kleiner zijn dan de originele  $\hat{\beta}_1$ . Verder worden ook de waarden voor  $q$  en  $p$  weergegeven.

Koksijde	Fractie $\hat{\beta}_{1\text{gegenereerd}} < \hat{\beta}_{1\text{origineel}}$	$q = \frac{k}{s}$	$p = 2 \cdot \min(q, 1 - q)$
<b>Lente</b>	2549/10 000	0,2549	0.488
<b>Zomer</b>	4071/10 000	0,4071	0.820
<b>Herfst</b>	4429/10 000	0,4429	0.864
<b>Winter</b>	8091/10 000	0,8091	0.388
<b>Jaar</b>	5511/10 000	0,5511	0.898
<b>Middelkerke</b>			
<b>Lente</b>	0/10 000	0	0
<b>Zomer</b>	0/10 000	0	0
<b>Herfst</b>	2/10 000	0,0019	0.004
<b>Winter</b>	50/10 000	0,043	0.1
<b>Jaar</b>	0/10 000	0	0

Tabel 6.12: Bepalen van de significantie voor de permutatietest in Koksijde en Middelkerke

Uit deze resultaten volgt dat er voor de lente, zomer, herfst en winter in Koksijde geen significante trend is: geen van de  $p$ -waarden duikt onder 0.0125, de  $p$ -waarde die nodig is bij een individuele test om op het 5% significantieniveau de nulhypothese dat er geen enkel seizoen een trend in gemiddelde windsnelheden vertoont, met recht te mogen verwerpen.

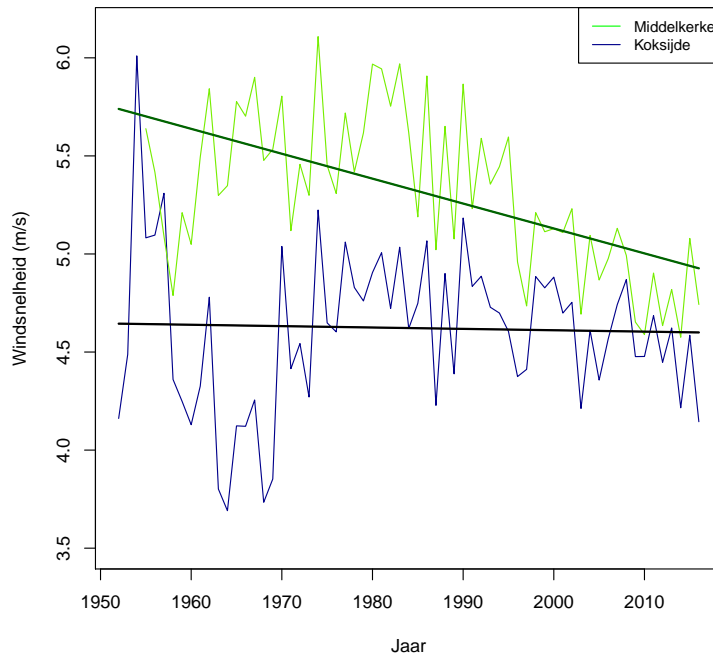
De resultaten in Middelkerke uit Tabel 6.12 laten een significant dalende trend zien in de lente, zomer en herfst ( $p$ -waarde  $< 0.0125$ ). In de winter daarentegen vinden we geen significant resultaat: de  $p$ -waarde bedraagt 0.1 en ligt dus een stuk boven de vereiste 0.0125.

De trend die in de jaargemiddelde windsnelheden in Koksijde en Middelkerke werd gevonden, wordt visueel voorgesteld in Figuur 6.13. De  $\hat{\beta}_0$  en  $\hat{\beta}_1$  verkregen van bij de lineaire regressie werden gebruikt om de rechten te plotten. Op deze figuur is duidelijk te zien dat de trend in Koksijde en Middelkerke sterk verschilt.

### 6.2.5 Mann-Kendalltest en Sen's slope

De Mann-Kendalltest is enkel toepasbaar wanneer er 'serial independence' is. Deze voorwaarde die de onafhankelijkheid van de metingen eist, komt erop neer dat de tijd tussen de steekproeven voldoende groot is zodat er geen correlatie is tussen de verschillende waarden. Aangezien bij jaar-, seizoens- of maandgemiddelden niet gewerkt wordt met rechtstreeks gemeten waarden, maar met jaargemiddelden, gaan we ervan uit dat de voorwaarde van onafhankelijkheid en identieke verdeling voldaan is en dat de Mann-Kendalltest bijgevolg toepasbaar is om een eventuele trend vast te stellen in de gemiddelden. De voorwaarde zou niet voldaan zijn mochten we per jaar met de volledige gedetailleerde verdeling van de wind rekening houden.

Figuur 6.15 en 6.16 tonen de seizoensgemiddelde windsnelheden voor Koksijde respectievelijk Middelkerke. De hoogste windsnelheden worden geobserveerd in de winter, wat uit de windrozen duidelijk zal worden (zie



Figuur 6.13: Trend in de jaarlijkse gemiddelde windsnelheid in Koksijde (zwart) en Middelkerke (donkergroen).

sectie 8.1). Aan de hand van Mann-Kendall en Sen's slope zal onderzocht worden of een trend zichtbaar is in de jaarlijkse gemiddelde windsnelheid.

Omdat de Mann-Kendalltest en Sen's slope in R zijn ingebouwd, zullen de berekeningen niet handmatig gedaan worden. Om toch een idee te hebben over hoe de  $S$ ,  $\text{Var } S$ ,  $z_{MK}$ , ... hun waarden verkrijgen, worden deze eenmaal (deels) met de hand berekend voor de lente in Koksijde.

**Mann-Kendall en Sen's slope voor de lente in Koksijde** Er zijn  $n = 65$  jaarlijkse gemiddelden beschikbaar. Dit betekent dat van  $\binom{65}{2} = 2080$  verschillen het teken bepaald moet worden om de teststatistiek  $S$  te bepalen. Men vindt dat  $S = -111$ . Nu kan ook de variantie van deze teststatistiek bepaald worden:

$$\text{Var } S = \frac{1}{18} (n(n-1)(2n+5)) = 31\,199.$$

Omdat  $S < 0$ , wordt de volgende formule gebruikt om de gestandaardiseerde teststatistiek te bepalen:

$$z_{MK} = \frac{S+1}{\sqrt{\text{Var } S}} = \frac{-110}{\sqrt{31\,199}} = -0,62.$$

Volgens Tabel 5.1 kan de nulhypothese niet verworpen worden. Op het 'minst strenge' significantieniveau moet  $|z_{MK}| > 1.645$ , maar dit is hier niet het geval. Dit betekent dat de nulhypothese, dat er geen trend is, zelfs niet verworpen wordt op het 10% significantieniveau. Er is m.a.w. geen significant dalende trend vast te stellen in de jaarlijkse gemiddelde windsnelheden in de lente in Koksijde.

Om de rechte door de datapunten te bepalen, zullen we ook nog Sen's slope berekenen. Voor  $m$  vindt men de waarde  $-0,0018$ . Een 95% betrouwbaarheidsinterval wordt gegeven door  $[-0,0103, 0,0065]$ . De intercept is gelijk aan 4,6741. Een plot van deze rechte door de datapunten staat weergegeven in Figuur 6.15.

**Andere resultaten** In Tabel 6.14 staan de resultaten van de Mann-Kendalltest en de waarden voor Sen's slope weergegeven van de jaar- en seizoensgemiddelden in Koksijde en Middelkerke. Deze werden berekend met behulp van R.

Koksijde	$z$	$p$ -waarde	$m$	$b$	95% BI $m$
<b>Lente</b>	-0,6	0.533	-0,0028	4,6401	[-0,0109; 0,0055]
<b>Zomer</b>	0,3	0.734	0,0011	3,9361	[-0,0059; 0,0075]
<b>Herfst</b>	0,4	0.700	0,0014	4,4929	[-0,007; 0,0082]
<b>Winter</b>	0,9	0.3849	0,0041	4,9658	[-0,0057; 0,0135]
<b>Jaar</b>	-0,2	0.879	-0,0005	4,6364	[-0,0063; 0,0058]
Middelkerke	$z$	$p$ -waarde	$m$	$b$	95% BI $m$
<b>Lente</b>	-3,8	0.00015	-0,0154	5,6639	[-0,0229; -0,0082]
<b>Zomer</b>	-3,7	0.00018	-0,0118	5,1695	[-0,0174; -0,0061]
<b>Herfst</b>	-3,3	0.00093	-0,0141	5,7284	[-0,02212; -0,006]
<b>Winter</b>	-1,8	0.0693	-0,0096	6,0447	[-0,0192; 0,0012]
<b>Jaar</b>	-4,1	$6.96 \cdot 10^{-6}$	-0,0111	5,65	[-0,0161; -0,0065]

Tabel 6.14: Trendanalyse van de jaarlijkse gemiddelde windsnelheid per seizoen in Koksijde en Middelkerke in de periode 1952/1955 – 2016. In de tabel staan volgende gegevens weergegeven: de gestandaardiseerde Mann-Kendallteststatistiek  $z$ , grootte van de  $p$ -waarde t.o.v. de nulhypothese dat er geen trend is, schattingen voor Sen's slope  $m$  en intercept  $b$  en een betrouwbaarheidsinterval voor  $m$ .

In Figuren 6.15 en Figuur 6.16 wordt de trendlijn, die geschat werd met de methode van Sen, getoond bovenop de jaar- en seizoensgemiddelde windsnelheden.

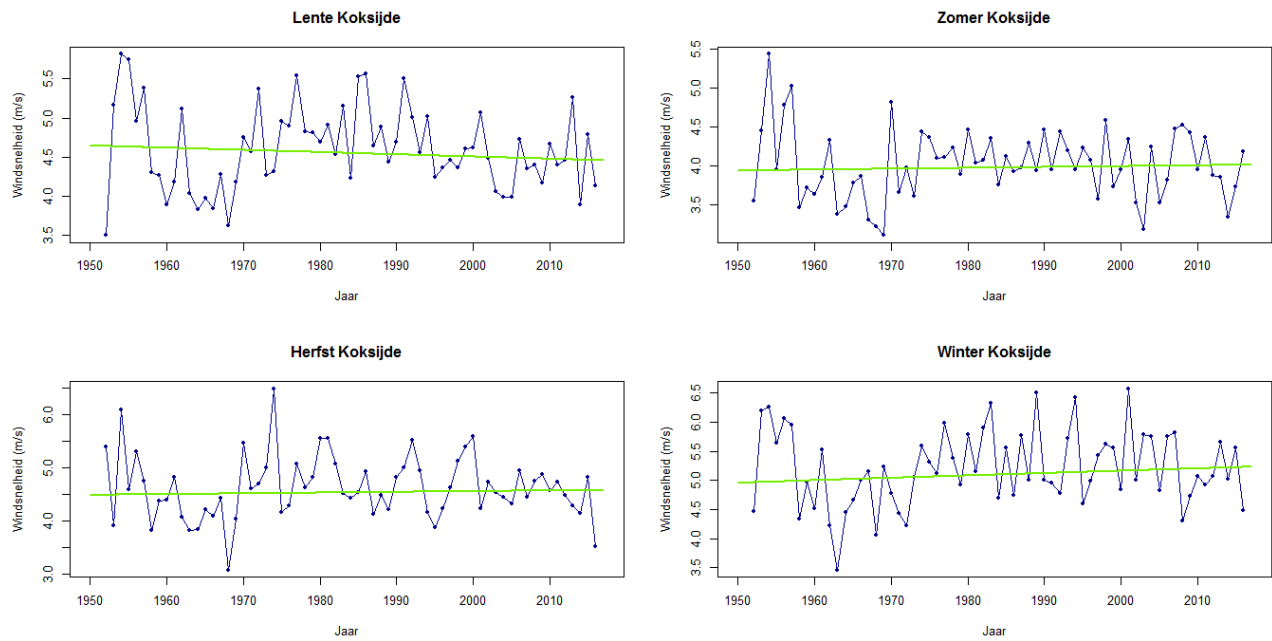
### 6.2.6 Samenvattende tabel en conclusies

De parametrische lineaire regressie, niet-parametrische lineaire regressie en Mann-Kendalltest zijn het met elkaar eens. In Koksijde is er geen enkele significante trend te vinden, noch bij de jaargemiddelden, noch bij de seizoensgemiddelden. In Middelkerke detecteren de drie testen een significante trend bij de jaargemiddelde windsnelheden en ook bij de seizoensgemiddelden in de lente, zomer en herfst. Schattingen van de regressie-coëfficiënt geven aan dat dit telkens gaat om een dalende trend. Wat betreft de winter vinden de drie testen een  $p$ -waarde die groter is dan 0.05 maar net onder 0.1 blijft. In de winter kan dus niet gesproken worden van een significante trend, noch van een significant dalende trend wanneer de Bonferroni-correctie in acht wordt genomen. De resultaten die bekomen werden (schattingen voor  $\beta_1$  en  $p$ -waarden) staan weergegeven in Tabel 6.17.

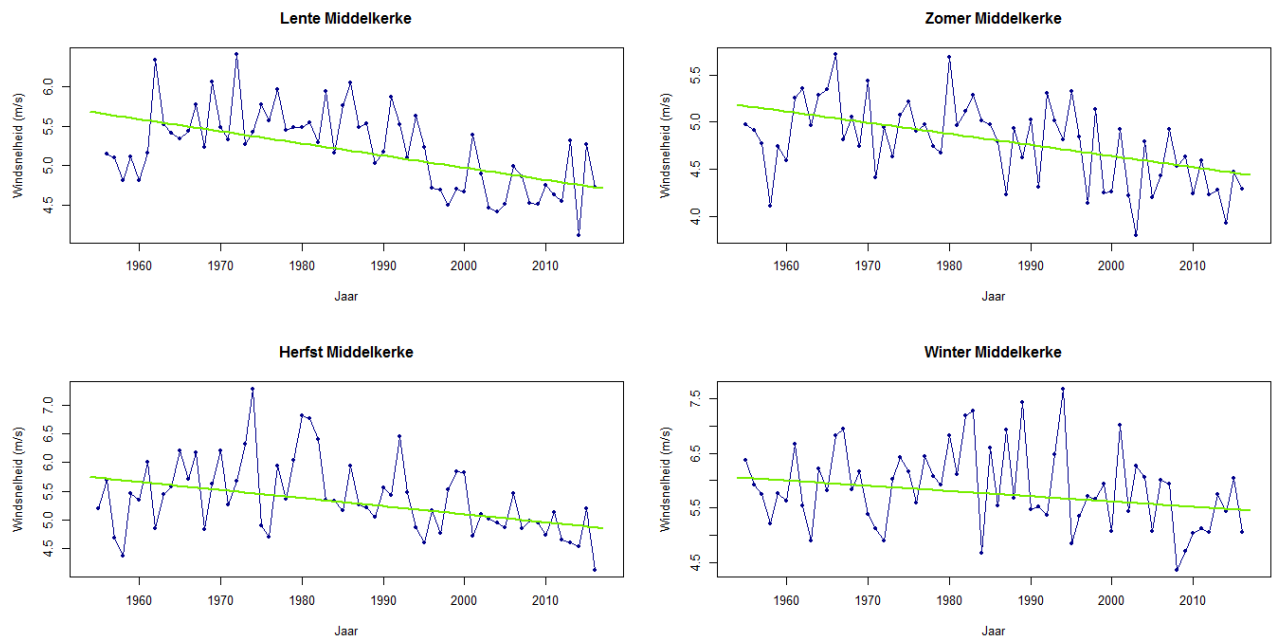
## 6.3 Maandgemiddelde windsnelheden

Tot dusver werden alleen evoluties bekeken van jaar- en seizoensgemiddelden. In deze sectie zullen ook de maandgemiddelden van 1952 in Koksijde en 1955 in Middelkerke tot 2016 kort onder de loep genomen worden.

In Figuur 6.20 werd voor elk jaar de maandelijkse gemiddelde windsnelheid berekend in Koksijde. Op deze figuur valt onmiddellijk op dat de windsnelheden in de zomermaanden een stuk lager liggen dan in de wintermaanden. Bovendien vertonen de windsnelheden in de wintermaanden grotere schommelingen dan in mei



Figuur 6.15: Seizoensgemiddelde windsnelheid in Koksijde van 1955 – 2016 met daarbij de trendlijn berekend met Sen's slope.



Figuur 6.16: Seizoensgemiddelde windsnelheid in Middelkerke van 1955–2016 met daarbij de trendlijn berekend met Sen's slope.

	Lineaire regressie		Permutatiemethode		Mann-Kendall	
<b>Koksijde</b>	$p$	$\hat{\beta}_1$	$p$	$\hat{\beta}_1$	$p$	$\hat{\beta}_1$
<b>Lente</b>	0.521	-0,0022	0.488	-0,0022	0.533	-0,0028
<b>Zomer</b>	0.807	-0,0007	0.82	-0,0007	0.734	0,0011
<b>Herfst</b>	0.884	-0,00058	0.864	-0.00058	0.700	0,0014
<b>Winter</b>	0.238	0,005	0.388	0,005	0.3849	0,0041
<b>Jaar</b>	0.881	0,0004	0.898	0,0004	0.879	-0,0005
<b>Middelkerke</b>	$p$	$\hat{\beta}_1$	$p$	$\hat{\beta}_1$	$p$	$\hat{\beta}_1$
<b>Lente</b>	$1.74 \cdot 10^{-5}$	-0,0142	0	-0,0142	0.00015	-0,0154
<b>Zomer</b>	$3.22 \cdot 10^{-5}$	-0,012	0	-0,012	0.00018	-0,0118
<b>Herfst</b>	0.003	-0,013	0.004	-0,013	0.00093	-0,0141
<b>Winter</b>	0.079	-0,009	0.1	-0,009	0.0693	-0,0096
<b>Jaar</b>	$2.91 \cdot 10^{-6}$	-0,012	0	-0,012	$6.96 \cdot 10^{-6}$	-0,0111

Tabel 6.17: Samenvattende tabel met waarden voor  $\hat{\beta}_1$  en  $p$ -waarden voor testen van trends in jaar- en seizoensgemiddelde windsnelheden in Koksijde en Middelkerke

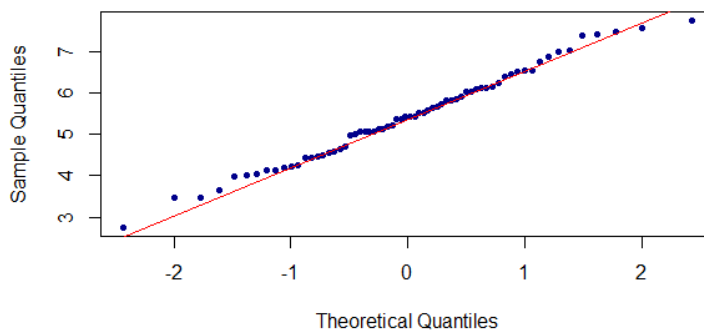
tot en met augustus. Op elk van de figuren wordt ook een trendlijn  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$  aangeduid, waarbij de schatters voor  $\beta_0$  en  $\beta_1$  werden bepaald aan de hand van parametrische lineaire regressie. Zoals te vermoeden viel uit de figuur is geen van de trends significant. Rekening houdend met de Bonferroni-correctie (zie sectie 5.6) kan pas van een significante trend gesproken worden wanneer de  $p$ -waarde lager zou zijn dan  $\frac{0.05}{12} = 0.004$ . In Tabel 6.21 is te zien dat dit voor geen enkele  $p$ -waarde het geval is. Dit betekent m.a.w. dat voor geen enkele maand de nulhypothese dat er geen monotone trend is in maandelijkse gemiddelde windsnelheid kan verworpen worden. De laagste  $p$ -waarde die gevonden wordt, is 0.0566 en de op één na laatste 0.267.

De schatters voor  $\beta_0$  en  $\beta_1$  werden berekend onder de assumpties die door een parametrische lineaire regressie verondersteld worden. De vijf assumpties die voldaan moeten zijn [41, 39], worden hieronder nagegaan.

We kunnen alvast opmerken dat met multicollineariteit geen rekening gehouden moet worden aangezien er slechts één responsvariabele is. Een QQ-plot in Figuur 6.18 laat zien dat de maandgemiddelden vrij goed de normale verdeling volgen. De scatterplots uit Figuur 6.20 en 6.23 tonen geen aanwijzingen dat een eventuele trend een niet-lineaire vorm zou aannemen. Om *serial correlation* na te gaan maken we gebruik van de Durbin-Watson-test. De vuistregel wordt gehanteerd dat waarden voor deze teststatistiek tussen 1,5 en 2,5 duiden op afwezigheid van correlatie. De Durbin-Watson-test geeft voor de maandgemiddelden in Koksijde en Middelkerke waarden tussen 1,57 en 2,26 wat betekent dat nulhypothese niet kan verworpen worden en dat de residuen dus normaal verdeeld zijn. De waarden worden getoond in Tabel 6.19. Tot slot moet ook nagegaan worden of er homoscedasticiteit is. We moeten met andere woorden nagaan of de stochastische variabelen allen dezelfde variantie hebben en niet toenemen in de tijd. Een test die hiervoor gebruikt kan worden, is de Goldfeld-Quandttest [68]. Deze test bevestigt dat er mag van uitgegaan worden dat de variabiliteit in de residuen niet stijgt over de tijd.

Eenzelfde analyse wordt gedaan voor Middelkerke. Maandgemiddelden met regressierechte  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$  worden voorgesteld in Figuur 6.23. Hier zien we echter heel wat maanden waar de trend significant lijkt te zijn. De  $p$ -waarden die gevonden worden in Middelkerke liggen een stuk lager dan deze in Koksijde. Dit was te verwachten: de trend in de jaarlijkse gemiddelden en seizoensgemiddelden waren significant in Middelkerke, maar niet in Koksijde. De  $p$ -waarden die gevonden werden voor de  $\hat{\beta}_1 = 0$  staan weergegeven in Tabel 6.22.

Op basis van deze tabel is te zien dat in 8 van de 12 maanden een significante trend is vast te stellen op het

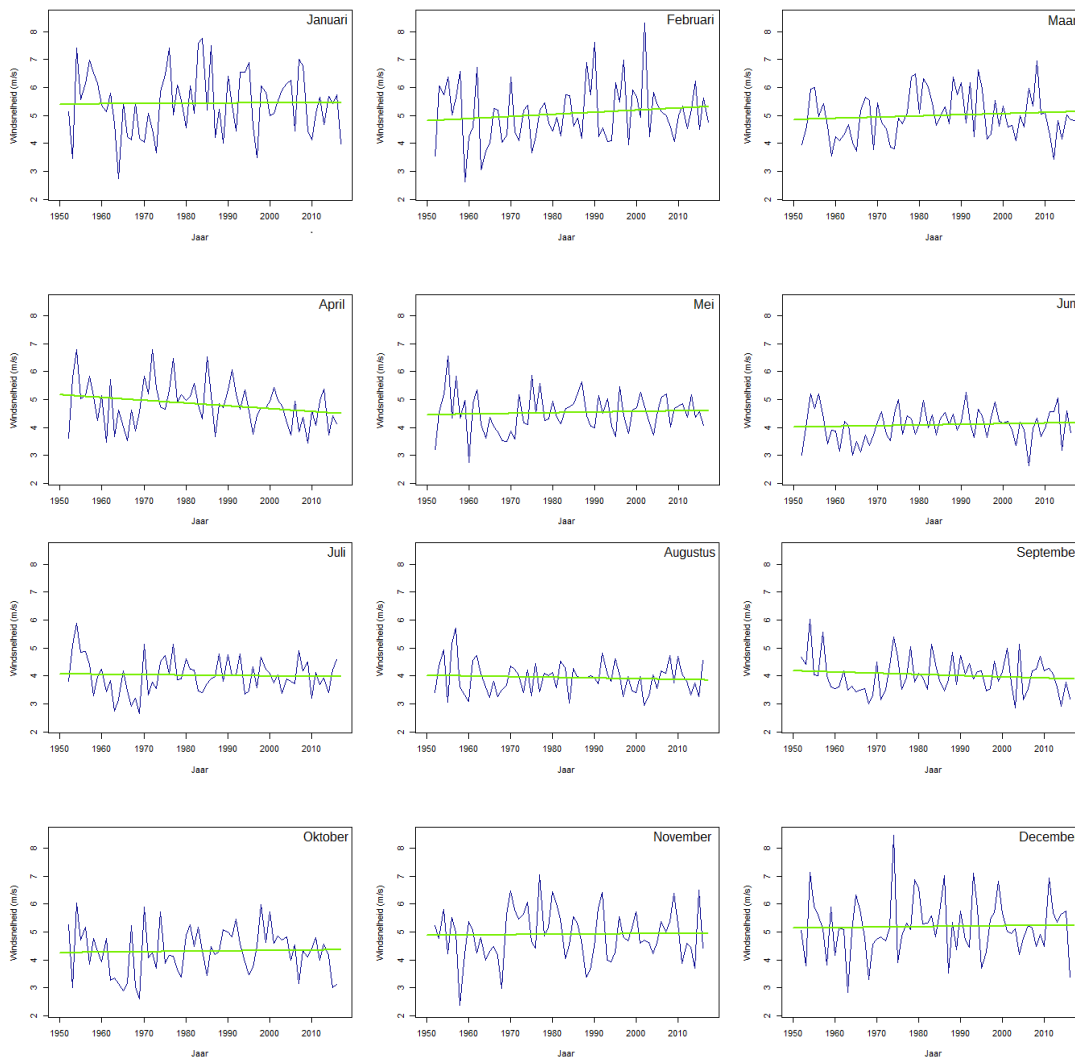


Figuur 6.18: Normale-QQ-plot van de kwantielen van de verdeling van de maandelijkse gemiddelde windsnelheden uitgezet tegenover de kwantielen uit de theoretische normale verdeling.

Koksijde				Middelkerke			
Maand	DW	Maand	DW	Maand	DW	Maand	DW
<b>Januari</b>	1,75	<b>Juli</b>	1,57	<b>Januari</b>	1,94	<b>Juli</b>	2,17
<b>Februari</b>	2,17	<b>Augustus</b>	1,98	<b>Februari</b>	2,26	<b>Augustus</b>	1,79
<b>Maart</b>	1,93	<b>September</b>	1,81	<b>Maart</b>	1,22	<b>September</b>	2,05
<b>April</b>	1,70	<b>Oktober</b>	1,78	<b>April</b>	1,92	<b>Oktober</b>	1,85
<b>Mei</b>	1,83	<b>November</b>	1,61	<b>Mei</b>	2,14	<b>November</b>	1,57
<b>Juni</b>	1,67	<b>December</b>	2,02	<b>Juni</b>	2,05	<b>December</b>	1,80

Tabel 6.19: Waarden van de Durbin-Watson-teststatistiek (DW) voor de maandgemiddelde windsnelheden in Koksijde en Middelkerke





Figuur 6.20: Maandelijks gemiddelde windsnelheden per jaar voor Koksijde

Maand	$p$ -waarde $\hat{\beta}_1 = 0$	Maand	$p$ -waarde $\hat{\beta}_1 = 0$
Januari	0.904	Juli	0.732
Februari	0.267	Augustus	0.511
Maart	0.434	September	0.311
April	0.0566	Oktober	0.786
Mei	0.618	November	0.843
Juni	0.527	December	0.803

Tabel 6.21:  $p$ -waarden voor de hypothese dat  $\hat{\beta}_1 = 0$  in Koksijde

Maand	$p$ -waarde $\hat{\beta}_1$	Maand	$p$ -waarde $\hat{\beta}_1$
Januari	0.237	Juli	0.0053
Februari	0.529	Augustus	0.00087
Maart	0.0499	September	0.0057
April	0.0000026	Oktober	0.0957
Mei	0.0131	November	0.0498
Juni	0.00698	December	0.103

Tabel 6.22:  $p$ -waarden voor de hypothese dat  $\hat{\beta}_1 = 0$  in Middelkerke

5% significantieniveau voor een individuele test. Na het toepassen van de Bonferroni-correctie zijn dit er nog slechts twee, namelijk in april en augustus. Op de vraag of er maanden zijn waarop de wind in de loop van de jaren afnam, is het antwoord dus ja, namelijk in april en augustus.

## 6.4 Dagen met krachtige wind

Tot nu toe werd enkel gekeken naar de verdeling van de wind en de trends die de jaarlijkse gemiddelde windsnelheden vertonen. Het is naar aanleiding van de klimaatverandering dat deze trends bestudeerd werden: het lijkt interessant om te weten of net zoals de temperatuur, ook de windsnelheden toenemen. Dit bleek echter niet zo te zijn. Integendeel, gemiddelde windsnelheden nemen af in Middelkerke en vertonen geen significante evolutie in Koksijde.

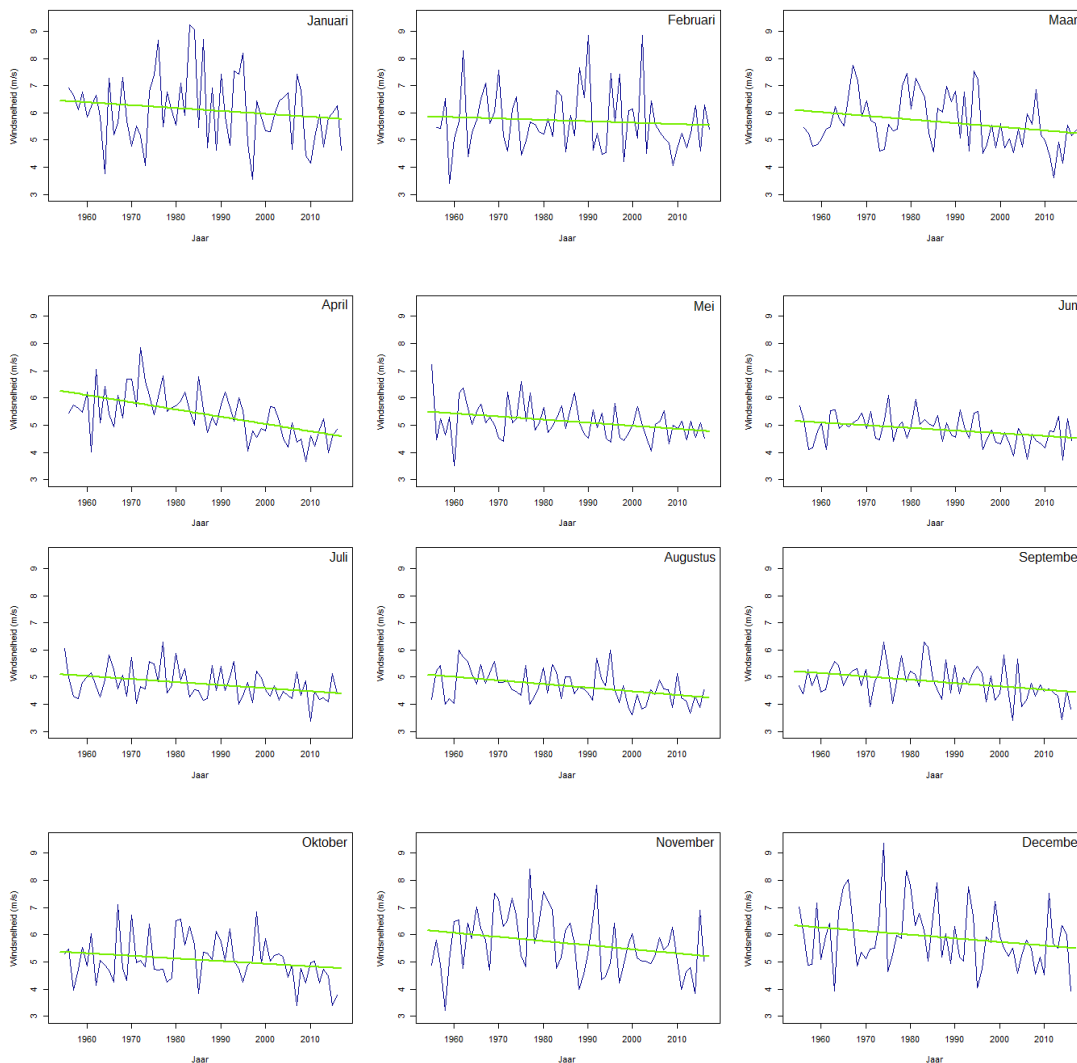
In deze sectie wordt een trendanalyse gedaan van het jaarlijks aantal metingen waarbij de windsnelheid een bepaalde grenswaarde overschrijdt. In Tabel 6.25 en 6.26 wordt deze analyse weergegeven.

Als eerste threshold werd gekozen voor 10 m/s. Dit is de windsnelheid vanaf dewelke men spreekt over een krachtige wind. Als laatste threshold werd gekozen voor 15 m/s. Dit komt overeen met een ‘harde wind’ (net geen stormachtige wind). Verder dan 15 m/s werd geen analyse meer gedaan omdat er per jaar nog maximum 8 metingen geweest zijn waarbij de windsnelheid 16 m/s overschreed. De aantallen zijn dan te klein om hierop nog een betekenisvolle analyse te doen die betrouwbare, interpreteerbare resultaten oplevert.

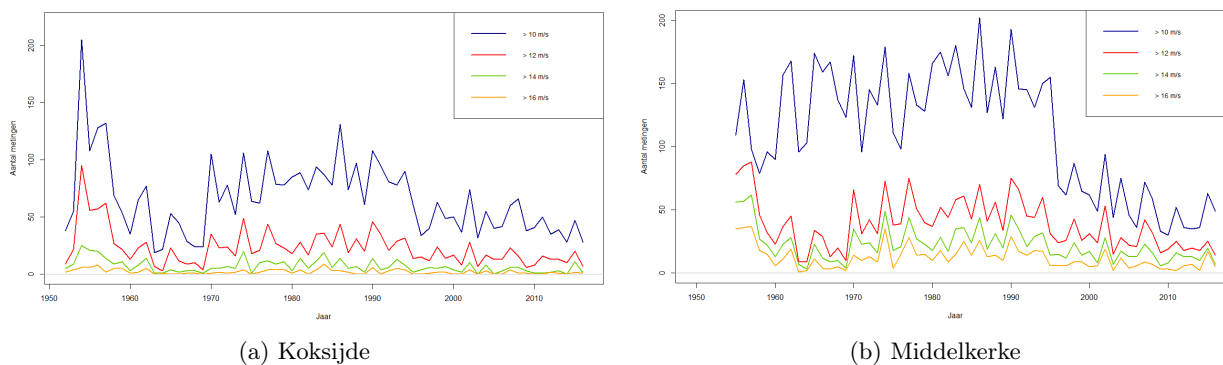
De resultaten van de trendanalyse weergegeven in Tabel 6.25 en Tabel 6.26 tonen allen hetzelfde resultaat: de nulhypothese dat er geen trend is, kan voor elke threshold tussen 10 m/s en 15 m/s verworpen worden op het 5% significantieniveau. Dit betekent dat het aantal metingen per jaar waarop de windsnelheid groter is dan de vooropgelegde threshold afneemt. In het geval waarbij de threshold bijvoorbeeld 10 m/s is, verwacht men dat het aantal windmetingen (met windsnelheid  $> 10$  m/s) per jaar gemiddeld afneemt met 0,557 meetmomenten (op een 66-tal per jaar).

In Figuur 6.24 worden voor 4 verschillende thresholds het aantal metingen per jaar voorgesteld waarbij de windsnelheid die bepaalde threshold overschrijdt. Op de figuur is duidelijk te zien dat er vóór het jaar 1970 eerst zeer veel dagen zijn waarop de thresholds 10 m/s en 12 m/s worden overschreden en daarna – tussen 1960 en 1970 – een dal. Ook hier is het opnieuw duidelijk dat dit dal de trendanalyse zal beïnvloeden: de trend zal minder significant dalend zijn wanneer we de volledige periode van 1952 tot 2016 in beschouwing nemen, dan wanneer we de analyse zouden doen voor de periode van pakweg 1972 tot 2016. Dit is een vaststelling die we ook reeds konden doen bij de jaar- en seizoensgemiddelde windsnelheden. Hierbij moet opgemerkt worden dat hier zeker geen eenvoudige lineaire trend uit naar voren komt, maar een grillige evolutie die in eerste-orde benadering wel een dalende trend vertoont.

**Opmerking** Naarmate de threshold groter wordt, zijn er jaren waarop het aantal dagen met windsnelheden hoger dan de vooropgelegde threshold dezelfde zijn. In dit geval zijn er wel *tied ranks*. De Mann-Kendalltest die



Figuur 6.23: Maandelijks gemiddelde windsnelheden per jaar voor Middelkerke



Figuur 6.24: Aantal metingen per jaar waarbij de windsnelheid groter was dan een zekere threshold (10 m/s, 12 m/s, 14 m/s en 16 m/s). De grijze horizontale lijn duidt aan waar het aantal dagen nul is.

Koksijde	$z$	$p$ – waarde	$m$	$b$	95% BI $m$
> 10 m/s	-2,3	0.01991	-0,3009	48,4453	[-0,5854; -0,04]
> 11 m/s	-2,5	0.012872	-0,2083	30,9167	[-0,4375; -0,0541]
> 12 m/s	-2,5	0.012929	-0,1389	19,1111	[-0,27272; -0,0182]
> 13 m/s	-2,3	0.018987	-0,1	11,7	[-0,1852; 0]
> 14 m/s	-2,5	0.011313	-0,0583	6,379	[-0,1132; 0]
> 15 m/s	-2,9	0.0042704	-0,0385	3,8846	[-0,0857; 0]

Tabel 6.25: Trendanalyse van het aantal dagen per jaar waarbij de windsnelheid een zekere waarde overschrijdt in Koksijde in de periode 1952 – 2016. In de tabel staan volgende gegevens weergegeven: de gestandaardiseerde Mann-Kendallteststatistiek  $z$ , grootte van de  $p$ -waarde, schattingen voor Sen's slope  $m$  en intercept  $b$  en een betrouwbaarheidsinterval voor  $m$ .

Middelkerke $z$	$p$ -waarde	$m$	$b$	95% BI $m$	
>10 m/s	-4,9	$1.23 \cdot 10^{-6}$	-0,8684	88,2237	[-1,1765; -0,55]
>11 m/s	-4,1	$3.77 \cdot 10^{-5}$	-0,5926	58,6111	[-0,8636; -0,34]
>12 m/s	-4,2	$2.26 \cdot 10^{-5}$	-0,4286	39,4286	[-0,6154; -0,2281]
>13 m/s	-3,8	0.00012	-0,2727	24,9545	[-0,4118; -0,1429]
>14 m/s	-4	$6.86 \cdot 10^{-5}$	-0,1887	15,0755	[-0,2692; -0,1071]
>15 m/s	-3,8	0.00012	-0,1346	10,4327	[-0,2; -0,0698]

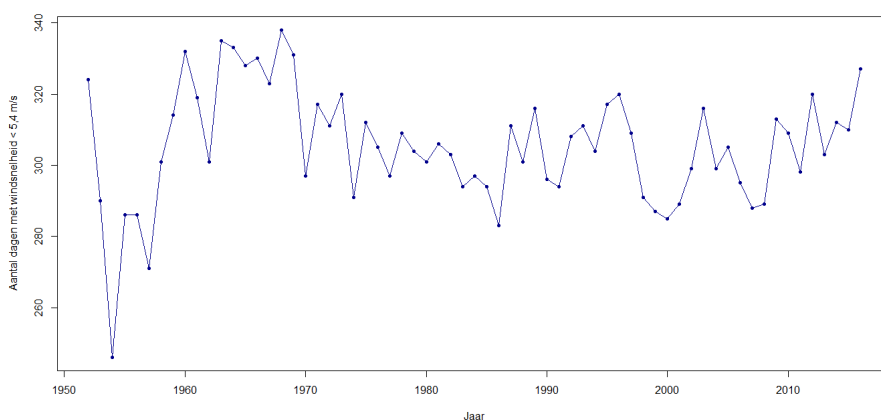
Tabel 6.26: Trendanalyse van het aantal dagen per jaar waarbij de windsnelheid een zekere waarde overschrijdt in Middelkerke in de periode 1955 – 2016. In de tabel staan volgende gegevens weergegeven: de gestandaardiseerde Mann-Kendallteststatistiek  $z$ , de  $p$ -waarde t.o.v. de nulhypothese dat er geen trend is in de data, schattingen voor Sen's slope  $m$  en intercept  $b$  en een 95% betrouwbaarheidsinterval voor  $m$ .

geïmplementeerd is in R houdt echter wel rekening met deze *tied ranks* bij het berekenen van de teststatistiek.

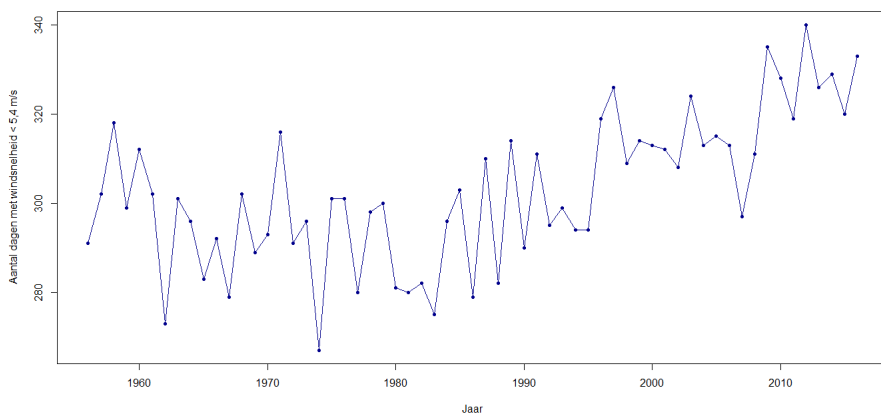
## 6.5 Dagen met zwakke tot matige wind

Voor de windmolenindustrie en de energiesector kan het van belang zijn te weten hoeveel dagen op een jaar de windsnelheden onder 5,4 m/s blijven. Dit is namelijk de windsnelheid waarbij windmolens nog amper zullen draaien. Uit de dataset zullen we het aantal meetmomenten halen waarop de windsnelheid onder 5,4 m/s blijft.

In Koksijde is de trend van het aantal meetmomenten waarop de windsnelheid onder 5,4 m/s bleef (van de ongeveer 1460 meetmomenten per jaar) te zien in Figuur 6.27. De Mann-Kendalltest geeft in Koksijde als gestandaardiseerde teststatistiek  $z = -1$ . Het aantal dagen in Koksijde waarop de gemiddelde windsnelheid onder 5,4 m/s duikt, is dus niet significant stijgend of dalend.



Figuur 6.27: Aantal meetmomenten waarop de windsnelheid onder 5,4 m/s bleef in Koksijde van 1952 tot 2016.



Figuur 6.28: Aantal dagen waarop de windsnelheid onder 5,4 m/s bleef in Middelkerke van 1956 tot 2016.

In Middelkerke daarentegen geeft de Mann-Kendalltest aan dat  $z = 4,7$ , wat wijst op een zeer significante (stijgende, want  $S > 0$ ) trend. Als waarde voor  $m$  vinden we 0,575, voor  $b$  vinden we 285,68. Dit betekent dat de verwachting is dat per jaar gemiddeld 0,575 metingen bijkomen waarop de windsnelheid kleiner is dan 5,4 m/s. Een plot van het aantal metingen per jaar<sup>1</sup> is te zien in Figuur 6.28. Voor de windenergiesector heeft

<sup>1</sup>Bij deze jaren is het jaar 1955 niet inbegrepen omdat er te veel waarden ontbreken in 1955.

dit als gevolg dat windmolens minder vaak zullen draaien en bijgevolg minder energie zullen produceren. We mogen hier echter niet te snel de conclusie trekken dat hierdoor de windmolens effectief minder zullen draaien. Daarvoor moeten we in feite ook het aantal dagen waarop het meer waait dan 10 Bft (= 24,5 m/s) in rekening brengen. In Middelkerke is er echter geen enkele windsnelheid die hoger is dan 24,4 m/s. Dit leidt ertoe dat de enige oorzaak die relevant is voor het niet draaien van windmolens de lage windsnelheden zijn. Aangezien het aantal dagen met zwakke tot matige wind toeneemt, doet dit vermoeden dat in Middelkerke met de jaren minder energie zou worden opgewekt, ten minste mochten er windmolens staan.

**Opmerking** In feite is het beter om per dag de maximale gemeten waarde te berekenen en dan te kijken hoeveel keer de windsnelheid onder 5,4 m/s blijft. Pas dan weten we echt hoeveel dagen de windmolens daardoor niet gedraaid hebben. Het is beter wanneer deze analyse kan uitgevoerd worden op een uurlijkse dataset dan op een zesuurlijkse dataset. Een andere bron vermeldt dat windmolens worden stilgelegd wanneer de windsnelheid kleiner wordt dan 3,5 m/s (zogenaamde cut-in speed) [66]. Bovendien wordt ook te kort door de bocht gegaan wanneer via de originele dataset wordt nagegaan hoeveel dagen de windsnelheden onder deze waarde bleven: windsnelheden op grote hoogte zijn namelijk verschillend van deze op 10 meter. Wanneer op ongeveer 40 meter hoogte de wind een snelheid heeft van 3,5 m/s kan met de power law (vergelijking 1.1) berekend worden dat windsnelheden op 10 meter dan ongeveer 2,87 m/s bedragen. Een correctere analyse zou bijgevolg zijn om de trend te onderzoeken in het aantal dagen waarop de windsnelheid 2,87 m/s niet overschrijdt.

## 6.6 Rukwinden

In het licht van de klimaatverandering en daarmee samenhangende ‘extreme weather events’ lijkt het interessant om de evolutie van de rukwinden te bestuderen. Een mogelijke vraag die bestudeerd zou kunnen worden is of de rukwinden heviger zijn geworden de afgelopen 70 jaar.

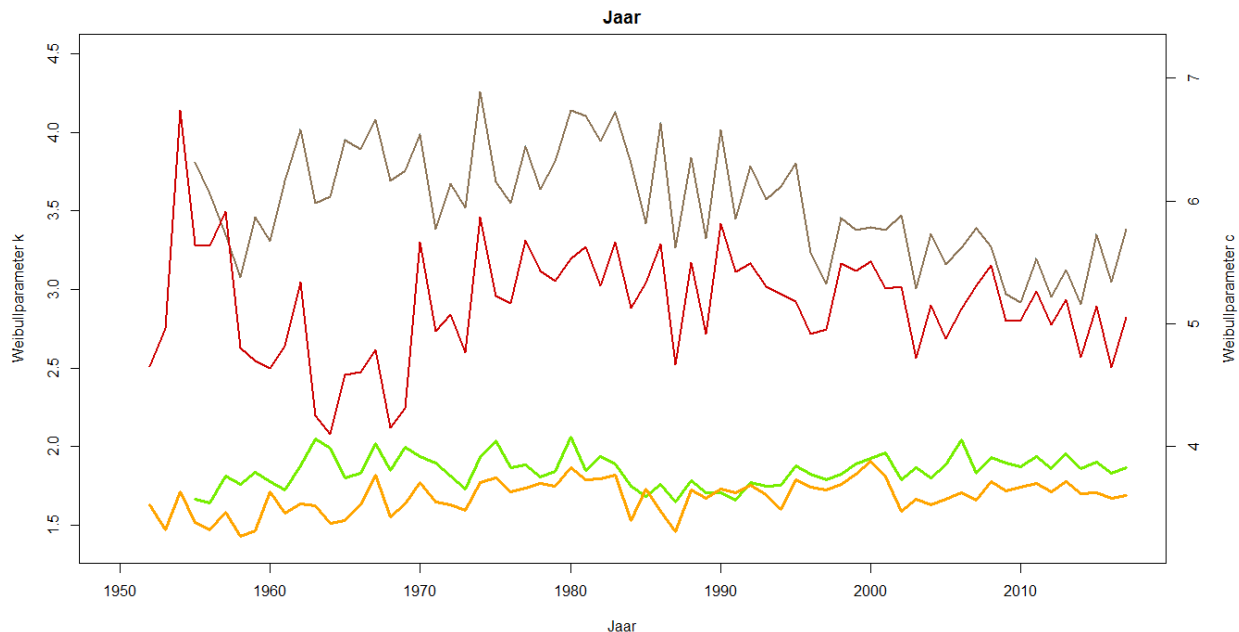
Zoals in Hoofdstuk 3 reeds werd aangehaald beschikken de datasefts van Koksijde en Middelkerke slechts over een beperkt aantal gegevens wat rukwinden betreft. De beschikbare gegevens laten toe om maar een analyse te doen van de rukwinden vanaf 1994. Deze analyse zou dus een periode van een kleine 20 jaar omvatten. Wanneer men spreekt over klimaat en in het bijzonder over een trend in een bepaalde variabele, bedoelt men een periode van minimum 30 jaar. Het houdt in dit geval dus weinig steek om een trendanalyse te doen op de rukwinden. We laten het hier dan ook achterwege.

## 6.7 Trendanalyse op de verdelingsparameters

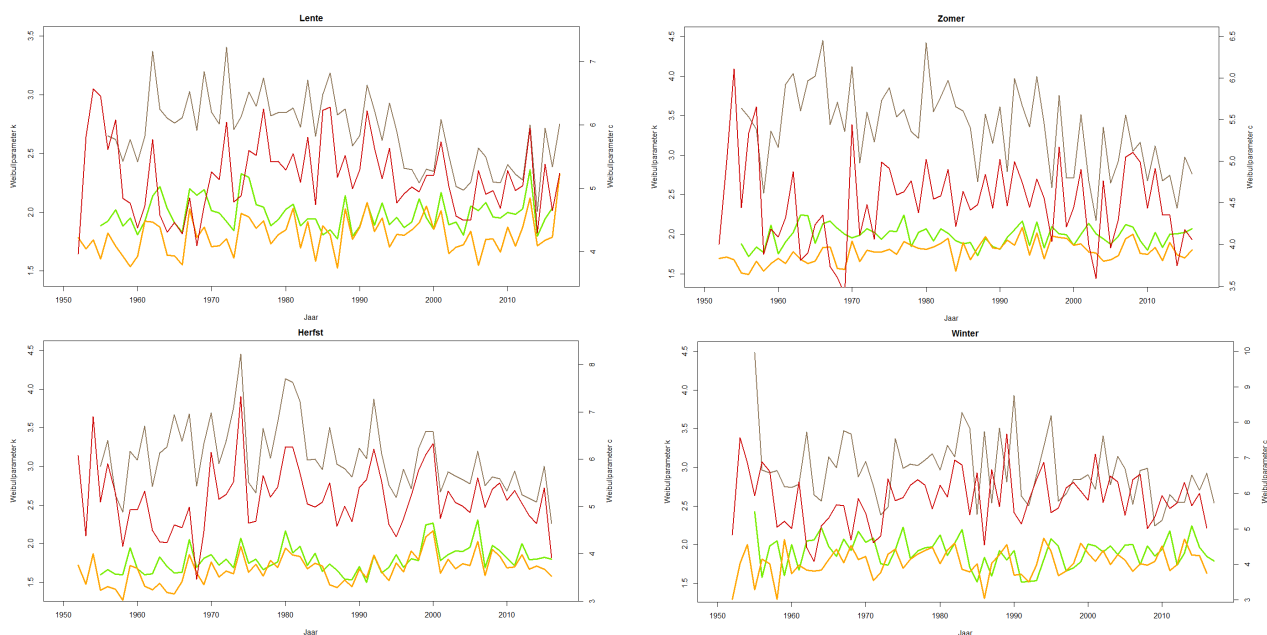
### 6.7.1 Visuele inspectie

In hoofdstuk 4 werden de Weibullparameters  $k$  en  $c$  geschat voor de volledige dataset van Koksijde en Middelkerke. Uit sectie 4.6 was al duidelijk dat de gemiddelde windsnelheden iets hoger liggen in Middelkerke. Dit wijst erop dat de schaalparameter in Middelkerke gemiddeld hoger ligt dan deze in Koksijde; dit is te zien in Figuur 6.29a. Eveneens zouden de schaal- en vormparameter geschat kunnen worden per jaar. Dit geeft voor Koksijde 65 en voor Middelkerke 62 waarden voor  $\hat{k}$  en  $\hat{c}$  die uitgezet kunnen worden in de tijd. Het kan nu interessant zijn om de evolutie over de jaren heen van deze parameters te bekijken. Een zekere evolutie in de vormparameter zou kunnen wijzen op het meer of juist minder ‘uitgesmeerd’ zijn van de windsnelheden. Een stijgende trend in de waarden van deze parameter zou dan duiden op windsnelheden die meer geconcentreerd zitten. Een stijgende of dalende trend in de schaalparameter  $c$  zou wijzen op hogere respectievelijk lagere windsnelheden. In Figuur 6.29a worden de jaarlijkse parameters  $\hat{k}$  en  $\hat{c}$  uitgezet in de tijd; in Figuur 6.29b wordt hetzelfde gedaan, maar voor de seizoenen.

In de figuren zijn een aantal bijzonderheden op te merken.



(a) Geschatte Weibullparameters  $k$  en  $c$  per jaar.



(b) Geschatte Weibullparameters  $k$  en  $c$  per seizoen.

Figuur 6.29: Geschatte Weibullparameters  $k$  en  $c$  voor Koksijde en Middelkerke. In rood en grijs staat de schaalparameter  $c$  voor Koksijde respectievelijk Middelkerke; in geel en groen de vormparameter  $k$  voor Koksijde respectievelijk Middelkerke.

- In de winter van 1955 in Middelkerke vinden we een zeer hoge waarde voor  $c$ . Deze werd berekend op basis van december 1955 en januari tot en met maart 1956. Hiervoor waren alle gegevens aanwezig, dus het gaat hier effectief om een correcte waarde.
- We merken op dat het verloop van de parameter  $c$  in Koksijde in elk seizoen een dal vertoont tijdens de jaren '60. Deze vaststelling wordt ook gedaan bij de jaarlijkse gemiddelde windsnelheden in Koksijde. Wanneer de gemiddelde windsnelheid lager is, betekent dit dat de Weibullkansverdelingsfunctie meer opgeschoven is naar links en dit gaat samen met een kleinere waarde voor de parameter  $c$ . Dus we zien hier een 'overeenstemmend' resultaat.

Om te weten te komen of er een trend zit in één van de parameters  $k$  of  $c$ , doen we een trendanalyse met de methode van Mann-Kendall en Sen's slope (sectie 5.5), net zoals in [50]. Voor de trendanalyse werden hier eveneens lineaire regressie als een permutatietest uitgevoerd.

### 6.7.2 Analyse per jaar

Uit Tabel 6.30 en Tabel 6.31 zien we dat in Middelkerke een significant dalende trend in de schaalparameters te vinden is. Dit was te verwachten aangezien in Middelkerke ook voor elk seizoen een dalende trend te bemerken was in de windsnelheden. Voor Koksijde vinden we een significant stijgende trend in de vormparameter. De resultaten van de lineaire regressie en permutatietest zijn terug te vinden in Tabel 6.30 en deze van de Mann-Kendalltest en Sen's slope in Tabel 6.31.

### 6.7.3 Analyse per seizoen

#### Koksijde

Voor de schaalparameter  $c$  vinden we in geen enkel seizoen een significant dalende of stijgende trend. Voor de parameter  $k$  daarentegen vinden we driemaal een significant stijgend resultaat, namelijk voor de lente, zomer en herfst. Voor de lente is  $\hat{\beta}_1 = 0,002$  en de  $p$ -waarde 0.0418; voor de zomer is  $\hat{\beta}_1 = 0,003$  en de  $p$ -waarde 0.0003 en voor de herfst vinden we dat  $\hat{\beta}_1 = 0,004$  en als  $p$ -waarde vinden we 0.0004. Dit betekent dat voor de zomer en de herfst een zeer significante stijging te zien is in de waarden van  $\hat{k}$ . Concreet betekent dit dat de Weibullkansdichtheidsfunctie iets 'gepiekter' wordt, wat betekent dat de windsnelheden iets meer geconcentreerd liggen. Voor de winter is bij de parameter  $k$  geen trend vast te stellen.

#### Middelkerke

In Middelkerke vinden we een tegenovergesteld resultaat. Hier vinden we in elk seizoen een significant dalende trend voor de schaalparameter  $c$ . Wanneer de schaalparameter afneemt, betekent dit dat de windsnelheden gemiddeld lager liggen. Dit kon ook worden vastgesteld bij de trendanalyse op de jaarlijkse gemiddelde windsnelheden in Middelkerke. Voor de vormparameter  $k$  is er alleen in de herfst een significant stijgende trend te zien.

## 6.8 Discussie

Zowel de parametrische lineaire regressie, de Mann-Kendalltest als de permutatietest geven aan dat de jaarlijkse gemiddelde windsnelheid in Koksijde niet significant daalt, terwijl dit in Middelkerke wel het geval is. Om een beter zicht te krijgen op dit resultaat, bekijken we de jaarlijkse gemiddelde windsnelheid van zowel Koksijde als Middelkerke gedurende de laatste 70 jaar (Figuur 6.13). In deze figuur, waar de jaarlijkse gemiddelde windsnelheden staan weergegeven, is te zien dat de jaarlijkse gemiddelde windsnelheden in Koksijde een dal



	Lineaire regressie				Permutatie			
	$k$	$c$	$k$	$c$	$k$	$c$	$k$	$c$
<b>Koksijde</b>	$\hat{\beta}_1$	$p$ -waarde	$\hat{\beta}_1$	$p$ -waarde	$\hat{\beta}_1$	$p$ -waarde	$\hat{\beta}_1$	$p$ -waarde
<b>Lente</b>	0,002045	0.0418	-0,0025	0.523	0,002045	0.0372	-0,0025	0.5168
<b>Zomer</b>	0,0031	0.0003	-0,0016	0.652	0,0031	0.0004	-0,0016	0.617
<b>Herfst</b>	0,0042	0.00042	0,00017	0.971	0,0042	0	0,00017	0.963
<b>Winter</b>	0,0014	0.221	0,0057	0.249	0,0014	0.1161	0,0057	0.2472
<b>Jaar</b>	0,0025	0.00016	0,00097	0.747	0,0025	0.001	0,00097	0.7502
<b>Middelkerke</b>	$\hat{\beta}_1$	$p$ -waarde	$\hat{\beta}_1$	$p$ -waarde	$\hat{\beta}_1$	$p$ -waarde	$\hat{\beta}_1$	$p$ -waarde
<b>Lente</b>	0,00118	0.350	-0,0184	$2.8 \cdot 10^{-6}$	0,00118	0.356	-0,0184	0
<b>Zomer</b>	0,000175	0.848	-0,0134	$3.2 \cdot 10^{-5}$	0,000175	0.846	-0,0134	0
<b>Herfst</b>	0,0033	0.0069	-0,0144	0.0042	0,0033	0.005	-0,0144	0.004
<b>Winter</b>	-0,0014	0,3057	-0,0145	0.0192	-0,0014	0.3214	-0,0145	0.0248
<b>Jaar</b>	0,00071	0.318	-0,01335	$3.4 \cdot 10^{-6}$	0,00071	0.2942	-0,01335	0

Tabel 6.30: Resultaten van de trendanalyse aan de hand van parametrische en niet-parametrische lineaire regressie op de geschatte Weibullparameters in Koksijde en Middelkerke

	$k$				$c$			
	$z_{MK}$	$p$ -waarde	$m$	$b$	$z_{MK}$	$p$ -waarde	$m$	$b$
<b>Koksijde</b>	$z_{MK}$	$p$ -waarde	$m$	$b$	$z_{MK}$	$p$ -waarde	$m$	$b$
<b>Lente</b>	1,2	0.215	0,0013	1,762	-0,5	0.591	-0,003	5,227
<b>Zomer</b>	3,3	0.0011	0,0032	1,694	0	0.905	0	4,447
<b>Herfst</b>	3,1	0.00168	0,0042	1,507	0,6	0.559	0,0027	5,007
<b>Winter</b>	0,8	0.4117	0,0011	1,727	0,9	0.368	0,0051	5,625
<b>Jaar</b>	3,2	0.00128	0,0025	1,60	0,1	0.959	$3 \cdot 10^{-4}$	5,161
<b>Middelkerke</b>	$z_{MK}$	$p$ -waarde	$m$	$b$	$z_{MK}$	$p$ -waarde	$m$	$b$
<b>Lente</b>	0	1	0	1,952	-4,1	$4 \cdot 10^{-5}$	-0,0189	6,49
<b>Zomer</b>	0	1	0	1,99	-3,8	0.00017	-0,0136	5,845
<b>Herfst</b>	3,1	0.0017	0,0033	1,654	-3,2	0.0013	-0,016	6,43
<b>Winter</b>	-1	0.331	-0,0014	1,98	-1,7	0.0824	-0,0104	6,89
<b>Jaar</b>	1,4	0.174	0,001	1,815	-4,5	$6.2 \cdot 10^{-6}$	-0,0148	6,43

Tabel 6.31: Resultaten van de trendanalyse aan de hand van Mann-Kendalltest en Sen's slope op de geschatte Weibullparameters op de jaar- en seizoensgemiddelde windsnelheden in Koksijde en Middelkerke

vertonen in de jaren '60. In Middelkerke is dit dal niet aanwezig. We zien ook dat de jaarlijkse gemiddelde windsnelheden vanaf 1970 hetzelfde patroon vertonen in zowel Koksijde als Middelkerke. De gemiddelde windsnelheid ligt in Middelkerke dan wel systematisch hoger, maar vertoont een zeer gelijkaardig patroon als dat van Koksijde. Het is zeer aannemelijk dat de aanwezigheid van dit dal in Koksijde en de afwezigheid ervan in Middelkerke het verschil in resultaten van beide testen verklaart: de gemiddelde windsnelheid neemt in Middelkerke meer af omdat er geen dal is dat de trendlijn beïnvloedt. Bovendien heeft het dal dat in Koksijde is waar te nemen een invloed op de squared error: voor de jaargemiddelde windsnelheden in Koksijde vinden we dat de squared error gelijk is aan 10,58, in Middelkerke is de squared error maar 6,74.

**Complexer model** Het is duidelijk dat in dit hoofdstuk over trendanalyse een vrij eenvoudig model werd gebruikt. Door te werken met jaar- en seizoensgemiddelden moet geen rekening gehouden worden met onder andere 'serial correlation'. Mocht er gewerkt worden met alle individuele observaties, dan is er wel 'serial correlation' en dat moet dan ook in het model verwerkt worden met behulp van een variantie-covariantiematrix. Verder bevatten de originele, individuele uitkomsten veel detailinformatie waar rekening mee gehouden moet worden, daar waar in een model waar enkel met gemiddelden wordt gewerkt deze gedetailleerde informatie niet meer bevat zit.

Een andere mogelijkheid is een model waarbij de data over de vier seizoenen in hetzelfde model gestopt worden. In dit geval moet gewerkt worden met een dummyvariabele, een indicatorfunctie die de waarde 0 of 1 aanneemt, voor elk seizoen. Dit model kan zowel gebruikt worden in het geval enkel met gemiddelde waarden wordt gewerkt als in het geval waarbij met alle individuele waarden wordt gewerkt. De data waarmee wordt gewerkt, hangt af van de onderzoeksvraag die beantwoord wil worden.

**Seriële correlatie** In de trendanalyse werd steeds gewerkt met jaar- of seizoensgemiddelden. Hierbij werd verondersteld dat de observaties niet gecorreleerd waren over de tijd, omdat juist met gemiddelden gewerkt wordt, met samenvattingsmaten die geen gedetailleerde informatie meer bevatten. Deze assumptie van ongecorreleerd zijn was ook één van de vereenvoudigingen die werd doorgevoerd in het model, maar deze verdient hierbij toch wat meer aandacht.

Een geschikte test om 'serial correlation' te detecteren in de residuen van een regressieanalyse is de Durbin-Watson test. De teststatistiek  $d$  neemt een waarde aan tussen 0 en 4. Hierbij geeft een waarde van  $d = 2$  aan dat er geen correlatie is. De nulhypothese  $H_0$  houdt in dat de autocorrelatie gelijk is aan nul; de alternatieve hypothese zegt dat de autocorrelatie groter is dan nul<sup>2</sup>. De bekomen teststatistiek  $d$  moet vervolgens vergeleken worden met de waarden voor  $d_L$  en  $d_U$  die vastliggen voor een welbepaald significantieniveau  $\alpha$  en in tabellen worden voorgesteld. We kiezen als significantieniveau  $\alpha = 0.05$  en  $n = 60$ . De tabel geeft hiervoor als kritische waarden  $d_L = 1,55$  en  $d_U = 1,62$ . Hierbij staat  $d_L$  voor een kritische ondergrens en  $d_U$  voor een kritische bovengrens. De nulhypothese wordt niet verworpen wanneer  $d > d_U$ . Wanneer  $d < d_L$  wordt de nulhypothese wel verworpen en is er sprake van positieve autocorrelatie. Wanneer de waarde voor  $d$  ligt in het interval  $[d_L, d_U]$ , dan is de test onduidelijk.

Bij de jaar- en seizoensgemiddelden van de windsnelheden wordt in sommige gevallen de nulhypothese verworpen, in andere gevallen niet.

De Durbin-Watson teststatistiek [39] geeft voor de jaargemiddelden in Koksijde een waarde van 1,30 en voor die van Middelkerke 1,39. Beide waarden vallen buiten het interval  $[1,55; 1,62]$ , wat betekent dat de nulhypothese verworpen wordt en de residuen positief gecorreleerd zijn. De geschatte correlatie tussen opeenvolgende residuen is voor Koksijde 0,33, voor Middelkerke is dat 0,30.

Voor de seizoensgemiddelde windsnelheden zijn de resultaten van de Durbin-Watson test uiteenlopend: in Koksijde en Middelkerke vinden we voor de lente respectievelijk 1,33 en 1,46, voor de zomer 1,77 en 1,92, voor de herfst 1,59 en 1,38 en voor de winter 1,73 en 1,98. Het is opvallend dat voor de zomer en de winter de nulhypothese niet verworpen wordt en de test dus geen positieve seriële correlatie aangeeft. De waarden

<sup>2</sup>Het is blijkbaar zo dat omdat de meeste regressieproblemen tijdsreeksen bevatten die een positieve autocorrelatie vertonen, geopteerd wordt voor deze eenzijdige alternatieve hypothese.

voor  $d$  liggen in deze gevallen ook vrij dicht bij 2 en de correlatiecoëfficiënt<sup>3</sup> ligt voor de vier gevallen tussen 0,0048 en 0,12. Voor de lente (zowel in Koksijde als in Middelkerke) geeft de Durbin-Watson-teststatistiek met een waarde van 1,33 en 1,46 dan weer aan dat de nulhypothese verworpen wordt ten voordele van de alternatieve hypothese dat er positieve seriële correlatie is. De correlatiecoëfficiënt is hier 0,30 voor Koksijde en 0,26 voor Middelkerke. Voor de herfst is in Koksijde  $d = 1,59$ , wat betekent dat de test onbeslist is. Als correlatiecoëfficiënt vinden we 0,17. Voor de herfst in Koksijde verkrijgen we de waarde  $d = 1,38$ , wat opnieuw betekent dat er positieve seriële correlatie is en met 0,29 als grootte van de geschatte correlatie.

Met deze test hebben we opgemerkt dat de jaar- en seizoensgemiddelden niet volledig ongecorrleerd zijn en dat voor een eventuele latere trendanalyse hiermee rekening moet gehouden worden bij het kiezen van optimale methoden en testen.

**Homogenisering** Een ander aspect waar rekening mee gehouden kan worden, is de homogenisering van tijdsreeksen. Hiervoor is een grondige kennis nodig van de verplaatsingen van de meettoestellen en wijzigingen ervan. Een test die gebruikt kan worden om een sprong in de data te detecteren, is deze van Pettitt [40, 8, 48]. Zie ook sectie 9.3.1.

---

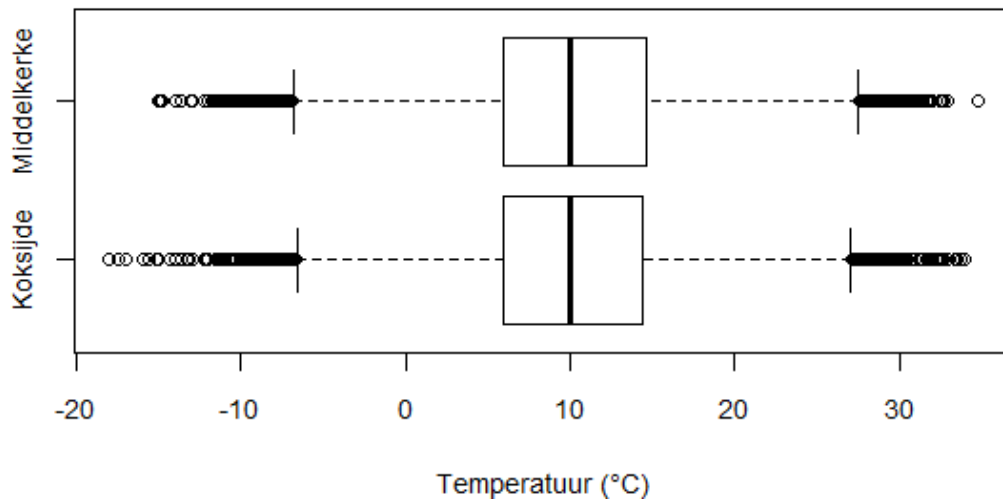
<sup>3</sup>Om de correlatie tussen opeenvolgende residuen te bepalen, werd telkens de correlatiecoëfficiënt bepaald van de punten  $(e_i, e_{i+1})$ .

# Hoofdstuk 7

## Kusttemperatuur

### 7.1 Beschrijvende analyse

Voor de periode van 1952 tot 2017 bedraagt de gemiddelde temperatuur in Koksijde  $10,00^{\circ}\text{C}$  en in Middelkerke  $10,02^{\circ}\text{C}$ . De steekproefstandaardafwijking in Koksijde bedraagt  $6,20^{\circ}\text{C}$ , in Middelkerke is dat  $6,04^{\circ}\text{C}$ . In Figuur 7.1 worden boxplots van de zeesurlijkse gemiddelde temperaturen in Koksijde en Middelkerke weergegeven. Zoals te verwachten was, zijn deze zeer gelijkaardig.

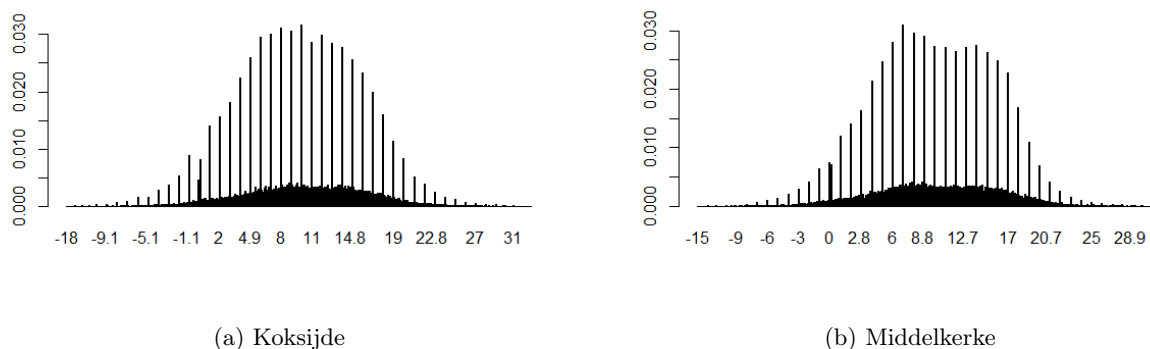


Figuur 7.1: Boxplot van de temperaturen in Koksijde (1952–2016) en Middelkerke (1955–2016)

#### 7.1.1 Verdeling van de temperatuur

Vooraleer een trendanalyse gedaan zal worden op de temperaturen in Koksijde en Middelkerke, zullen eerst een aantal figuren gemaakt worden die een beeld geven van de verdeling van de temperatuur. Het staafdiagram in Figuur 7.2 geeft een idee over het aantal verschillende waarden dat voor de temperatuur werd opgemeten. Hierop is te zien dat er veel meer gehele waarden werden geregistreerd dan decimale waarden, en dit zowel in Koksijde als in Middelkerke. Decimale waarden werden in beide kustgemeenten pas genoteerd vanaf 1 januari 1982.

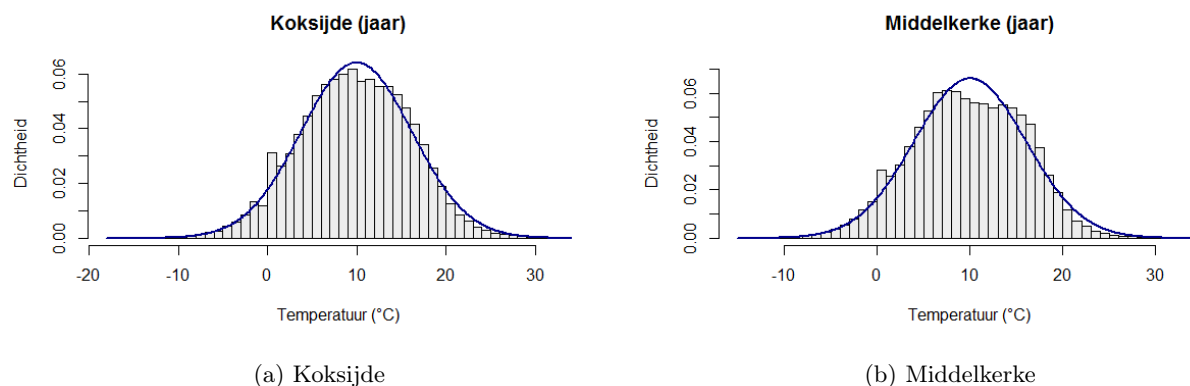
Figuur 7.3 toont histogrammen van de temperatuur in Koksijde en Middelkerke. Het steekproefgemiddelde en de standaarddeviatie zijn in Koksijde  $10,00$  en  $6,20$ ; in Middelkerke is dat  $10,02$  en  $6,04$ . Op deze histogrammen



Figuur 7.2: Staafdiagram voor de temperatuur

is te zien dat de temperaturen wel min of meer een normale verdeling volgen – in Koksijde meer dan in Middelkerke – maar er is ook te zien dat er rond nul te veel waarden zijn<sup>1</sup> en in Middelkerke zit er een ‘deuk’ in de top van het histogram. Om te bekijken in welke mate de verdeling afwijkt van een normale verdeling is het interessant om een QQ-plot voor beiden te maken. In Figuur 7.4 is inderdaad te zien dat de temperaturen in Koksijde de normale verdeling beter volgen dan deze in Middelkerke. Bij de QQ-plot van Middelkerke zien we een vrij grote afwijking aan de rechterkant. Bij Koksijde beperken de afwijkingen zich tot de staarten van de verdeling.

Zoals welbekend is de normale verdeling een van de meest natuurlijk voorkomende verdelingen, mede door de centrale limietstelling, die stelt dat steekproefgemiddelden naar een normale verdeling neigen als hun steekproefgrootte stijgt. Het is de familie verdelingen met parameters  $\mu$  (gemiddelde, eerste moment) en  $\sigma$  (standaarddeviatie, tweede moment) en alle andere hogereordemomenten gelijk aan 0.

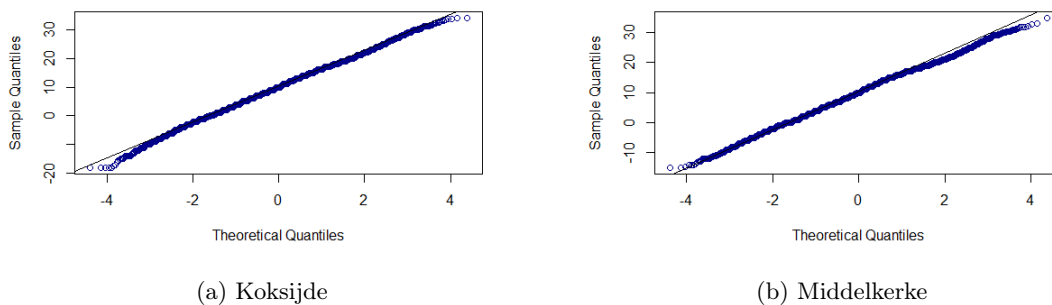


Figuur 7.3: Histogrammen van de zesuursgemiddelde temperaturen met daarboven de normale verdeling met parameters geschat door de maximum-likelihoodschatters (steekproefgemiddelde en standaarddeviatie).

## 7.2 Trendanalyse op de temperatuur

De opwarming van het klimaat is onweerlegbaar aanwezig en sinds de jaren '50 doen zich veranderingen voor die ongezien zijn de voorbije decennia en eeuwen. De atmosfeer en de oceanen zijn warmer geworden, de hoeveelheden ijs en sneeuw zijn geslonken, het zeeniveau is gestegen en de concentratie broeikasgassen is

<sup>1</sup>De waarde 0 komt 77 keer voor in de dataset, de waarden 0,1 en  $-0,1$  elk meer dan 600 keer.



Figuur 7.4: QQ-plot voor de temperaturen ten opzichte van de normale verdeling

toegenomen [55]. Wetenschappers zijn van oordeel dat deze stijging onder andere verband houdt met een grotere concentratie broeikasgassen in de atmosfeer die de energiebalans van de Aarde verstoren [55]. Omdat deze stijgende trend overal ter wereld waarneembaar is, is het interessant na te gaan of deze ook aan de kust waarneembaar is. We vermoeden natuurlijk dat dit zo zal zijn, maar het loont zeker de moeite om ook hiervoor nog eens een trendanalyse te doen.

Net zoals voor de windsnelheid werd gedaan, zal ook voor de jaar- en seizoensgemiddelde temperatuur gekeken worden of een eventuele trend is vast te stellen. Omdat uit de beschrijvende analyse blijkt dat de temperaturen zowel in Koksijde als Middelkerke de normale verdeling vrij goed benaderen, kan voor het onderzoeken van een trend parametrische lineaire regressie gebruikt worden. We merken hierbij echter op dat door het werken met gemiddelde windsnelheden ook hier de centrale limietstelling haar rol zal spelen en deze dus bij benadering normaal verdeeld zullen zijn. Op de jaar- en seizoensgemiddelde temperaturen zullen ook de twee andere methoden voor trendanalyse worden toegepast.

### 7.2.1 Lineaire regressie

Met lineaire regressie bekomen we de waarden weergegeven in Tabel 7.5. Op basis van deze gegevens kunnen we besluiten dat in Koksijde over de jaren heen én in elk van de seizoenen apart een globaal stijgende trend is vast te stellen, ook na Bonferroni-correctie. In elk seizoen kan de nulhypothese dat er geen globale trend is, verworpen worden op het 1% significantieniveau. In Figuur 7.6 worden de rechten getoond die de trend van het verloop van de jaarlijkse gemiddelde temperatuur weergeven.

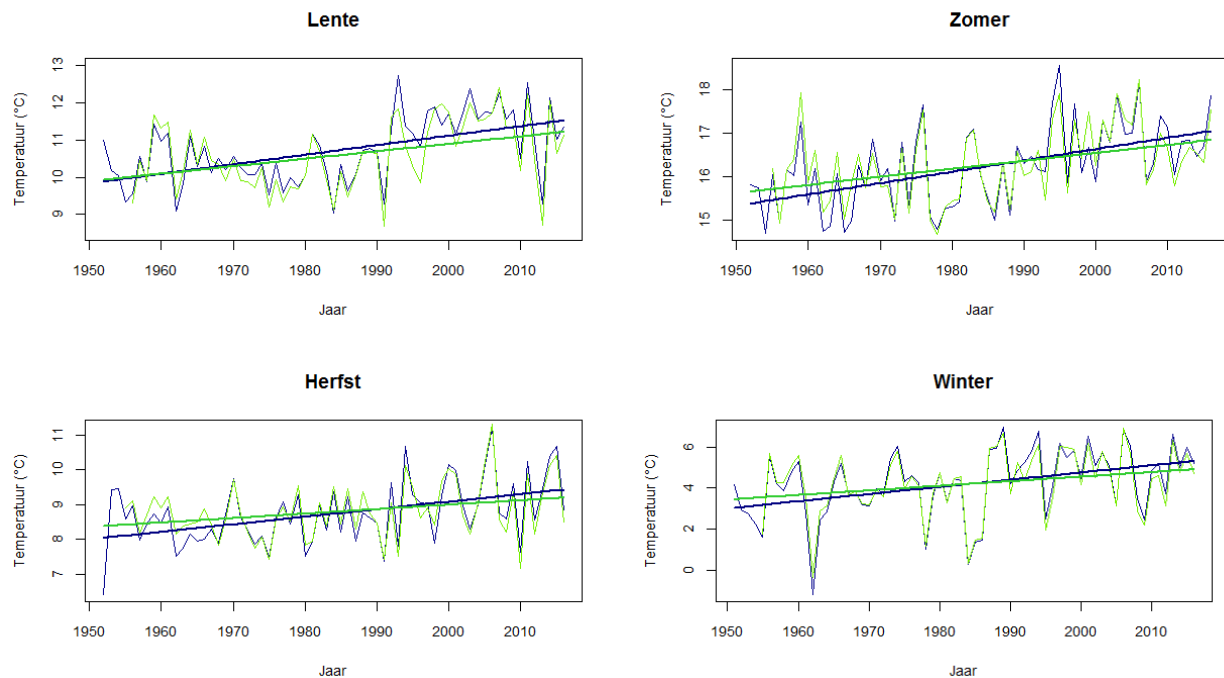
Ook in Middelkerke is in de vier seizoenen een significant stijgende trend waar te nemen. In de herfst en de winter ligt de  $p$ -waarde echter een stuk hoger en kan de nulhypothese, na Bonferroni-correctie, niet verworpen worden op het 5% significantieniveau: de  $p$ -waarden liggen niet onder 0.0125. In de lente en de zomer is de trend duidelijker en kan  $H_0$  verworpen worden op het 5% significantieniveau, zelfs op het 1% significantieniveau.

### 7.2.2 Niet-parametrische lineaire regressie

Naast een parametrische lineaire regressie voeren we ook een permutatietest uit. Hierbij worden de jaar- of seizoensgemiddelden 10 000 keer gepermuteed. De resultaten van deze niet-parametrische lineaire regressie staan weergegeven in Tabel 7.7. Na toepassen van de Bonferroni-correctie vinden we dat in zowel Koksijde als Middelkerke de temperatuur een significante stijgende, niet noodzakelijk lineaire, trend vertoont in de lente en de zomer op het 5% significantieniveau. Juist omdat de Bonferroni-correctie wordt toegepast, kan de nulhypothese dat de seizoensgemiddelde temperaturen toene- noch afnemen niet verworpen worden op het 5% significantieniveau in de herfst en de winter. Kijken we naar de jaargemiddelde temperaturen dan kan de nulhypothese zowel in Koksijde als in Middelkerke verworpen worden op dit significantieniveau.

Koksijde	$\hat{\beta}_0 \pm SE$	$\hat{\beta}_1 \pm SE$	$p$ -waarde
Lente	$9,885 \pm 0,1897$	$0,0255 \pm 0,00500$	$3.31 \cdot 10^{-6}$
Zomer	$15,376 \pm 0,202$	$0,0261 \pm 0,00531$	$6.62 \cdot 10^{-6}$
Herfst	$8,0442 \pm 0,205$	$0,0216 \pm 0,00539$	0.00016
Winter	$3,0756 \pm 0,378$	$0,0348 \pm 0,00982$	0.000736
Jaar	$9,131335 \pm 0,155937$	$0,026955 \pm 0,004171$	$1.82 \cdot 10^{-8}$
<b>Middelkerke</b>			
Lente	$9,990 \pm 0,223$	$0,0200 \pm 0.00621$	0.0021
Zomer	$15,712 \pm 0,206$	$0,0184 \pm 0,00569$	0.00194
Herfst	$8,418 \pm 0,2008$	$0,0128 \pm 0,00554$	0.0242
Winter	$3,544 \pm 0,391$	$0,0222 \pm 0.0108$	0.0435
Jaar	$9,356 \pm 0,189$	$0,0190 \pm 0,00490$	0.000271

Tabel 7.5: Resultaten van de lineaire regressie op de temperaturen in Koksijde (1952 –2016) en Middelkerke (1955 – 2016).



Figuur 7.6: Seizoensgemiddelde temperatuur per jaar met voor Koksijde (blauw) en Middelkerke (groen). Op elke figuur worden ook beide trendlijnen weergegeven.

Koksijde	Fractie $\hat{\beta}_{1\text{gegenererd}} < \hat{\beta}_{1\text{origineel}}$	$q = \frac{k}{s}$	$p = 2 \cdot \min(q, 1 - q)$
Lente	9988/10 000	0,9988	0.0024
Zomer	9985/10 000	0,9985	0.0030
Herfst	9879/10 000	,09879	0.0242
Winter	9786/10 000	0,9786	0.0428
Jaar	9999/10 000	0,9999	0.0002
<b>Middelkerke</b>			
Lente	9991/10 000	0,9991	0.0018
Zomer	9989/10 000	0,9989	0.0022
Herfst	9895/10 000	0,9895	0.0210
Winter	9786/10 000	0,9786	0.0448
Jaar	9999/10 000	0,9999	0.0002

Tabel 7.7: Bepalen van de significantie voor de permutatietest in Koksijde en Middelkerke

Bij het doorlopen van deze procedure voor de vier seizoenen, krijgen we histogrammen voor de schatters van  $\beta_1$  en deze staan voor Koksijde weergegeven in Figuur 7.8. De histogrammen bekomen bij Middelkerke zien er nagenoeg hetzelfde uit en worden niet meer getoond.

Met enkelvoudige lineaire regressie vinden we dat de oorspronkelijke  $\hat{\beta}_1$  in de lente 0,02564 is. Op Figuur 7.8 is duidelijk te zien dat deze waarde zich in de uiterste rechtse staart van de verdeling van gesimuleerde  $\hat{\beta}_1$  bevindt. Dit komt overeen met een lage  $p$ -waarde met name 0.0024, precies het resultaat dat we bij de Mann-Kendalltest ook hadden gevonden. Ook voor de zomer, herfst en winter bekomen we een gelijkaardig resultaat als bij de Mann-Kendalltest. Voor de zomer vinden we door enkelvoudige lineaire regressie dat  $\hat{\beta}_1$  een waarde van 0,02627 aanneemt, voor de herfst is dat 0,01904 en voor de winter 0,03293. Op Figuur 7.8 is te zien dat deze geobserveerde schatters zich steeds in de uiterste rechtse staart van de verdeling bevinden, wat wijst op een significant stijgende trend in de temperaturen.

### 7.2.3 Mann-Kendalltest en Sen's slope

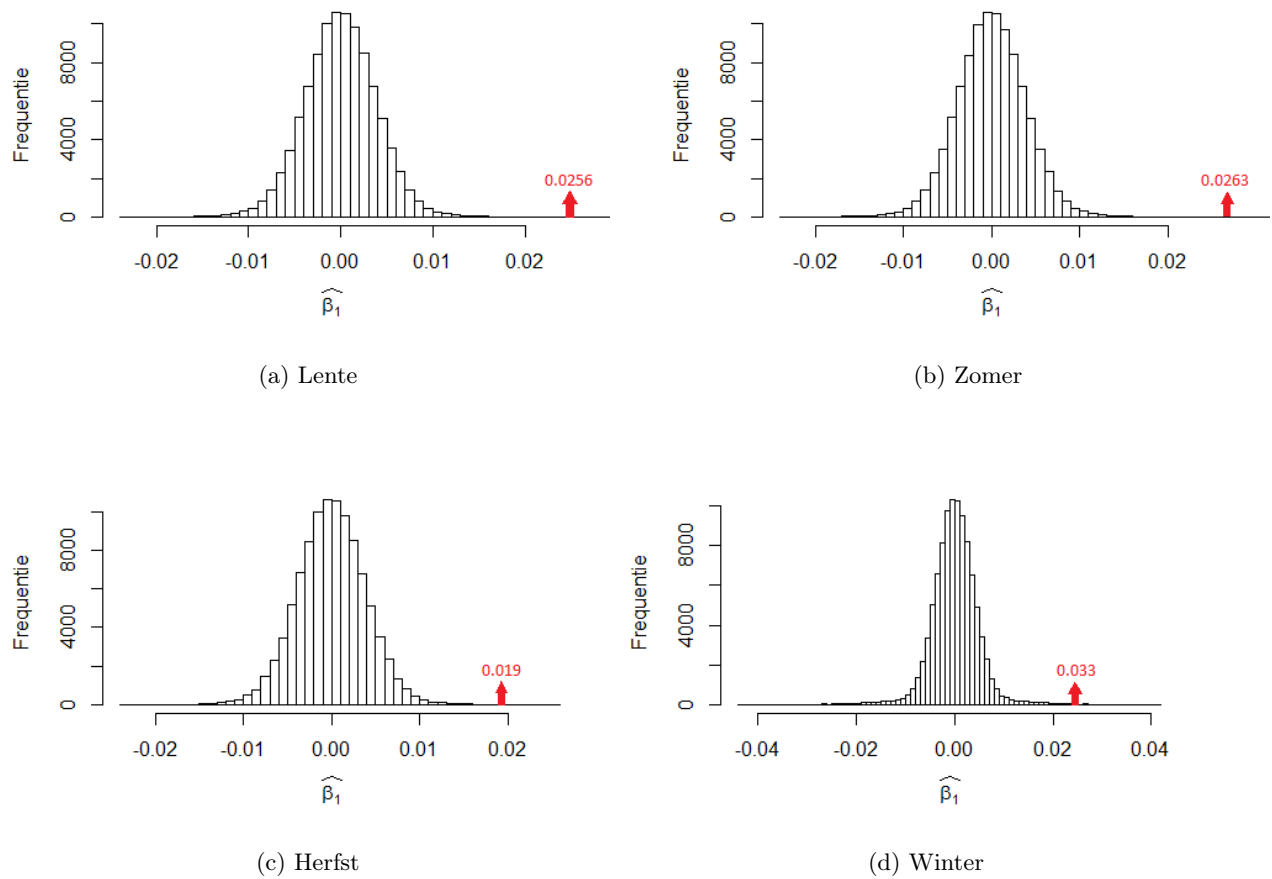
We voeren de Mann-Kendalltest uit voor de vier seizoenen. De waarden die bij deze test gevonden werden, staan weergegeven in de Tabel 7.9.

## 7.3 Conclusie

In Figuur 7.6 is te zien dat de variatie rond de trendlijn op het eerste zicht niet al te groot is. De MSE voor de seizoensgemiddelde temperaturen in Koksijde voor de lente, zomer, herfst en winter zijn respectievelijk 0,554; 0,626; 0,644 en 2,24. In de winter blijken de gemiddelde temperaturen dus meer variatie rond de trendlijn te vertonen. In de andere seizoenen is de MSE vrij klein, maar toch een stuk groter dan de MSE die we bij de gemiddelde windsnelheden in Koksijde kunnen vinden<sup>2</sup>.

<sup>2</sup>Ter vergelijking: voor de lente 0,275; voor de zomer 0,200; voor de herfst 0,35; voor de winter 0,418





Figuur 7.8: Van links naar rechts: histogrammen van de verdeling van schatters voor  $\beta_1$  voor de 4 seizoenen in Koksijde. Deze werden gegenereerd op basis van 10 000 permutaties. Op elk histogram wordt aangeduid waar de geobserveerde waarde voor  $\hat{\beta}_1$  terug te vinden is in het histogram.

Uit de analyses kunnen we besluiten dat er een significante stijging van de jaar- en seizoensgemiddelde temperatuur te zien is. Deze stijging is het duidelijkst aanwezig (kleinste  $p$ -waarden) in de lente en de zomer, zowel in Koksijde als Middelkerke.

<b>Koksijde</b>	$z$	$p$ -waarde	$m$	$b$	95% BI $m$
<b>Lente</b>	4,4	$9.80 \cdot 10^{-6}$	0,0283	9,75	[0,0171; 0,0378]
<b>Zomer</b>	4,4	$1.239 \cdot 10^{-5}$	0,0265	15,3475	[0,0151; 0,0377]
<b>Herfst</b>	3,2	0.00161	0,02	8,0409	[0,0085; 0,0329]
<b>Winter</b>	3,5	0.000450	0,0361	3,2253	[0,0168; 0,0537]
<b>Jaar</b>	5,2	$1.7887 \cdot 10^{-7}$	0,0282	9,2442	[0,0202; 0,0349]
<b>Middelkerke</b>					
<b>Lente</b>	3,1	0.00190	0,0224	9,956	[0,0089; 0,0363]
<b>Zomer</b>	3	0.0023891	0,0193	15,5657	[0,0082; 0,0305]
<b>Herfst</b>	1,8	0.066589	0,0116	8,3901	[-0,0007; 0,0234]
<b>Winter</b>	2,2	0.028767	0,0212	3,5702	[0,0029; 0,0436]
<b>Jaar</b>	3,5	0.000401	0,0215	9,453	[0,0102; 0,0332]

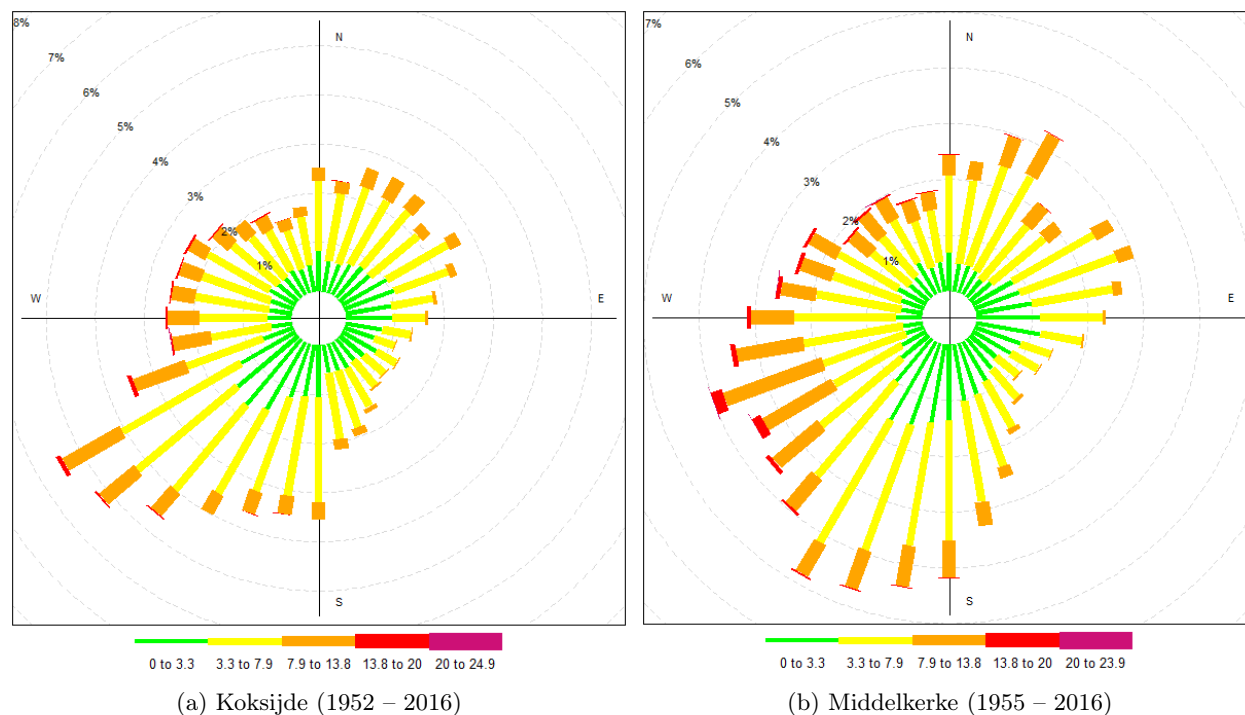
Tabel 7.9: Trendanalyse van de seizoensgemiddelde temperatuur per seizoen in Koksijde in de periode 1952 – 2016 en in Middelkerke in de periode 1955 – 2016. In de tabel staan volgende gegevens weergegeven: gestandaardiseerde Mann-Kendallteststatistiek  $z$ ,  $p$ -waarde van de teststatistiek, Sen's slope  $m$ , intercept  $b$  en een betrouwbaarheidsinterval voor  $m$ .

# Hoofdstuk 8

## De zeebries

De zeebries is een typisch kustweersfenomeen dat een belangrijke rol speelt in het microklimaat van de kust. Met behulp van de beschikbare data kan deze bestudeerd worden. Naast de windsnelheid is hiervoor ook de windrichting relevant. Onderzoek naar een evolutie in de zeebries is nog niet eerder gedaan in België.

### 8.1 Beschrijvende analyse van windrichtingen

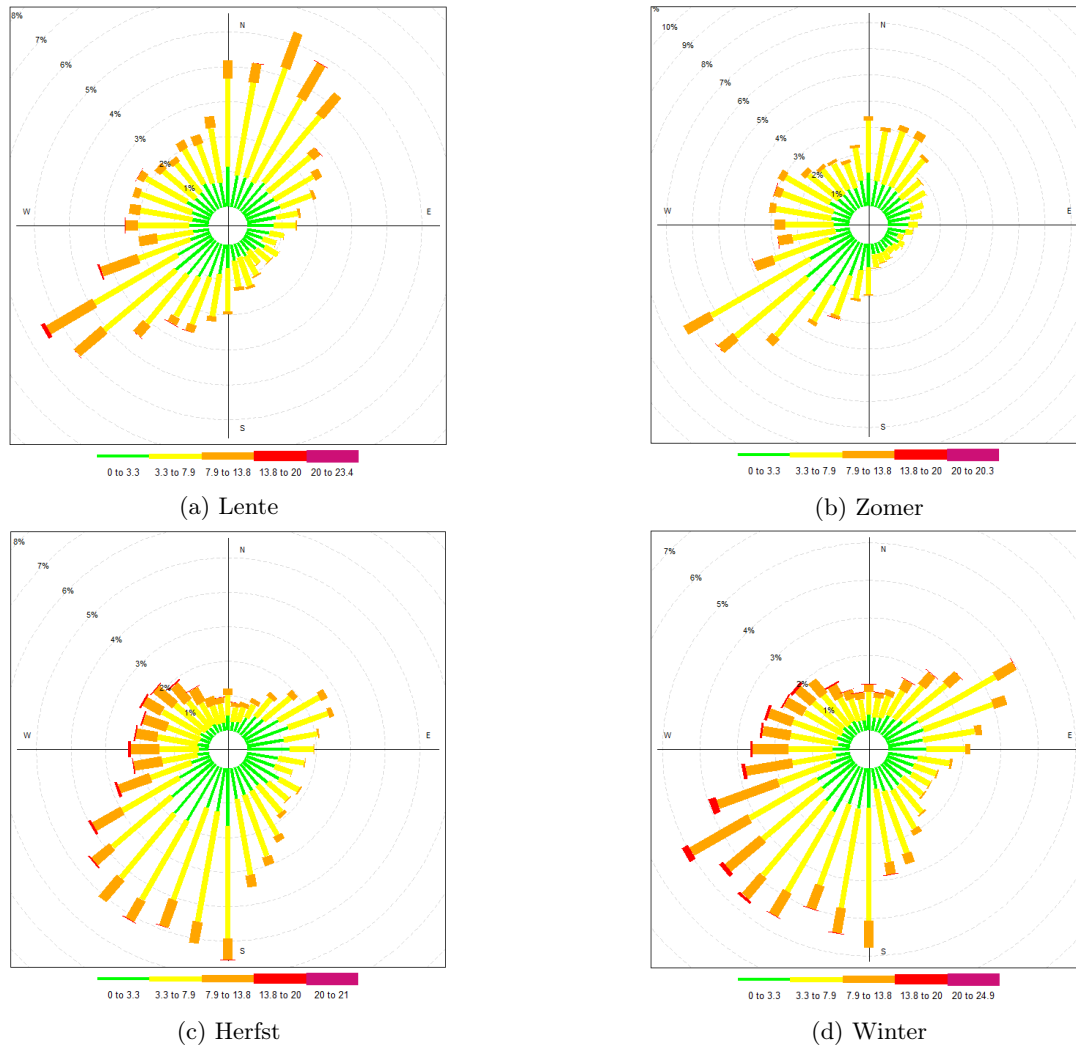


Figuur 8.1: Windrozen die alle windgegevens gebruiken.

Windrozen laten toe om een zicht te krijgen op de gezamenlijke verdeling van de windsnelheid en de windrichting in Koksijde en Middelkerke. Op de windros is af te lezen welk percentage van de wind uit welke windrichting komt. Verder geeft de kleur weer hoe sterk de windsnelheid was. De kleurschaal werd gekozen volgens de Beaufortschaal (zie Tabel 2). De wind in Koksijde gaat volgens deze schaal van ‘windstil’ tot ‘krach-

tige wind', waarbij een meer donkere kleur op sterkere wind duidt. De gebruikte gegevens zijn de zesuurlijkse uurgemiddelden zoals besproken in sectie 2.2.2.

De windrozen van Koksijde en Middelkerke worden vergeleken op twee vlakken: Figuur 8.1 toont windrozen die alle gegevens van 1952 (voor Koksijde)/1955 (voor Middelkerke) tot 2016 bevat. Figuren 8.2 en 8.3 tonen de windrozen per seizoen.

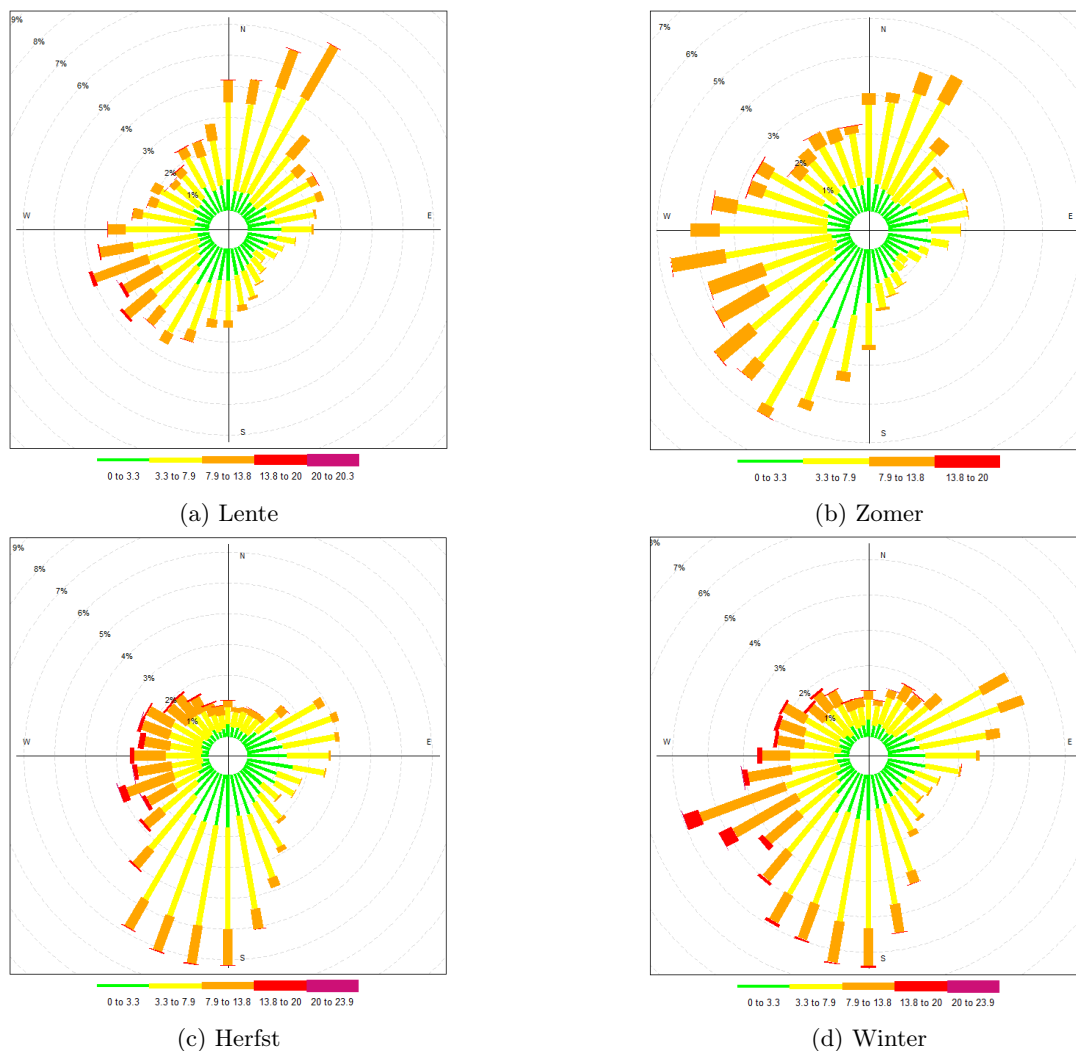


Figuur 8.2: Windrozen voor de vier seizoenen in Koksijde

In de windrozen voor zowel Koksijde als Middelkerke valt op dat de wind in de lente vaak uit het noordnoordoosten komt. Ook is te zien dat in elk seizoen de wind vaak waait uit zuidwestelijke richting en dat de winden uit deze richting vaak sterker zijn. Verder kan men opmerken dat er in de winter meer wind is dan in de lente en de zomer. Tot slot merken we op dat wanneer de wind van over zee komt, deze vaak sterk is.

## 8.2 Wat is de zeebries?

De zeebries is een landinwaartse wind die in de loop van de dag ontstaat bij een voldoende groot temperatuurverschil tussen een koudere zee en een warmer landoppervlak, en zich bijgevolg meestal voordoet in het voorjaar en een deel van de zomer [32]. Tijdens het eerste deel van de dag (voormiddag) waait er een zwakke,



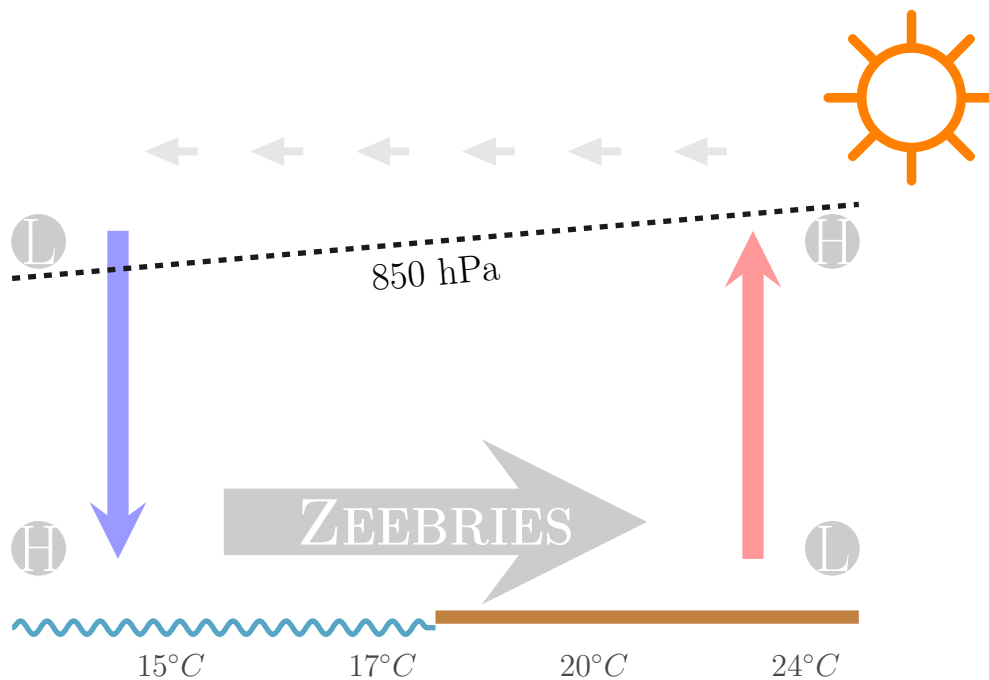
Figuur 8.3: Windrozen voor de vier seizoenen in Middelkerke

warme wind vanuit het binnenland. Later op de dag, wanneer het warmer wordt, kan die landwind verdwijnen en plaatsmaken voor de veel koelere zeebries.

Op warme, heldere dagen kan het land sterk worden opgewarmd door de zon. Het land absorbeert kortegolfstraling en de temperatuur boven land stijgt door conductie. Omdat land sneller opwarmt dan water, dat een veel grotere warmtecapaciteit heeft, ontstaat in de loop van de dag een groot temperatuurverschil tussen land en zee.

De mechanismen achter de zeebries staan geïllustreerd in Figuur 8.4. Door een grotere verwarming boven land, zet de luchtkolom boven land meer uit dan boven zee, waardoor een drukverschil tussen beide ontstaat. Dicht boven land ontstaat een lokaal lagedrukgebied ('warme lucht stijgt'). Bijgevolg ontstaat hogerop, laten we zeggen op het 850 hPa drukniveau, een hogedrukgebied boven land. Op diezelfde hoogte (in  $m$ ) vinden we boven zee een lagedrukgebied, net omdat de luchtkolom boven zee niet zoveel uitzet als boven land. Door de drukverschillen ontstaat op deze hoogte een luchtstroming van het land naar de zee. Wanneer deze getransporteerde lucht zich boven zee bevindt, zal deze beginnen dalen (koudere lucht). Door de dalende lucht ontstaat vlak boven zee een hogedrukgebied. Aan het aardoppervlak hebben we dan te maken met een hogedrukgebied boven zee en een lagedrukgebied boven land. Ook hier zal door het drukverschil opnieuw een stroming ontstaan (wind) die waait van een hogedrukgebied naar een lagedrukgebied (of dus van zee naar land). Dit is de zogenaamde zeebries [20]. Deze komt meer voor in de zomer en zelden in de winter, net omdat

er in de zomer zonnige, warme dagen zijn. De zeebries steekt meestal op in de loop van de namiddag, zo rond 14 uur lokale kloktijd.



Figuur 8.4: Schematische schets van het ontstaan van de zeebries. H staat voor hogedrukgebied, L voor lagedrukgebied. De zeebries waait van over zee naar het land.

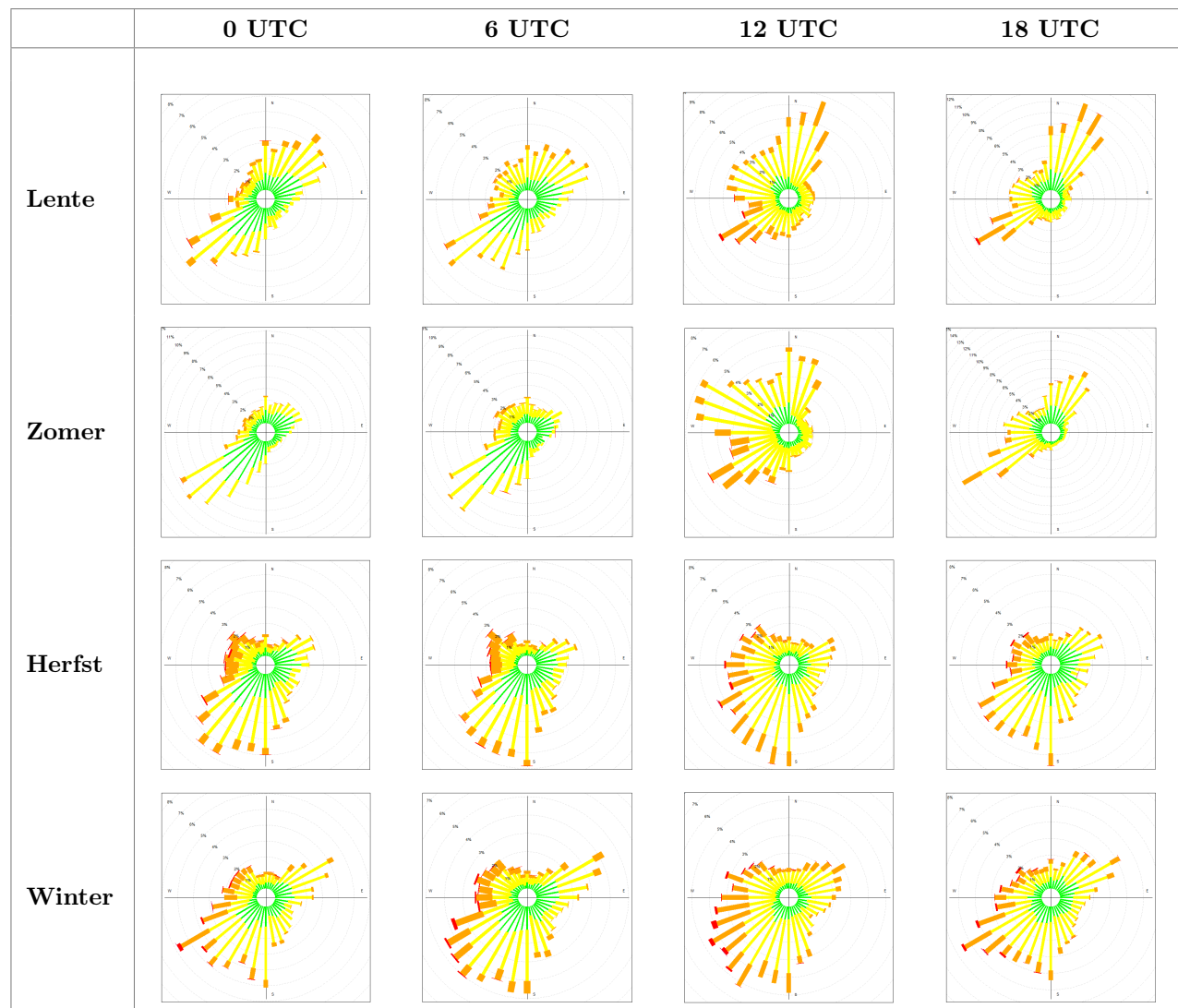
### 8.3 De relevantie van trends in de zeebries

In tijden van klimaatverandering stelt zich ook de vraag of de zeebries aan het veranderen is. Voor de zeebries zijn a priori verschillende scenario's mogelijk. Kijken we naar de analyse van de gemiddelde windsnelheid, dan zouden we verwachten dat de zeebries ook in dalende lijn gaat. Het is echter geweten dat de gemiddelde temperatuur in de lente en zomer stijgt [8]. Dit zou betekenen dat het land beter kan opwarmen, wat kan leiden tot een groter verschil tussen land- en zeetemperatuur en dus ook voor een sterkere zeebries. Er moet echter rekening mee gehouden worden dat door de temperatuurstijging niet alleen het land opwarmt, maar eveneens het zeewater. Dit zou een groter verschil tussen land- en zeetemperatuur dan weer kunnen tegengaan waardoor geen significante trend in de zeebries zichtbaar is.

### 8.4 Zeebries identificeren in de dataset: zeebriesdagen

In de volgende sectie gaan we na of het mogelijk is om in de dataset van Koksijde en Middelkerke, meer bepaald uit de windgegevens, de zeebries te ontwaren. Wanneer we voor elk seizoen op de beschikbare tijdstippen een windroos tekenen, zouden we mogelijks kunnen afleiden wanneer een zeebries meer aanwezig is. Aan de hand van karakteristieken van de zeebries verwachten we aan de Belgische kust vooral een zeebries in de lente en de zomer en dat in de loop van de namiddag, rond 12 UTC. We gaan dus op zoek naar een verhoogde frequentie in winden uit westelijke, noordwestelijke of noordelijke richting, voornamelijk in de lente en de zomer.

In Figuur 8.5<sup>1</sup> staan de verschillende windrozen getekend.



Tabel 8.5: Windrozen voor de verschillende seizoenen op 0, 6, 12 en 18 UTC in Koksijde. De windrozen werden gemaakt aan de hand van de gegevens van 1952 tot 2016.

Uit de windrozen is af te lezen dat de wind in de zomer om 12 UTC vaker uit westelijke tot noordelijke richting komt dan tijdens de andere seizoenen of op andere tijdstippen. Op die manier is de zeebries alvast te zien in de dataset.

De vraag stelt zich of het mogelijk is om de dagen waarop een zeebries ontstond uit de dataset te halen. Om dit te doen, moet een goede definitie van ‘zeebries’ voorhanden zijn. Het Oceanografisch Meteorologisch Station

<sup>1</sup>Bij het maken van deze windrozen in RStudio stootte ik op een eigenaardig probleem: voor de hele dataset en enkele subsets kon ik een windroos maken, voor andere subsets niet. Door met veel verschillende voorbeelden en gevallen te werken heb ik de volgende twee conclusies getrokken:

- Wanneer van een bepaalde subset geen windroos kan gemaakt worden, kunnen er ook geen windrozen van subsets van deze subset worden gemaakt.
- Het probleem kon (na lang zoeken) eenvoudig verholpen worden door namen van de kolommen ‘Wind speed (m/s)’ en ‘Wind direction (°)’ te vervangen door respectievelijk ‘ws’ en ‘wd’. Het blijft echter een raadsel waarom dit bij sommige subsets voor problemen zorgde en bij andere niet.

in Oostende hanteert de volgende definitie (persoonlijke communicatie met David Dehenauw per e-mail op 10 april 2017):

*De typische zeebries doet zich aan de Vlaamse kust voor als er 's ochtends een zwakke wind (2 Bft) staat uit veranderlijke richting. Afhankelijk van de duur van de opwarming zal de wind in de late voormiddag of rond de middag vrij snel draaien naar NNO en toenemen tot matig (3-4 Bft). Onder invloed van de corioliskracht ruimt de wind verder naar noordoosten, tot het geostrofisch evenwicht bereikt is en de zeebries evenwijdig waait met de kustlijn.*

Aan de hand van deze definitie en andere kennis over de zeebries die besproken werd in sectie 8.2, stellen we criteria op die moeten voldaan zijn om van een zeebries te kunnen spreken. Om deze zeebriesdagen te bepalen, wordt gebruik gemaakt van een uurlijkse dataset uit Middelkerke die verkregen werd via Belgocontrol. Middelkerke is het station bij uitstek om de zeebries te bestuderen: hier is een vrij lange meetreeks beschikbaar (1955 – 2016) en dit is bovendien het synoptische station dat het dichtst bij zee gelegen is. Het station van Koksijde zou minder ideaal zijn om de zeebriesdagen te bepalen: de luchtmachtbasis ligt bijna vier keer verder van de zee dan de luchthaven van Oostende, waardoor zwakke zeebriesen niet steeds tot daar doordringen.

De periode waarin gekeken wordt, loopt van 1 maart tot 1 september. Dit is een arbitraire, maar logische keuze: vóór 1 maart is er een lange, koude periode geweest waarin het zeewater behoorlijk is afgekoeld. Dit betekent ook dat wanneer zich in maart een eerste mooie lentedag voordoet, het land goed kan opwarmen, waardoor het temperatuurverschil tussen land en zee groot wordt. Dit laatste is een voorwaarde om van een echte zeebries te kunnen spreken [6]. De zee zal gedurende de lente- en zomermaanden verder opwarmen. Vanaf september zal de zee nog steeds zijn warmte vasthouden (door een grote warmtecapaciteit [1]), terwijl het land niet meer in die mate zal opwarmen dat het verschil in temperatuur tussen land en zee zo groot wordt dat een zeebries zal ontstaan. Het is echter wel mogelijk dat het in september nog vrij warm wordt, bijvoorbeeld wanneer zich een late nazomer voordoet. Maar het is te verantwoorden om met de maand september geen rekening meer te houden omdat de zeebriesdagen in deze maand schaars zijn.

Van een zeebries is geweten dat deze ontstaat op dagen waarbij 's ochtends een zwakke wind staat. Als definitie voor 's ochtends wordt gekozen voor 3 t.e.m. 7 UTC. Onder zwakke wind worden windsnelheden gerekend van 1,6 m/s tot 5,4 m/s. Bovendien wordt verondersteld dat deze winden waaien uit noordoostelijke tot zuidwestelijke richting ( $60^\circ$  tot  $230^\circ$ ). Een eerste subdataset die gemaakt wordt, bevat bijgevolg die dagen waarop de wind tussen 4 en 7 UTC uit noordwestelijke tot zuidoostelijke richting kwam en waarbij de windsnelheid lag tussen 1,6 m/s en 5,4 m/s.

Voor de middag moeten ook een paar veronderstellingen gemaakt worden. Ten eerste worden onder 'middag' de tijd tussen 11 en 15 UTC gerekend. Volgens de definitie van zeebries zal de wind toenemen. De windsnelheden die gebruikt worden zijn deze tussen 3,4 m/s en 10,7 m/s. Verder weten we dat de wind gedraaid is naar het noordoordoosten. Een tweede subdataset bevat dus de dagen waarop tussen 11 en 15 UTC de windsnelheden tussen 3,4 en 10,7 m/s lagen en waarbij de wind uit westelijke tot noordoostelijke richting ( $270^\circ$  tot  $30^\circ$ ) kwam. In Figuur 8.6 wordt schematisch voorgesteld met welke windsnelheden en windrichtingen rekening wordt gehouden op beide momenten van de dag.

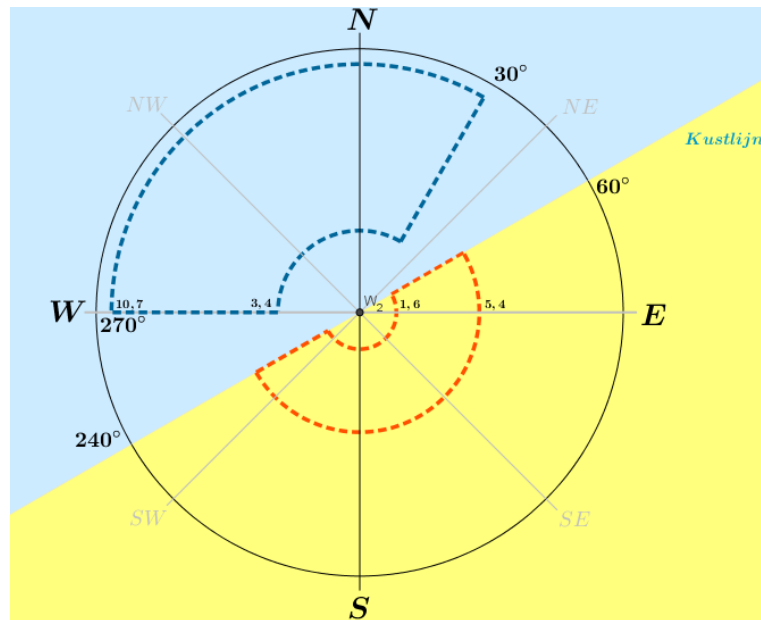
Een volgende stap is het bepalen van de doorsnede van de subdataset van 's ochtends en die van 's middags. We zijn namelijk op zoek naar die dagen waarop zowel de voorwaarden voor 's morgens als voor 's middags voldaan zijn. De dataset die we op die manier vinden bevat 2967 van de 23 725 dagen. Dit komt neer op gemiddeld 48 dagen per jaar waarop een zeebries ontstaat.

We kunnen deze gevonden dataset nog verder verfijnen. Omdat ook de neerslaghoeveelheden om de twaalf uur beschikbaar zijn, zullen de dagen waarop aan de criteria voor 's ochtends en 's avonds voldaan is, maar waarop er neerslag viel, niet in rekening gebracht worden. De kans is namelijk klein dat een zeebries zal ontstaan op een regenachtige dag<sup>2</sup>: de zon zal op dergelijke dagen niet voldoende kunnen schijnen om het land op te warmen zodat het temperatuurverschil tussen land en zee in die mate groot kan worden dat een zeebries kan ontstaan.

<sup>2</sup>Hierbij gaat het dan vooral over regen die in de voormiddag en de vroege namiddag valt, dus in de dataset gaat dit over de meting om 6 UTC.



De uiteindelijke dataset die, volgens de vooropgestelde criteria, de zeebriesdagen van 1955 tot 2016 bevat, bestaat uit dagen waarop het droog was en waarbij zowel 's morgens als 's middags aan de voorwaarden werd voldaan. Deze dataset heeft nog 2205 'zeebriesdagen', wat neerkomt op gemiddeld 36 dagen per jaar. Dit is een realistisch aantal en op deze dataset kunnen analyses uitgevoerd worden die iets vertellen over de zeebries.



Figuur 8.6: Het gebied tussen de stippellijnen geeft aan in welke richting de wind waait 's morgens en op de middag. Bovendien wordt ook aangeduid met welke windsnelheden op de verschillende momenten van de dag rekening wordt gehouden.

### Samenvatting van de criteria

- 1 maart  $\leq$  dag  $\leq$  1 september
- Windsnelheden op 4 UTC én 5 UTC én 6 UTC én 7 UTC  $\in [1,6; 5,4]$  m/s
- Windsnelheden op 12 UTC én 13 UTC én 14 UTC én 15 UTC  $\in [3,4; 10,7]$  m/s
- Windrichting 's ochtends  $\in [60^\circ, 240^\circ[$
- Windrichting 's middags  $\in [270^\circ, 30^\circ]$
- Neerslag = 0 mm om 6 UTC

**Opmerkingen bij de criteria** De voorwaarden die vooropgesteld werden om uit de dataset de zeebriesdagen te vinden, steunen op de definitie van de zeebries en op de karakteristieken ervan. Uiteraard blijven deze voorwaarden arbitrair gekozen en zijn er geen vaste criteria beschikbaar om de zeebriesdagen uit een dataset te halen. Bovendien bevat de dataset die hier gebruikt werd enkel informatie die verkregen werd nabij het grondoppervlak. Borne et al. [6] maken gebruik van een collectie van zes filters (cumulatieve criteria) om te bepalen of een dag een zeebriesdag is of niet. Hierbij wordt rekening gehouden met de verandering op grote hoogte, waar de geostrofische wind waait. Bovendien zou ook de kennis van de temperatuur boven het zeeoppervlak interessant zijn. Op die manier zou het temperatuurverschil tussen land en zee bekend zijn en zouden de dagen waarop dit verschil kleiner is dan  $3^\circ$  uit de dataset kunnen verwijderd worden. Verder zou

ook rekening gehouden kunnen worden met de bewolgingsgraad. Meer wolken in de lucht zorgen ervoor dat minder zonnestraling het land kan bereiken, wat leidt tot een koeling van het landoppervlak en wat op zijn beurt leidt tot een kleiner verschil in land- en zeetemperatuur.

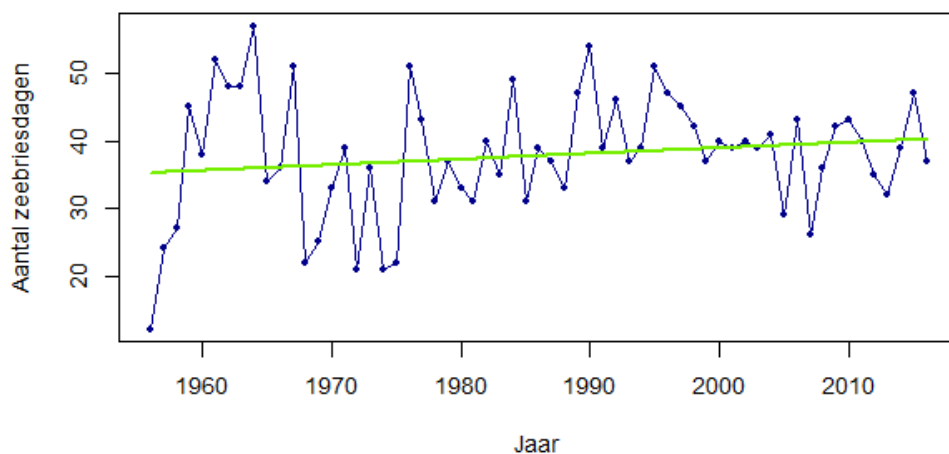
Er is dus geen zekerheid of in de 2205 dagen alle zeebriesdagen bevat zijn. Hoogstwaarschijnlijk is dit niet zo en allicht bevinden zich hierin ook nog dagen waarop er geen zeebries was. Het zou bijzonder nuttig zijn mocht bijgehouden worden op welke dagen een zeebries ontstond. Aan de hand van zo'n dataset zou dan kunnen afgetoetst worden welke de beste criteria zijn om, aan de hand van een uurlijkse dataset met enkel info nabij het grondoppervlak, de zeebriesdagen te filteren.

Tot slot merken we op dat deze analyse enkel gedaan werd voor Middelkerke: alleen voor dit meetstation was een uurlijkse dataset beschikbaar. Voor Koksijde is enkel de zesuurlijkse dataset beschikbaar. Aangezien voor de criteria verschillende uren gebruikt worden, was de dataset van Koksijde te beperkt om daarop een analyse uit te voeren.

## 8.5 Resultaat van de analyse

### 8.5.1 Aantal zeebriesdagen

De vraag die we hier beantwoord willen zien, is of het aantal zeebriesdagen toeneemt of afneemt. In Figuur 8.7 wordt weergegeven hoeveel zeebriesdagen per jaar worden geteld in Middelkerke<sup>3</sup>. De permutatietest levert



Figuur 8.7: Aantal zeebriesdagen van 1955 t.e.m. 2016

een  $p$ -waarde van 0.20. De Mann-Kendalltest, die door Romani et al. in [50] ook gebruikt wordt om een trend vast te stellen in het aantal dagen waarop de windsnelheid een bepaalde waarde overschrijdt, geeft als gestandaardiseerde teststatistiek  $z = 1,2$  en als  $p$ -waarde 0.24. De twee testen zijn het erover eens dat het aantal zeebriesdagen niet significant toe- of afneemt; op het 5% significantieniveau kan de nulhypothese dat het aantal zeebriesdagen stagneert niet verworpen worden. Op Figuur 8.7 werd eveneens de regressierechte geplotted waarvan de coëfficiënten geschat werden met Sen's slope ( $m = 0,0833$ ,  $b = 35,1667$ , BI  $[-0,06; 0,2222]$ ).

<sup>3</sup>We merken op dat het aantal zeebriesdagen in 1955 niet meer was dan zeven. Om in 1955 het aantal zeebriesdagen te bepalen, werden enkel de maanden mei, juni, juli en augustus gebruikt en niet maart en april (omdat in Middelkerke pas vanaf 1 mei metingen werden geregistreerd). In Figuur 8.7 werd dit datapunt reeds weggelaten omdat voor een verdere analyse geen rekening zal gehouden worden met dit gegeven. Het eerste datapunt is het aantal zeebriesdagen in het (volledige) jaar 1956, met name 12, een dieptepunt.

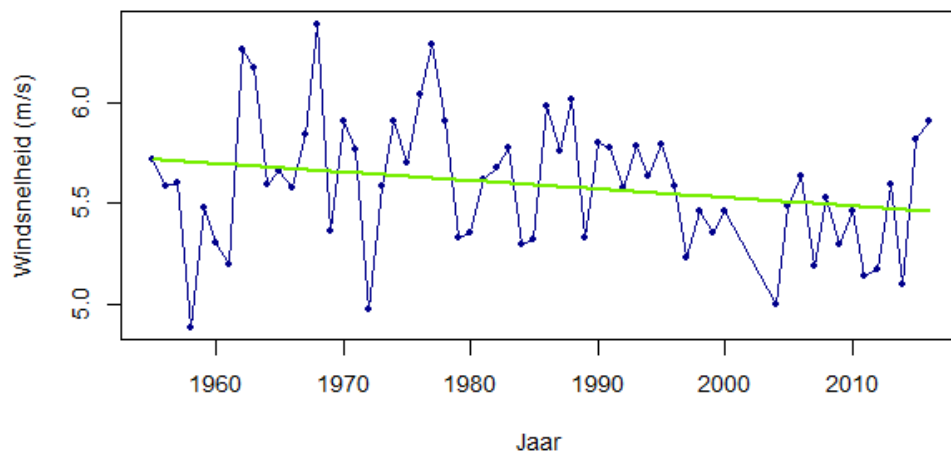
### 8.5.2 Gemiddelde sterkte van de zeebries

Van alle dagen per jaar, die beoordeeld werden als zeebriesdag, wordt de gemiddelde windsnelheid berekend van 10 UTC t.e.m. 15 UTC; dit zou dus in feite de gemiddelde zeebries per jaar voorstellen en hierbij vragen we ons af of de zeebries gemiddeld sterker wordt over de jaren heen. De Mann-Kendalltest voor trendanalyse geeft een gestandaardiseerde teststatistiek  $z = -2,1$  ( $p$ -waarde 0.038). Een permutatietest (100 000 permutaties) geeft voor  $\hat{\beta}_1$  een waarde van  $-0,00446$ . Uit de test blijkt dat het significantieniveau van deze test gelijk is aan 0.052. Een parametrische lineaire regressie schat dat  $\beta_1$  gelijk is aan  $-0,004458$  ( $p$ -waarde 0.0512) en dat  $\beta_0$  gelijk is 5,719574.

Volgens de Mann-Kendalltest kan de nulhypothese dat er geen trend is in de gemiddelde windsnelheden van de zeebries dus verworpen worden ten voordele van de alternatieve hypothese dat de gemiddelde zeebries een (stijgende of dalende, niet noodzakelijk lineaire) trend vertoont. Bij de permutatietest kan de nulhypothese niet verworpen worden op het 5% significantieniveau. Een histogram van de gegenereerde  $\hat{\beta}_1$  wordt getoond in Figuur 8.9. Ook bij de parametrische lineaire regressie kan  $H_0$  niet verworpen worden.

De twee niet-parametrische methoden, Mann-Kendall en de niet-parametrische lineaire regressie, geven een ander resultaat. We benadrukken hierbij nogmaals dat beide testen iets anders onderzoeken: bij de permutatietest wordt op zoek gegaan naar een lineaire trend daar waar bij de Mann-Kendalltest gezocht wordt naar een monotone, maar daarom niet noodzakelijk rechtlijnige trend.

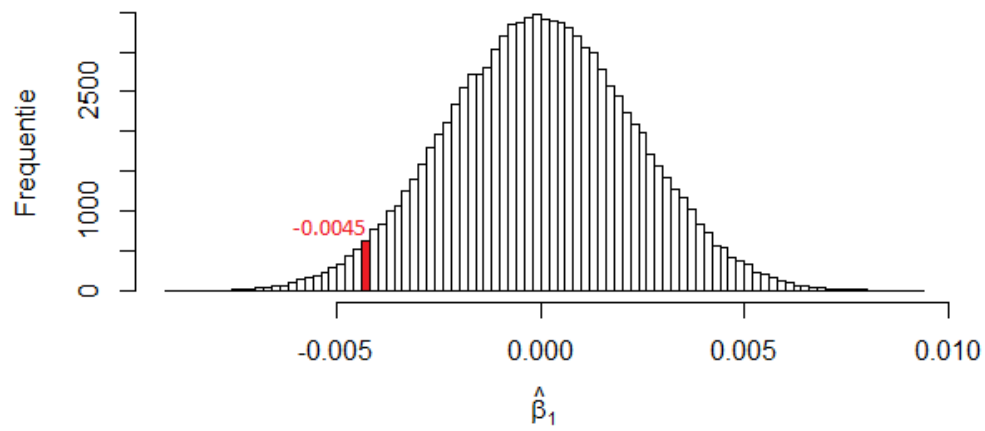
In Figuur 8.8 is de jaarlijkse gemiddelde sterkte van de zeebries te zien met daarboven de regressierechte waarbij  $\hat{\beta}_0$  en  $\hat{\beta}_1$  geschat werden aan de hand van Sen's slope ( $m = -0,0054$ ,  $b = 5,7641$ ).



Figuur 8.8: Jaarlijkse gemiddelde sterkte van de zeebries

## 8.6 Conclusie

Uit dit hoofdstuk kunnen we besluiten dat het aantal zeebriesdagen in de loop van de voorbije 62 jaar niet significant toe- of afgenomen is. Ook voor de gemiddelde sterkte van de zeebries is geen evidentie voor een significante trend. De Mann-Kendalltest gaf aan dat de nulhypothese dat de gemiddelde sterkte van de zeebries stagneert verworpen mag worden op het 5% significantieniveau, maar de andere testen spraken dit tegen. Dat de stijgende temperaturen een invloed zouden hebben op de gemiddelde sterkte van de zeebries, zoals vermoed



Figuur 8.9: Histogram van de schatters voor  $\beta_1$  onder de permutatie. In het rood staat aangegeven in welk deel van het histogram de geobserveerde waarde van  $\hat{\beta}_1$  zich bevindt.

zou kunnen worden, kon hier dus niet aangetoond worden. Hierbij is het echter opnieuw belangrijk om op te merken dat dit mogelijk ligt aan de beperkte criteria die werden vooropgesteld.

# Hoofdstuk 9

## Conclusies

### 9.1 Samenvatting van het onderzoek

In de literatuur is er veel te doen in verband met trends in het klimaat. Ook in deze studie wordt een steentje bijgedragen aan dergelijk onderzoek. Over heel de wereld wordt onderzoek gedaan naar evoluties van weervariabelen en naar statistische verdelingen, testen en methoden om deze te modelleren en te evalueren.

Het trekken van betrouwbare conclusies uit omvangrijke weerdatasets vraagt een aanzienlijke tijdsinvestering en is niet triviaal. Een gedegen clean-up om foutieve waarden te identificeren, vergt een groot deel van het werk. Een bescheiden aanzet in dit onderzoek staat in Hoofdstuk 3, waar de missing data in kaart gebracht werden en 12 windsnelheden na uitvoerige motivering uit de dataset verwijderd werden.

Er werd aangetoond dat de windsnelheden vrij goed gemodelleerd kunnen worden door een Weibullverdeling. Uit Hoofdstuk 4 blijkt het rechtvaardigbaar om aan te nemen dat de windsnelheden Weibull verdeeld zijn, ondanks mogelijke reserves (zie pagina 53). In de literatuur zijn talrijke methoden beschikbaar om de twee parameters van de Weibullverdeling te schatten. Uit onze analyse (sectie 4.3) blijkt dat de maximum-likelihoodmethode voor de windsnelheden in Koksijde en Middelkerke te verkiezen is boven de power-densitymethode of een empirische methode op basis van het eerste en tweede moment.

In de literatuur zijn een groot aantal methoden ontwikkeld om trends te identificeren en te kwantificeren, zoals parametrische lineaire regressie (de oerstandaard met talrijke variaties en veralgemeningen, zoals polynomiale regressie) en permutatietesten (een niet-parametrische regressiemethode), de Mann-Kendalltest (een verdelingsvrije trendtest waarbij de trendgrootte geschat wordt door Sen's slope) en LOWESS (voor een kwalitatief, visueel beeld). Zie hiervoor Hoofdstuk 5. Voor de analyse van klimatologische reeksen worden voornamelijk de Mann-Kendalltest en lineaire regressie gebruikt. Deze methoden werden aangewend om de trend in de windsnelheden in Koksijde en Middelkerke te onderzoeken.

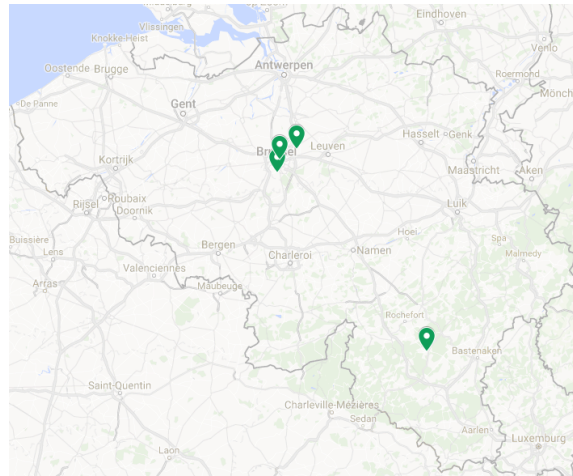
Uit de resultaten van de trendanalyse is duidelijk geworden dat de trend van jaar- en seizoensgemiddelde windsnelheden in Koksijde anders is dan in Middelkerke. Een significante daling kon in Middelkerke vastgesteld worden, die bovendien statistisch significant was in bijna alle seizoenen (in de winter net niet). In Koksijde is er geen enkele trend in de grootte van de jaar- en seizoensgemiddelde windsnelheden, wat mogelijks te verklaren valt door een dal in de gemiddelde windsnelheden in de jaren 1960. Bij het bestuderen van de seizoensgemiddelde windsnelheden was in Middelkerke in de lente, zomer en herfst een significant dalende trend vast te stellen ( $p < 0.05$ ) en in de winter was geen significante trend aanwezig ( $p < 0.10$ ). In Koksijde daarentegen was in geen van de seizoenen een significante trend te bespeuren. Een soortgelijke conclusie vinden we voor dagen met bijzondere wind: in Middelkerke daalt het aantal dagen met krachtige wind en stijgt het aantal dagen met zwakke wind, in Koksijde is dat niet zo.

Aan de kust de temperatuur het gemiddeld  $10^{\circ}\text{C}$ , maar de temperaturen stijgen significant in elk seizoen, waarbij deze stijging het duidelijkst is in Koksijde. De temperaturen stijgen met meer dan  $2^{\circ}\text{C}$  per eeuw. Alle voornoemde trendanalysemethoden zijn het hierover met elkaar eens.

De datasets van het KMI en Belgocontrol laten toe om de zeebries te identificeren in hun gegevens. Een tendens in het aantal dagen met een zeebries of in de grootte van die landinwaartse kustwind kunnen de gebruikte datasets (net) niet vaststellen, maar misschien is dat wel zo met sterkere datasets die ook over zeewatertemperatuur beschikken.

## 9.2 Vergelijking met binnen- en buitenland

In *Oog voor het klimaat* (2015), het klimaatrapport van het KMI, wordt de algemene evolutie van de weervariabelen bestudeerd. Hiervoor werden metingen gebruikt van vier klimatologische meetmasten in België, namelijk Sint-Joost-ten-Node, Ukkel, Zaventem en Saint-Hubert (zie Figuur 9.1).



Figuur 9.1: Meetstations in Ukkel, Zaventem, Sint-Joost-ten-Node en Saint-Hubert.

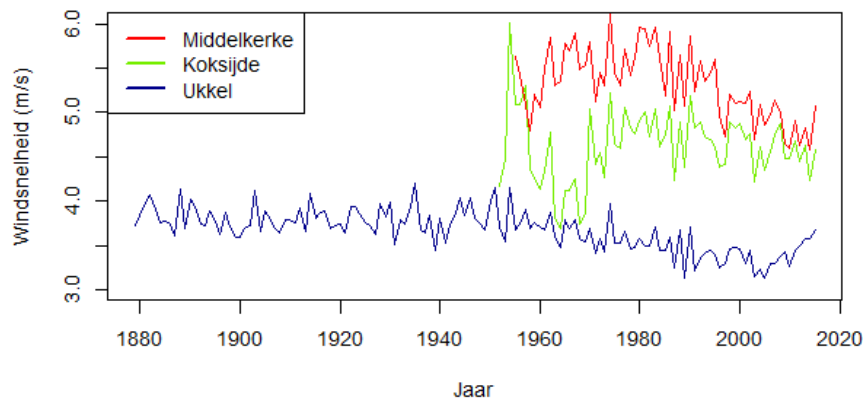
### 9.2.1 Windsnelheid

Bij de analyse van de gegevens van de windsnelheid kwam men tot de conclusie dat de gemiddelde windsnelheid in de streek rond Brussel vanaf het begin van de metingen in 1880 tot 1960 vrij stabiel gebleven is. Sinds 1960 is er een afname in gemiddelde windsnelheid met 10 tot 15% vastgesteld. Hierbij moet echter wel met de nodige voorzichtigheid omgesprongen worden met de gegevens: de omgeving van de meetstations is veranderd door de bouw van gebouwen en daar is dan ook rekening mee gehouden door homogenisering van de data met de test van Pettitt. De veranderde plantengroei in de buurt van het waarnemingspunt werd echter niet in rekening gebracht. De anemometrische reeksen in Zaventem en Saint-Hubert zijn een stuk korter en bevatten maar gegevens vanaf de jaren 1960.

Ook wanneer een studie werd gedaan van de seizoenen vanaf 1965, werd geconstateerd dat in de lente, zomer en herfst een significante ( $p$ -waarde  $< 0.05$ ) afname van gemiddelde windsnelheid te zien was sinds 1980. Dit was echter niet zo in de winter, het seizoen waarin de windsnelheden meestal het hoogst zijn [8].

De trend die de jaargemiddelde windsnelheden in Middelkerke volgen, is goed te vergelijken met de gegevens die in Ukkel verzameld werden. In *Oog voor het klimaat* [8] konden we vinden dat een analyse wees op een daling in gemiddelde windsnelheid.

Figuur 9.2 toont een visueel beeld van hoe de jaarlijkse gemiddelde windsnelheden in Koksijde en Middelkerke zich gedragen in de periode 1952 – 2016 in vergelijking met de jaargemiddelde windsnelheden in Ukkel. Op deze figuur is te zien dat de jaarlijkse gemiddelde windsnelheden in Koksijde en Middelkerke een stuk hoger liggen dan deze in Ukkel.



Figuur 9.2: Evolutie van de jaarlijkse gemiddelde windsnelheden in Ukkel, Koksijde en Middelkerke t.e.m. 2016

In Ukkel liggen de jaarlijkse gemiddelde windsnelheden in de periode 1950 – 2007 ongeveer in het interval  $[3,2; 4,2]$ . Voor Koksijde is dat in het interval  $[3,7; 6,0]$  en voor Middelkerke  $[4,6; 6,1]$ .

Bij vergelijking van de trend van de jaargemiddelde windsnelheden in Koksijde en Middelkerke met de jaargemiddelde windsnelheden in Ukkel, rijzen de volgende vaststellingen:

- Op het eerste, ‘ruwe’ zicht, zien we een duidelijk ‘dal’ bij de windsnelheden in Koksijde tussen 1960 en 1970. We zien echter geen duidelijke dalende trend na 1970, dewelke wel aanwezig is in Ukkel.
- Er is een vrij groot verschil in trend van jaarlijkse gemiddelde windsnelheid tussen Koksijde en Middelkerke. De trend die de windsnelheden in Middelkerke volgen, lijken beter overeen te komen met die uit Ukkel dan deze uit Koksijde. Zo kunnen we in Middelkerke vanaf 1980 ook een dalende trend vaststellen.

Naast de jaargemiddelde en seizoensgemiddelde windsnelheden, werd in Ukkel ook gekeken naar de evolutie van het jaarlijkse aantal stormdagen. Dit zijn dagen waarop de windstoten 70 km/h of 19 m/s overschrijden. Gegevens van windstoten zijn in Ukkel beschikbaar sinds 1940. Uit de trendanalyse kon men vaststellen dat er geen significante evolutie te zien was in het aantal jaarlijkse stormdagen. Er worden perioden opgemerkt met vrij veel stormdagen zoals rond 1947 en 1989, maar ook een laag aantal stormdagen wordt waargenomen in de jaren 1960. De laatste jaren wordt eveneens een relatief lage frequentie van stormdagen vastgesteld. In voorliggend werk kan geen analoge analyse worden gemaakt omdat de rukwinden pas vanaf 1994 bruikbaar zijn (zie Tabel 3.4).

### Dalende windsnelheden in Europa en daarbuiten

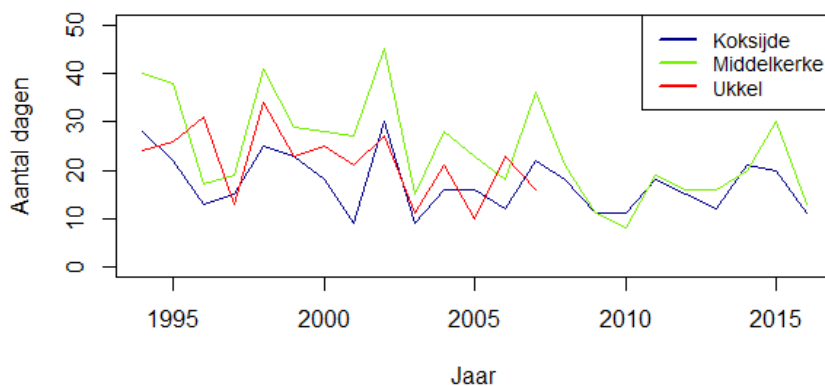
Niet alleen in België worden analyses uitgevoerd op anemometrische reeksen en andere klimatologische gegevens, maar ook in vele andere delen van Europa en de wereld [50, 17, 22, 54, 23, 43, 4]. Over het algemeen blijkt uit deze literatuur dat de windsnelheden nabij de grond afnemen. Een dalende trend in gemiddelde windsnelheid wordt niet alleen in België vastgesteld, maar over heel de wereld [4, 11]. In sommige artikelen worden oorzaken gezocht voor deze dalende trend. Hua Guo et al. stellen in [17] verschillende hypothesen voor waarom de gemiddelde windsnelheid zou kunnen dalen. Zij deden onder andere onderzoek naar de verandering in windsnelheid aan de zuidoostelijke kust van China. De auteurs menen dat de dalende windsnelheid in de *winter* veroorzaakt kan worden door de stijgende temperatuur. Door deze hogere temperaturen wordt de druk boven land kleiner en worden de temperatuur- en drukgradiënt tussen land en zee kleiner. Door een kleinere

drukgradiënt is er dus ook minder wind. Belangrijk om op te merken is dat deze redenering van toepassing is op de winter. In dit seizoen is het land vaker kouder dan de zee of oceaan omdat deze laatste een grotere warmtecapaciteit hebben. Wanneer de temperatuur boven land toeneemt, zal het temperatuurverschil tussen land en zee dus kleiner worden, waardoor ook de drukgradiënt zal afnemen.

Bichet et al. [5] gebruikten een globaal klimaatmodel om te onderzoeken welke impact veranderingen in ‘oppervlakteruwheid’ en veranderingen in o.a. zeeoppervlaktetemperatuur, uitstoot van aerosols en concentratie van broeikasgassen hebben op de windsnelheid op 10 m hoogte. De auteurs toonden aan dat verandering in oppervlakteruwheid gedeeltelijk bijdraagt aan een daling van de gemiddelde windsnelheid.

In [11] worden eveneens een aantal mogelijke oorzaken voor het dalen van de gemiddelde windsnelheid genoemd. De auteurs stellen dat zowel natuurlijke fenomenen als menselijke handelingen op lokale, regionale en globale schaal kunnen bijdragen tot een daling van de gemiddelde windsnelheid. Zij noemen als mogelijke oorzaken verstedelijking, veranderingen in luchttemperatuur en veranderingen in grootschalige atmosferische circulatiepatronen.

### 9.2.2 Rukwinden en stormdagen



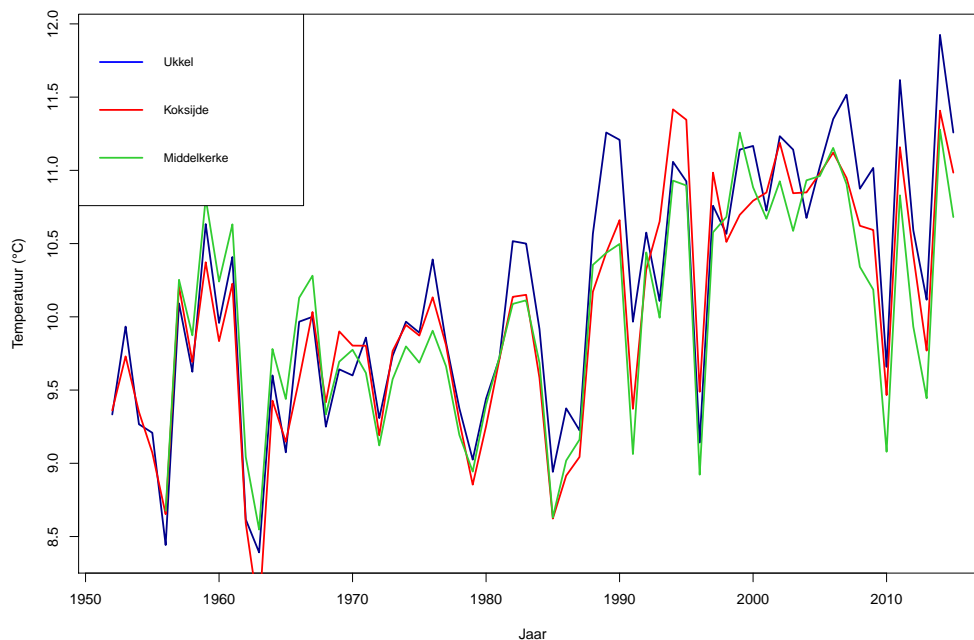
Figuur 9.3: Jaarlijks aantal dagen waarop de windstoten 70 km/h of 19,4 m/s overschrijden

Daar waar het in Ukkel te verantwoorden was om een analyse te doen op het jaarlijks aantal stormdagen, bleek dat zowel in Koksijde als in Middelkerke te weinig windstootgegevens beschikbaar waren om hierop een degelijke analyse te doen. Het is wel mogelijk om de evolutie van het aantal stormdagen in Koksijde en Middelkerke te vergelijken met de trend in Ukkel. Deze vergelijking kan echter enkel gebeuren gedurende de laatste 30 jaar, aangezien gegevens van rukwinden pas vanaf 1994 bruikbaar worden. Op Figuur 9.3 is te zien dat het jaarlijks aantal dagen waarop de windstoten 70 km/h overschrijden vrij gelijklopend is. Er zijn eveneens geen opmerkelijke verschillen met het aantal dagen in Ukkel. Het zou echter interessanter zijn, mochten ook windstootgegevens in Koksijde en Middelkerke beschikbaar zijn vanaf de jaren 1980. Op die manier zou kunnen gekeken worden of in die periode ook aan de kust een groot aantal stormdagen werd waargenomen.

### 9.2.3 Temperatuur

In Figuur 9.4 worden de jaarlijkse gemiddelde temperaturen voor Ukkel, Koksijde en Middelkerke vergeleken. Het valt onmiddellijk op dat de gemiddelde temperaturen in Ukkel, Koksijde en Middelkerke gelijkaardige patronen volgen. Vanaf ca. 1975 lijken de jaarlijkse gemiddelde temperaturen in Ukkel vaak iets hoger te liggen dan aan de kust. In *Oog voor het klimaat* wordt vastgesteld dat er een onweerlegbare stijging is in de





Figuur 9.4: Vergelijking van de jaarlijkse gemiddelde temperatuur in Ukkel, Koksijde en Middelkerke

jaar- en seizoensgemiddelde temperaturen vanaf de 19<sup>de</sup> eeuw [8]. Een gelijkaardig besluit vanaf de jaren '50 kon getrokken worden uit de analyse van de jaar- en seizoensgemiddelde temperaturen aan de kust. De meest significante stijging wordt vastgesteld in de lente en de zomer, zowel in Koksijde als Middelkerke.

De stijgende gemiddelde temperaturen zijn bijna over de hele wereld waarneembaar en uitgesproken [55].

#### 9.2.4 Zeebries

Wat betreft het aantal zeebriesdagen of de gemiddelde sterkte van de zeebries vonden we geen significante resultaten. Het is dus niet zo dat door een stijgende temperatuur ook de zeebries sterker wordt, zoals vermoed zou kunnen worden.

### 9.3 Mogelijkheden voor verder onderzoek

#### 9.3.1 Homogenisering

In een verder stadium zou het zinvol zijn om homogeniteitsbreuken in de data te detecteren en te corrigeren. Het gaat hier om gebeurtenissen die niet door weerkundige processen verklaard worden maar door veranderingen in de meetapparatuur of -omgeving.

Wanneer gedurende tientallen jaren metingen worden gedaan, is het haast onvermijdelijk dat meettoestellen vervangen en/of verplaatst moeten worden. Deze veranderingen en verplaatsingen kunnen de kwaliteit van de data beïnvloeden en leiden tot inhomogeniteit in de tijdsreeks. Zo kunnen bijvoorbeeld plotse sprongen of breekpunten voorkomen in de data.

Homogenisering van de tijdsreeksen is in dit geval belangrijk om op een betrouwbare en consistente manier de beschikbare data te kunnen analyseren. Om deze homogenisering te kunnen doorvoeren, is een goede

kennis van de geschiedenis van de meetapparatuur erg behulpzaam. Indien deze niet voorhanden is zijn er ook statistische tests die dergelijke breekpunten kunnen detecteren, zoals deze van Pettitt [8, 40] of de *Standard Normal Homogeneity Test* [9] en methoden om die te corrigeren.

De gemiddelde windsnelheden in Middelkerke vertonen een opmerkelijke sprong rond het jaar 1995 (zie Figuur 6.13). Om na te gaan of deze sprong te wijten is aan verandering van de omgeving of van het meettoestel of dat dit effectief gaat over sterk dalende gemiddelde windsnelheden, is kennis van verandering aan of verplaatsingen van het meetstation nuttig. In elk geval moet er iets aan het meettoestel veranderd zijn rond het jaar 1988 omdat het spectrum anders is dan tevoren (zie Figuur 3.2).

### 9.3.2 Koppelen van data voor analyse van de zeebries

Een meer gedetailleerde dataset met gegevens op grote hoogte ( $\pm 700$  hPa) en temperaturen van zeewater zouden kunnen bijdragen aan een verfijning en uitbreiding van de criteria voor zeebriesdagen. Deze nieuwe criteria zouden allicht een andere set van zeebriesdagen aangeven en dit zou mogelijks kunnen leiden tot een accuratere analyse. Het zou interessant zijn om de databanken van Koksijde en Middelkerke te koppelen met gegevens die verzameld werden door de boeien op zee. Deze bevatten onder andere de temperatuur van het zeewater. Een eventuele link tussen wind, zeebries of temperatuur en zeewater kunnen dan onderzocht worden.

### 9.3.3 Variantie-evolutie van de windsnelheid

Interessant om met deze data verder te onderzoeken is de variabiliteit van de windsnelheid. Wordt de windsnelheid meer of minder variabel met de jaren? Dergelijk onderzoek werd al eerder gedaan om een beeld te krijgen over het verschil in variabiliteit van de windsnelheden over de hele wereld [9].

### 9.3.4 Trendanalyse neerslag

In *Oog voor het klimaat* toonden analyses een significante stijging in de jaarlijkse hoeveelheid neerslag [8]. Het jaarlijks aantal dagen met neerslag daarentegen vertoonde geen opvallende evolutie. Het is interessant om ook deze analyse te gaan doen voor de kust.

### 9.3.5 Complexere modellen

Verbeteringen die zouden kunnen aangebracht worden om een meer gedetailleerde en correcte analyse te doen van jaar- en seizoensgemiddelde windsnelheden werden reeds besproken in Hoofdstuk 6. Gezien de beperkingen die ondervonden werden door het werken met jaar- en seizoensgemiddelden verdient het aanbeveling om de analyses te proberen met complexere modellen.

# Bijlage A

## English summary

The literature has a mouth full of climate trends. Around the world research is being conducted on weather variables and statistical distributions, tests and methods to model and evaluate them. Climate research and especially wind analysis are useful for climate policy, the energy sector, nature management and maybe even public health.

In this thesis, a number of climate variables on the Belgian coast, especially in Koksijde and Middelkerke, are studied in the period 1952 to 2016.

This thesis studies two longitudinal meteorological datasets, one from Koksijde (1952–2016, military airport) and one from Middelkerke (1955–2016, Ostend airport). Both measuring stations are near the Belgian coast, 19.58 km apart from each other. Recorded variables are time, hourly average wind speed and direction, six-hourly maximum wind gust speed, temperature (six-hour average, day minimum and day maximum), precipitation level and cloudiness (high, mid-and-low, total).

Drawing reliable conclusions from large (weather) datasets requires considerable time and is not trivial. A thorough clean-up to identify wrong values requires a great deal of the work. This studies comprises a modest start for clean-up, where the missing data were mapped and twelve wind speed values were removed from the data after extensive motivation.

We examine which probability distribution best describes wind speeds at the Belgian coast. It was shown that wind speeds can be modeled quite well by a Weibull distribution, despite possible concerns. In the literature, numerous methods are available to estimate the two parameters of the Weibull distribution. Our analysis shows that the maximum likelihood method for wind speeds in Koksijde and Middelkerke is preferable to the power density method or an empirical method based on the first and second moments.

In the literature a large number of methods have been developed to identify and quantify trends, such as parametric linear regression (the standard, with numerous variations and generalizations, such as polynomial regression) and permutation tests (a non-parametric regression method), the Mann-Kendall test (a distribution-free trend test – the quantification of the trend is estimated by Sen’s slope) and LOWESS (only for a qualitative visual impression). For the analysis of climatological time series, research mainly uses the Mann-Kendall test and linear regression. We used and compared these methods as well, together with a permutation test, to investigate the trend in coastal wind speeds.

Variables that were investigated are annual, seasonal and monthly average wind speeds and annual and seasonal average temperatures. The results of the trend analysis have revealed that the wind speed trends in Koksijde and Middelkerke are different. A significant decline in wind speeds could be found in Middelkerke, which was also statistically significant in spring, summer and autumn ( $p < 0.05/4$ ). Only in winter Middelkerke’s winds showed no significant trend ( $p = 0.07$ ). In Koksijde there is no trend in average wind speeds, annual nor seasonal. This might be explained by a fall in average wind speeds in the 1960s.

The trend found in the annual average wind speeds in Middelkerke corresponds more or less to those in Uccle, where a decrease in average wind speed has been established by 10 to 15% from 1960 to 2007. In spring, summer and fall a significant ( $p < 0.05$ ) decrease of average wind speed was observed in Belgian mainland since 1980. However, this was not the case in the winter. Furthermore, a declining trend in average wind speed is not only established in Belgium, but throughout the world.

The data sets of the Royal Meteorological Institute of Belgium and Belgocontrol allow to identify the existence of the seabreeze. A trend in the number of days with sea breezes or the magnitude of this inland coastal wind could not be established at the 5% significance level, but stronger data sets might be able to detect trends, for example when they contain sea water temperature.

Coastal temperatures average 10°C, but temperatures rise significantly over the last decades in each season. This rise is most apparent in Koksijde. The temperatures rise at a rate of on average 2°C per century. All used trend analysis methods gave the same result. Also in Uccle an undeniable increase in the annual and seasonal average temperatures since the 19<sup>th</sup> century was established. The trend we found here was to be expected. As the rising average temperatures are noticed and pronounced throughout the world, this trend was to be expected, but this does not make it less alarming.

# Bijlage B

## Tabellen met meetmomenten

### *Middelkerke*

- 1979-8-22:0
- 2004-1-21:0
- 2004-1-2:0
- 2004-11-13:18
- 2004-2-24:0
- 2004-2-27:6
- 2004-3-24:
- 2004-3-31:12
- 2004-3-3:0
- 2004-3-3:6
- 2004-4-16:12
- 2004-5-13:0
- 2004-5-19:0
- 2004-7-11:12
- 2004-7-15:0
- 2004-7-20:0
- 2004-7-22:0
- 2004-7-26:6
- 2004-8-1:12
- 2006-10-31:18
- 2006-9-27:6
- 2007-10-18:0

Tabel B.1: Momenten waarop de volledige meting ontbreekt in de weerdataset van Middelkerke

### *Koksijde*

- 1977-3-15:0
- 1977-3-15:12
- 2004-1-2:6
- 2004-11-15:6
- 2004-11-8:12
- 2004-12-4:12
- 2004-12-8:12
- 2004-2-25:0
- 2004-2-25:18
- 2004-2-25:6
- 2004-2-26:0
- 2004-3-29:12
- 2004-4-17:0
- 2004-4-17:18
- 2004-5-29:6
- 2004-6-23:12
- 2004-6-9:12
- 2004-8-14:6
- 2004-8-25:0
- 2004-9-29:12
- 2013-8-3:12
- 2013-8-26:12
- 2014-11-15:18

Tabel B.2: Momenten waarop de volledige meting ontbreekt in de weerdataset van Koksijde

Windsnelheid	Frequentie	Windsnelheid	Frequentie	Windsnelheid	Frequentie
3	9069	7.7	305	15.8	20
4	7874	11.7	263	13.9	16
2	7402	12.7	219	14.4	16
5	7188	12	217	16	14
1	4584	12.2	214	16.8	14
2.5	4273	8.7	206	17	13
3.5	3913	3.1	197	12.9	9
0	3579	2.1	179	15.4	9
6.1	3451	13.2	179	18.8	8
1.5	3159	9.2	167	14.9	7
4.5	3158	2.6	148	17.8	7
5.6	2668	13	138	13.4	6
6	2626	6.2	129	18.5	6
6.6	2583	9.7	127	19.3	6
7.1	2403	5.7	123	18.3	5
7.6	2130	10.2	119	13.8	4
7	1996	6.7	99	16.4	4
8.1	1620	11.3	99	18	4
8	1377	10.8	97	20.3	4
0.5	1263	14.2	90	16.9	3
8.6	1148	13.7	86	15.3	2
9.1	1022	15.2	71	15.9	2
9	931	11.8	58	19	2
5.1	874	14	48	19.8	2
3.6	801	12.3	47	20	2
4.1	788	14.7	42	21	2
4.6	733	9.8	34	1.02	1
10.1	721	15	33	6.63	1
9.6	626	9.3	32	11.22	1
10	601	12.8	32	17.4	1
11.2	426	13.3	32	17.5	1
7.2	384	16.3	30	19.5	1
11	354	10.3	25	23.4	1
10.7	351	17.3	21	24.9	1
8.2	310				

Tabel B.3: Frequentietabel van alle gemeten windsnelheden in Koksijde gesorteerd op frequentie

NR	Jaar	Maand	Dag	Uur	NR	Jaar	Maand	Dag	Uur
1	1953	3	02	06	35	1969	8	22	00
2	1953	3	02	18	36	1969	9	16	18
3	1953	3	03	00	37	1970	4	20	00
4	1953	3	06	12		1970	5	14	06
5	1953	3	07	00	38	1970	5	27	06
6	1953	3	07	06	39	1970	10	19	18
7	1953	3	07	18	40	1971	5	09	18
8	1953	3	09	06	41	1971	6	07	12
9	1953	3	09	18	42	1972	3	23	12
10	1953	3	10	00	43	1972	6	12	06
11	1953	3	10	12	44	1972	6	24	00
12	1953	3	11	00	45	1973	4	10	12
13	1953	3	11	06		1973	9	27	18
14	1953	3	11	12	46	1973	10	06	12
15	1953	3	11	18	47	1974	12	30	12
16	1953	3	14	12	48	1975	2	21	18
17	1953	4	04	00	49	1975	8	17	06
18	1953	5	06	18	50	1975	8	17	18
19	1953	10	01	00	51	1976	2	04	06
20	1959	3	28	00	52	1976	7	31	18
21	1959	7	07	00		1976	12	17	18
	1960	1	18	18	53	1977	5	02	00
22	1961	4	01	00	54	1979	10	18	12
23	1962	7	01	00	55	1985	5	15	00
24	1964	9	01	00	56	1985	5	16	00
25	1965	12	20	00	57	1985	5	17	00
26	1966	11	02	12	58	1987	9	17	00
27	1966	11	27	06	59	1987	9	17	06
28	1968	10	28	06	60	1995	9	08	12
29	1968	11	10	18	61	2005	6	29	00
30	1969	1	07	18	62	2005	6	29	06
31	1969	6	08	00	63	2005	6	30	00
32	1969	6	26	12	64	2005	12	04	00
33	1969	7	13	18	65	2005	12	04	06
34	1969	8	10	06	66	2005	12	05	12
	1969	8	5	00					

Tabel B.4: Meetmomenten waarop de windsnelheid ontbreekt, maar waarop wel andere variabelen werden gemeten. In het rood staan de momenten weergegeven waarop foute outliers werden gevonden.

NR	YEAR	MONTH	DAY	HOUR	Total nebulosity (octa)	Temperature	Precipitation (mm)
1	1953	3	02	06	NA	NA	NA
2	1953	3	02	18	NA	NA	NA
3	1953	3	03	00	NA	NA	NA
4	1953	3	06	12	NA	NA	NA
5	1953	3	07	00	NA	NA	NA
6	1953	3	07	06	NA	NA	NA
7	1953	3	07	18	NA	NA	NA
8	1953	3	09	06	NA	NA	NA
9	1953	3	09	18	NA	NA	NA
10	1953	3	10	00	NA	NA	NA
11	1953	3	10	12	NA	NA	NA
12	1953	3	11	00	NA	NA	NA
13	1953	3	11	06	NA	NA	NA
14	1953	3	11	12	NA	NA	NA
15	1953	3	11	18	NA	NA	NA
16	1953	3	14	12	NA	NA	NA
17	1953	4	04	00	NA	NA	NA
18	1953	5	06	18	7	9	0
19	1953	10	01	00	NA	NA	NA
20	1959	3	28	00	NA	NA	NA
21	1959	7	07	00	2	11	NA
22	1961	4	01	00	NA	NA	NA
23	1962	7	01	00	7	11	NA
24	1964	9	01	00	NA	NA	NA
25	1965	12	20	00	1	5	NA
26	1966	11	02	12	0	2	NA
27	1966	11	27	06	8	8	0
28	1968	10	28	06	7	10	0
29	1968	11	10	18	8	4	0.2
30	1969	1	07	18	8	5	0
31	1969	6	08	00	0	11	NA
32	1969	6	26	12	3	20	NA
33	1969	7	13	18	1	21	0
34	1969	8	10	06	2	16	0
35	1969	8	22	00	3	12	NA
36	1969	9	16	18	7	19	0
37	1970	4	20	00	3	7	NA
38	1970	5	27	06	2	12	0
39	1970	10	19	18	7	12	5
40	1971	5	09	18	5	13	4
41	1971	6	07	12	6	17	NA
42	1972	3	23	12	1	10	NA
43	1972	6	12	06	1	7	0
44	1972	6	24	00	2	11	NA
45	1973	4	10	12	5	6	NA
46	1973	10	06	12	7	17	NA
47	1974	12	30	12	6	8	NA
48	1975	2	21	18	6	5	0
49	1975	8	17	06	NA	NA	NA
50	1975	8	17	18	NA	NA	NA
51	1976	2	04	06	7	-2	0
52	1976	7	31	18	8	13	4
53	1977	5	02	00	0	NA	NA
54	1979	10	18	12	3	14	NA
55	1985	5	15	00	7	9.9	NA
56	1985	5	16	00	3	8.7	NA
57	1985	5	17	00	7	11.7	NA
58	1987	9	17	00	4	13.8	NA
59	1987	9	17	06	7	15.5	0
60	1995	9	08	12	5	18.7	NA
61	2005	6	29	00	7	16.8	NA
62	2005	6	29	06	5	18.1	12
63	2005	6	30	00	2	15.4	NA
64	2005	12	04	00	7	7.5	NA
65	2005	12	04	06	5	6.6	12
66	2005	12	05	12	7	8	NA

Tabel B.5: Weersomstandigheden (totale bewolking, temperatuur, neerslag) van meetmomenten waarop de wind ontbreekt in Koksijde



# Bijlage C

## R-code

### C.1 Inleiding

### C.2 Gegevens over wind aan de Belgische kust

#### C.2.3 Hulpvariabelen

##### Moment

```
# Genereren van de hulpvariabele moment
DATASET$moment <- paste(DATASET$YEAR,"-",DATASET$MONTH,"",DATASET$DAY,"-",DATASET$HOUR)
# Men kan er ook voor kiezen om bijvoorbeeld enkel jaar, maand en dag mee te nemen
```

##### Seizoen en seizoenjaar

```
# Genereren van de hulpvariabele SEIZOEN (hier met Koksijde2017, Middelkerke2017 analoog)
Koksijde2017$SEIZOEN <- "0"
Koksijde2017[(Koksijde2017$MONTH==3 & Koksijde2017$DAY >= 21) | Koksijde2017$MONTH==4 |
  Koksijde2017$MONTH==5 | (Koksijde2017$MONTH==6 & Koksijde2017$DAY < 21),"SEIZOEN"] <- "
  L"
Koksijde2017[(Koksijde2017$MONTH==6 & Koksijde2017$DAY >= 21) | Koksijde2017$MONTH==7 |
  Koksijde2017$MONTH==8 | (Koksijde2017$MONTH==9 & Koksijde2017$DAY < 21),"SEIZOEN"] <- "
  Z"
Koksijde2017[(Koksijde2017$MONTH==9 & Koksijde2017$DAY >= 21) | Koksijde2017$MONTH==10 |
  Koksijde2017$MONTH==11 | (Koksijde2017$MONTH==12 & Koksijde2017$DAY < 21),"SEIZOEN"] <-
  "H"
Koksijde2017[(Koksijde2017$MONTH==12 & Koksijde2017$DAY >= 21) | [Koksijde2017$MONTH==1 |
  Koksijde2017$MONTH==2 | (Koksijde2017$MONTH==3 & Koksijde2017$DAY < 21),"SEIZOEN"] <- "
  W"

# Genereren van de hulpvariabelen SEIZOENJAAR (hier met Koksijde2017, analoog voor
  Middelkerke2017)
Koksijde2017$SEIZOENJAAR <- "0"
Koksijde2017[Koksijde2017$MONTH > 3 | (Koksijde2017$MONTH == 3 & Koksijde2017$DAY >= 21),"
  SEIZOENJAAR"] <- "N"
```

```
Koksijde2017$SEIZOENJAAR <- ifelse(Koksijde2017$SEIZOENJAAR=="N", paste(
  Koksijde2017$SEIZOEN, Koksijde2017$YEAR),paste(Koksijde2017$SEIZOEN, Koksijde2017$YEAR
  -1)) # Winters van jaar X moeten behoren tot jaar X - 1
```

### Windrichting (hoek)

```
# Genereren van hulpvariabele windrichting
DATASET$richting <- -DATASET$wd + 90
```

### Verschil Koksijde - Middelkerke

```
# Genereren hulpvariabele windverschil
y<-merge(koksijde,Middelkerke,by="moment") # Momenten bepalen waarbij zowel in Koksijde als
  Middelkerke gegevens zijn, doorsnede van beide datasets bepalen op basis van de
  hulpvariabele 'moment'
y$verschiltemp <- y$'Temperature (C).x' - y$'Temperature (C).y' # Het verschil maken

# Variabele Tempschil voor Koksijde en Middelkerke
DATASET$tempschil <- c(NA,(DATASET[2:(nrow(DATASET)-1),"Temperature (C)"]-DATASET[1:(nrow(
  DATASET)-2),"Temperature (C)"])^2+(DATASET[2:(nrow(DATASET)-1),"Temperature (C)"]-
  DATASET[3:(nrow(DATASET)),"Temperature (C)"])^2, NA)
```

### Variabiliteit tussen opeenvolgende metingen

```
# Genereren hulpvariabele windschil en tempschil
DATASET$windschilSTATION <- c(NA,(DATASET[2:(nrow(DATASET)-1),"ws"]-DATASET[1:(nrow(DATASET
  )-2),"ws"])^2+(DATASET[2:(nrow(DATASET)-1),"ws"]-DATASET[3:(nrow(DATASET)),"ws"])^2, NA
  )

y<-merge(Koksijde2017,Middelkerke2017,by="moment") # Deze mergen zorgt voor een dataframe y
  met temp.x de temperatuur uit de eerste mergecomponent, hier Koksijde2017, temp.y is
  de temperatuur in Middelkerke2017. De naam van het dataframe y en temp.y hebben niets
  met elkaar te maken.
y$tempschilkok <- c(NA,(y[2:(nrow(y)-1),"temp.x"]-y[1:(nrow(y)-2),"temp.x"])^2+(y[2:(nrow(y
  )-1),"temp.x"]-y[3:(nrow(y)),"temp.x"])^2, NA)
y$tempschilmid <- c(NA,(y[2:(nrow(y)-1),"temp.y"]-y[1:(nrow(y)-2),"temp.y"])^2+(y[2:(nrow(y
  )-1),"temp.y"]-y[3:(nrow(y)),"temp.y"])^2, NA)
```

### Orthogonale componenten van de windsnelheid

```
# Genereren van de hulpvariabele u en v en omzetten naar graden (voor Koksijde, idem voor
  Middelkerke)
Koksijde2017$u <- Koksijde2017$ws*cos(Koksijde2017$richting*pi/180)
Koksijde2017$v <- Koksijde2017$ws*sin(Koksijde2017$richting*pi/180)
```

## C.3 Clean-up van de dataset

### C.3.1 Ontbrekende meetmomenten

```
# Uurverschil berekenen (om na te gaan welke rijen in de dataset ontbreken, analoog voor
  Middelkerke)
Koksijde2017$uurverschil <- c(NA,(Koksijde2017[2:(nrow(Koksijde2017)-1),"HOOR"]-Koksijde2017
  [1:(nrow(Koksijde2017)-2),"HOOR"])%%24,NA)
```

### C.3.2 Windsnelheden

#### Spectrum van gemeten waarden

```
# Genereren staafdiagrammen
counts <- table(DATASET$ws)
barplot(counts / sum(counts), col = "lightseagreen")
```

#### Ontbrekende waarden

```
# Aantal NA's bepalen
sum(is.na(DATASET$VARIABLE))

# Spectrumplot van gemeten windsnelheden per jaar
plot(DATASET$YEAR, DATASET$ws, ylim = c(0,10), col = "darkblue", xlab = "Jaar", ylab = "
  Windsnelheden (m/s)", pch = "-")
```

#### Foutieve outliers

```
# Histogrammen windsnelheden
hist(DATASET$ws, breaks = 40, col = "lightseagreen", xlab = "Windsnelheden (m/s)", ylab = "
  Frequentie", main = "")

# Scatterplot van gemeten windsnelheden per seizoen
plot(SEIZOEN$YEAR, SEIZOEN$ws, xlab = "Jaar", ylab = "Windsnelheid (m/s)", main = "
  Windsnelheden (SEIZOEN)", col = "darkblue", pch = 20) # Hier staat SEIZOEN voor
  bijvoorbeeld LenteKoksijde

# Scatterplot met kleurenschaal (Koksijde)
scatter2D(y$verschilwind, y$windschil, colvar = y$ws.x, pch = 20, xlab = expression(W[
  Koksijde] - W[Middelkerke]), ylab = expression((w[i] - w[i-1])^2 + (w[i] - w[i+1])^2))

# Correlatieplot
plot(y$ws.x, y$ws.y, col = "darkblue", pch = 20, xlab = "Windsnelheid Koksijde (m/s)", ylab
  = "Windsnelheid Middelkerke (m/s)", ylim = c(0,70))
# Alleen momenten waarbij zowel voor Koksijde als Middelkerke windsnelheden beschikbaar
  zijn, worden geplotted. Deze zitten in de doorsnede (gemerged volgens moment)
curve(x+0, col = "red", add = TRUE, lty = "dotted")
```

### C.3.3 Temperatuur

#### Foutieve outliers

```
# Histogrammen temperatuur
hist(DATASET$temp, breaks = 50, col = "lightseagreen", xlab = "Temperatuur (C)", ylab = "
  Frequentie", main = "")

# Plot verschil in temperatuur
scatter2D(y$verschiltemp, y$tempschil, colvar = y$temp.x, pch = 20, xlab = expression(Temp[
  Koksijde] - Temp[Middelkerke]), ylab = expression((T[i] - T[i-1])^2 + (T[i] - T[i-1])^2
  ))
```

```
# Correlatielot temperaturen in Middelkerke t.o.v. deze in Koksijde
plot(Tempschil$Temp.x, Tempschil$Temp.y, xlab = "Temperatuur Koksijde", ylab = "Temperatuur
      Middelkerke", pch = "*", xlim = c(-20,35), ylim = c(-20,35))
curve(x+0, col = "gray", add = TRUE)
curve(x+10, col = "red", add = TRUE)
curve(x-0, col = "red", add = TRUE)
```

### C.3.4 Neerslag

#### Ontbrekende waarden

```
# Histogrammen neerslaghoeveelheden
hist(DATASET$prec, breaks = 200, col = "lightseagreen", xlab = "Neerslag (mm)", ylab = "
      Frequentie", main = "")

# Percentage niet-geregistreerde metingen bepalen per maand (voor Koksijde, idem voor
      Middelkerke). Kok6u is een uitgedund dataframe met de metingen om 6 UTC en 18 UTC (de
      enige momenten waarop neerslag gemeten werd)
Kok6u <- Koksijde2017[which(Koksijde2017$HOUR == 6 | Koksijde2017$HOUR == 18),]
Kok6u$Date <- with(Kok6u, sprintf("%d-%02d", Kok6u$YEAR, Kok6u$MONTH))
A <- unique(Kok6u$Date)
Kokmaanden <- data.frame(matrix(0,783,1)) # vector met evenveel rijen als er maanden zijn
colnames(Kokmaanden)[1] <- "maand"
Kokmaanden$maand <- A
for(i in 1:783){Kokmaanden$aantalna[i] <- length(Kok6u[which(Kok6u$Date == A[i] & is.na(
      Kok6u$prec)],]$Date)}
for(i in 1:783){Kokmaanden$aantal[i] <- length(Kok6u[which(Kok6u$Date == A[i]),]$Date)}
Kokmaanden$fractiena <- (Kokmaanden$aantalna/Kokmaanden$aantal)*100
Kokmaanden$getal <- c(1:783)

# Plot van het percentage niet-geregistreerde metingen
plot(Midmaanden$getal, Midmaanden$fractiena, col = "white", xlab= "Maand na begin van de
      metingen", ylab= "Aantal ontbrekende waarden (%)", pch = ".", xaxt="n",ylim = c(0,90))
# geen waarden op de x-as weergeven
axis(1, at = seq(0, 750, by = 50), las=2) # waarden op de x-as zelf meegeven
lines(Midmaanden$getal, Midmaanden$fractiena, col = "darkblue")

# Gemiddeld percentage ontbrekende waarden berekenen vanaf 2004
mean(Midmaanden[which(Midmaanden$getal > 605),]$fractiena)
```

## C.4 Verdeling van de windsnelheden in Koksijde en Middelkerke

```
# QQ-plot Koksijde of Middelkerke normale verdeling
qqnorm(DATASET$ws, col = "darkblue", main = "")
```

### C.4.1 Rayleighdistributie

#### Wiskundige analyse van cumulatieve en kansdichtheidsfunctie

```
# Plotten cumulatieve Rayleigh voor variabele c
```

```

curve(1-exp((-x^2)/(2*(WAARDE VOOR C)^2)), xlim = c(0,10), col = "KLEUR", xlab = "", ylab =
  "", lwd = 2, col = "darkblue", add = TRUE)

# Legende toevoegen aan plot
legend("bottomright", c("c=0.5", "c=1", "c=2","c=3", "c=4"), lwd = c(1,2,3,2,2), col = c("
  black", "darkblue","blue","lightskyblue3","lightseagreen"))

# Plotten dichtheidsfunctie Rayleigh voor variabele c
curve(x/((1/2)^2)*exp((-x^2)/(2*(WAARDE VOOR C)^2)), xlim = c(0,10), col = "black", xlab =
  "", ylab = "", main = "Rayleighdistributie met variabele c")
legend("topright", c("c=0.5", "c=1", "c=2","c=3", "c=4"), lwd = c(1,2,3,2,2), col = c("
  black", "darkblue","blue","lightskyblue3","lightseagreen")) # Legende voor plot

# Correlatieplot van de u- en v- component
plot(Koksijde2017$u, Koksijde2017$v, xlab = "u-component", ylab="v-component", col = "
  darkblue", pch = '.' ,main = "Koksijde")
plot(Middelkerke2017$u, Middelkerke2017$v, xlab = "u-component", ylab="v-component", col =
  "darkblue", pch = '.', main = "Middelkerke")

```

### Geschied voor de modellering van windsnelheden?

```

# Rayleigh-QQ-plot voor Koksijde en Middelkerke
x.wei<-rweibull(n=N,shape=2,scale=C*sqrt(2))
qqplot(x.wei, DATASET$ws, main="", xlab = "Theoretic Quantiles", ylab = "Sample Quantiles")
abline(0,1, col = "red")

# QQ-plot u- en v-componenten (analoog voor Middelkerke)
qqnorm(Koksijde2017$ucomp, col = "darkblue", main = "")
qqline(Koksijde2017$ucomp, col = "red")
qqnorm(Koksijde2017$vcomp, col = "darkblue", main = "")
qqline(Koksijde2017$vcomp, col = "red")

# Correlatieplot van de orthogonale componenten van de windvector
plot(Koksijde2017$u, Koksijde2017$v, xlim = c(-25,25), ylim = c(-25,25))

# Significantie correlatie testen (Pearson's product-moment correlation)
cor.test(Middelkerke2017$u, Middelkerke2017$v)

# Test voor normaliteit (Anderson-Darling normality test)
ad.test(Koksijde2017$u)

# Test voor homogeniteit van varianties (Bartlett test of homogeneity of variances)
bartlett.test(list(Middelkerke2017$u,Middelkerke2017$v))

```

## C.4.2 Weibullverdeling

### Wiskundige analyse van cumulatieve en kansdichtheidsfunctie

```

# Plotten cumulatieve Weibull met variabele c
curve((1.8/WAARDE VOOR C)*(x/WAARDE VOOR C)^0.8*exp(-(x/WAARDE VOOR C)^1.8), col = "KLEUR",
  lty=1, ylim = c(0,0.23), xlim = c(0,24), lwd=1, xlab = "", ylab = "", main = "
  Weibulldistributie met variabele c en k = 1.8")

```

```

curve((1.8/WAARDE VOOR C)*(x/WAARDE VOOR C)^0.8*exp(-(x/WAARDE VOOR C)^1.8), col = "KLEUR",
      lwd=2,add=TRUE)

# Legende toevoegen aan plot
legend("topright", c("c=4", "c=5", "c=5.5","c=6", "c=7"), lwd = c(1,2,3,2,2), col = c("
  black","darkblue","blue","lightskyblue3","lightseagreen"))

# Plotten dichtheidsfuncties Weibull met variabele c
curve(1-exp(-(x/WAARDE VOOR C)^1.8), ylim = c(0,1), xlim = c(0,30), xlab = "", ylab = "",
      main = "Weibull CDF met variabele c en k=1.8")
curve(1-exp(-(x/WAARDE VOOR C)^1.8), col = "darkblue", lwd = 2, add= TRUE)
legend("bottomright", c("c=4", "c=5", "c=5.5","c=6", "c=7"), lwd = c(1,2,3,2,2), col = c("
  black", "darkblue","blue","lightskyblue3","lightseagreen"))

# Plotten cumulatieve verdelingsfunctie Weibull met variabele k
curve(1-exp(-(x/5)^WAARDE VOOR K), ylim = c(0,1), xlim = c(0,30), xlab = "", ylab = "",
      main = "Weibull CDF met variabele k en c=5")
curve(1-exp(-(x/5)^WAARDE VOOR K), col = "KLEUR", lwd = 2, add= TRUE)
legend("bottomright", c("k=1.1", "k=1.5", "k=2","k=3", "k=6"), lwd = c(1,2,3,2,2), col = c(
  ("black", "darkblue","blue","lightskyblue3","lightseagreen"))

# Plotten cumulatieve verdelingsfunctie Weibull met variabele k
curve((WAARDE VOOR K/5)*(x/5)^(WAARDE VAN K - 1)*exp(-(x/5)^WAARDE VOOR K), col = "black",
      lty=1, ylim = c(0,0.43), xlim = c(0,24), lwd=1, xlab = "", ylab = "", main = "
  Weibulldistributie met variabele k en c = 5")
curve((WAARDE VOOR K/5)*(x/5)^(WAARDE VOOR K - 1)*exp(-(x/5)^WAARDE VOOR K), col = "
  darkblue", lwd=1.5, add = TRUE)
legend("topright", c("k=1.1", "k=1.5", "k=2","k=3", "k=6"), lwd = c(1,2,3,2,2), col = c("
  black", "darkblue","blue","lightskyblue3","lightseagreen"))

```

#### C.4.4 Schatten van de Weibullparameters voor de windsnelheden in Koksijde en Middelkerke

##### Schatting met de maximum-likelihoodmethode

```

# Schatten van de Weibullparameters aan de hand van de maximum likelihoodmethode
fit_weibull(na.omit(koksijde[which(koksijde$YEAR==X),]$'Wind speed (m/s)'))
# Indien gewenst kan het seizoen, de maand, het jaartal,... meegegeven worden.

# Histogrammen met daarboven de Weibullkansverdelingsfunctie met geschatte parameters
x=sort(na.omit(GEWENSTE DATASET$wps)) # wps staat voor de rukwinden (Wind Peak Speed)
X <- na.omit(GEWENSTE DATASET$wps)
hist(X, prob=TRUE, breaks = 40, xlim = c(0,30), xlab = "Windsnelheid (m/s)", ylab = "
  Dichtheid", main = "STATION (JAAR/SEIZOEN)", col = "lightgray")
lines(x, dweibull(x, shape=GESCHATTE k, scale=GESCHATTE c),col="blue", lwd = 2)

# Randow Weibullverdeelde dataset genereren
rweib <- rweibull(220 000, 1.75, 5.19)
rweib <- data.frame(rweib)
rweibklein <- data.frame(rweib[which(rweib$rweib<0.5),])
rweibklein2 <- rweibklein[-c(1,2,3,4,5,6), ]
koksijdezondernullen <- koksijde

```

```

nullen <- koksijdezondernullen[which(koksijdezondernullen$ws == 0), 7]
koksijdezondernullen[which(koksijdezondernullen$ws == 0), 7] <- rweibklein2
koksijdezondernullen[which(koksijdezondernullen$ws == 0), 7]

# Aantal nulmetingen bepalen
length(DATASET[which(DATASET$ws==0),]$ws)

# Dataset opsplitsen in twee delen aan de hand van random true en false toe te kennen aan
  elke entry
l <- length(koksijdezondernullen$ws)
koksijdezondernullen$eerstehelft <- (runif(l) < rep(0.5,l))
eerstehelft <- koksijdezondernullen[which(koksijdezondernullen$eerstehelft==TRUE),]
hist(koksijdezondernullen$ws)
hist(koksijdezondernullen[which(koksijdezondernullen$ws<=2),]$ws)

# Iteraties uitvoeren met behulp van while-lus
knieuw = 2
koud <- knieuw
knieuw <- (sum(lentezonder0$'Wind speed (m/s)'^~koud*lentezonder0$ln)/sum(lentezonder0$'Wind
  speed (m/s)'^~koud)-sum(lentezonder0$ln)/21986)^(-1)
A <- while(abs(knieuw-koud)<0.001) {knieuw <- koud; print(knieuw);}

# Plot met iteraties voor de maximum-likelihoodmethode
plot(A, col = "darkblue", pch = 20, xlab = "i-de iteratie", ylab = "shape parameter k")

Schatting met de power density method

# Berekening Energy Pattern Factor
[mean((na.omit(Koksijde2017$ws))^3)/(mean(na.omit(Koksijde2017$ws))^3)]/3,69 + 1

# Plot gefitte Weibullkansverdeling, geschat via MLM en via power density methode.
x=sort(na.omit(Koksijde2017$ws))
X <- na.omit(Koksijde2017$ws)
hist(X, prob=TRUE, breaks = 30, ylim = c(0,0.17), xlab = "Windsnelheid (m/s)", ylab = "
  Dichtheid", main = "Weibull vs Rayleigh")
curve(x/3.83^2*exp(-x^2/(2*3.83^2)), add = TRUE, col = "red", lwd = 2)
lines(x, dweibull(x, shape=1.63, scale=5.15),col="blue", lwd = 2) # schaal-en vormparameter
  berekend met MLM

# Histogrammen boven elkaar
b=sort(na.omit(Koksijde2017[which(Koksijde2017$SEIZOEN=="X"),]$wps))
B <- na.omit(Koksijde2017[which(Koksijde2017$SEIZOEN=="X"),]$wps)
c=sort(na.omit(Middelkerke2017[which(Middelkerke2017$SEIZOEN=="X"),]$wps))
C <- na.omit(Middelkerke2017[which(Middelkerke2017$SEIZOEN=="X"),]$wps)

hist(B, prob=TRUE, breaks = 40, xlim = c(0,30), xlab = "Windsnelheid (m/s)", ylab = "
  Dichtheid", main = "Lente", col = rgb(0.8,0.8,0.8,0.4))
hist(C, prob=TRUE, breaks = 40, col = rgb(0.5,0.9,0.2,0.2), add = T)

# Weibullkansverdelingsfuncties plotten bovenop de histogrammen
lines(b, dweibull(b, shape=GESCHATTE k, scale=GESCHATTE c),col="blue", lwd = 2)

```

```
# Vergelijking power density en maximum-likelihoodmethode
par(mfrow = c(1,2))
x=sort(na.omit(Koksijde2017$ws))
X <- na.omit(Koksijde2017$ws)
hist(X, prob=TRUE, breaks = 30, ylim = c(0,0.2), xlab = "Windsnelheid (m/s)", ylab = "
  Dichtheid", main = "Koksijde (jaar)", col = "gray91")
lines(x, dweibull(x, shape=1.63, scale=5.15),col="blue", lwd = 2)
lines(x, dweibull(x, shape=2.5667, scale=5.187),col="red", lwd = 2)
y=sort(na.omit(Middelkerke2017$ws))
Y <- na.omit(Middelkerke2017$ws)
hist(Y, prob=TRUE, breaks = 30, ylim = c(0,0.2), xlab = "Windsnelheid (m/s)", ylab = "
  Dichtheid", main = "Middelkerke (jaar)", col = "gray91")
lines(y, dweibull(y, shape=1.84, scale=5.98),col="blue", lwd = 2)
lines(y, dweibull(y, shape=2.7294, scale=5.96729),col="red", lwd = 2)
```

### C.4.5 Geschied voor de modellering van windsnelheden?

```
# Weibull-QQ-plot
x.wei<-rweibull(n=N,shape=k,scale=c)
qqplot(x.wei, DATASET$ws, pch = 20, col = "darkblue", xlab = "Theoretical Quantiles", ylab
  = "Sample Quantiles")
abline(0,1, col = "red")

x=sort(na.omit(DATASETKOKSIJDE$ws))
X <- na.omit(DATASETKOKSIJDE$ws)
y=sort(na.omit(DATASETMIDDELKERKE$ws))
Y <- na.omit(DATASETMIDDELKERKE$ws)
hist(X, prob=TRUE, breaks = 40, xlab = "Windsnelheid (m/s)", ylab = "Dichtheid", col = rgb
  (0,0,1,0.2), main = "Weibull Koksijde vs Middelkerke (SEIZOEN)")
hist(X, prob=TRUE, breaks = 40, xlim = c(0,20), xlab = "Windsnelheid (m/s)", ylab = "
  Dichtheid", col = rgb(0,0,1,0.2), main = "Weibull Koksijde vs Middelkerke (SEIZOEN)")
```

### C.4.7 Verdeling van rukwinden

```
# Histogrammen van de rukwinden met daarboven de Weibullkansdichtheidsfunctie
hist(X, prob=TRUE, breaks = 70, ylim = c(0,0.10), xlim = c(0,35), xlab = "Windsnelheid (m/s
  )", ylab = "Dichtheid", main = "Koksijde (jaar)", col = "lightgray")
lines(x, dweibull(x, shape=2.201, scale=10.039),col="blue", lwd = 2)
Y <- na.omit(Middelkerke2017$wps)
hist(Y, prob=TRUE, breaks = 70, ylim = c(0,0.10), xlim = c(0,35), xlab = "Windsnelheid (m/s
  )", ylab = "Dichtheid", main = "Middelkerke (jaar)", col = "lightgray")
lines(x, dweibull(x, shape=2.211, scale=10.649),col="blue", lwd = 2)
```

## C.5 Testen en methoden voor trendanalyse

### C.5.7 Vergelijking van de power

**Benchmarking: schatting kans op type-I-fout**

```
P <- data.frame(matrix(0, 1000,3)) # 1000 X 3
NORM <- data.frame(matrix(0,65,1))
```



```

NORM$jaar <- c(1952:2016)
S <- matrix(0, 200,1)

# Lineaire regressie
for(i in 1:1000){
NORM[,1] <- rnorm(65,mean=GEMIDDELDE WINDSNELHEID, sd=STANDAARDDEVIATIE WINDSNELHEID)
P[i,1] <- summary(lm(NORM$NORM~NORM$jaar))$coefficients[[8]] #p-waarde

# Mann-Kendalltest
for(i in 1:1000){
  NORM[,1] <- rnorm(65,mean=GEMIDDELDE WINDSNELHEID, sd=STANDAARDDEVIATIE
WINDSNELHEID)
  ts <- ts(NORM$NORM, start = 1952, end = 2016, frequency = 1)
  P[i,2] <- mk.test(ts)[5]
}

# Permutatietest
for(i in 1:1000){
  NORM[,1] <- rnorm(65,mean=GEMIDDELDE WINDSNELHEID, sd=GEMIDDELDE WINDSNELHEID)
  for (j in 1:200){
    NORM$perm <- sample(NORM[,1])
    S[j,1] <- summary((lm(NORM$perm ~ NORM$jaar)))$coefficients[[2]]
  }
  P[i,3] <- ecdf(S)(summary((lm(NORM$NORM ~ NORM$jaar)))$coefficients[[2]])
}

```

### Testen van de power voor een verschil van 0,01 m/s per jaar

```

# Genereren en testen onder alternatieve hypothese (voor Koksijde)
alter <- data.frame(matrix(0,65,1))
alter$jaar <- c(1952:2016)
K <- data.frame(matrix(0, 1000,3))
V <- matrix(0, 100,1)
for(i in 1:1000){
alter[,1] <- 4.2757 + 0.01*alter$jaar + rnorm(65,mean=0, sd=0.4452)
K[i,1] <- summary(lm(alter$col1~alter$jaar))$coefficients[[8]]

for(i in 1:1000){
alter[,1] <- 4.2757 + 0.01*alter$jaar + rnorm(65,mean=0, sd=0.4452)
ts <- ts(alter$col1, start = 1952, end = 2016, frequency = 1)
K[i,2] <- mk.test(ts)[5]}

for(i in 1:1000){
alter[,1] <- 4.2757 + 0.01*alter$jaar + rnorm(65,mean=0, sd=0.4452)
  for(j in 1:100){
    alter$perm <- sample(alter$col1)
    V[j,1] <- summary((lm(alter$perm ~ alter$jaar)))$coefficients[[2]] #
beta_1
  }
  K[i,3] <- ecdf(V)(summary((lm(alter$col1. ~ alter$jaar)))$coefficients[[2]])
}

```

## C.6 Trends in de windsnelheden

### C.6.1 Boxplots voor de windsnelheden in de verschillende seizoenen

```
# Jaargemiddelde wind/temperatuur berekenen per seizoen
SEIZOENgemiddeldetemp <- aggregate(koksijde[koksijde$SEIZOEN=="X"],$'Temperature', by =
  list(koksijde[koksijde$SEIZOEN=="X"],$YEAR), FUN=mean, na.rm=TRUE)

# Seizoensgemiddelde windsnelheden berekenen
par(mfrow=c(2,2))
plot(SEIZOENgemSTATION$JAAR, SEIZOENgemSTATION$x, col = "white", pch = ".", xlab = "Jaar",
  ylab = "Windsnelheid (m/s)", main = "SEIZOEN STATION")
lines(SEIZOENgemSTATION$JAAR, SEIZOENgemSTATION$x, col = "darkblue")
curve(beta_0 - beta_1*(x-195X), add = T , col = "chartreuse2", lwd = 2)

# Boxplots per seizoen voor Koksijde en Middelkerke
par(mfrow=c(2,2))
boxplot(DATASET[which(DATASET$SEIZOEN=="X"),]$ws ~ DATASET[which(DATASET$SEIZOEN=="X"),]
  $YEAR, ylim = c(0,25), main = "Windsnelheden SEIZOEN")
```

### C.6.2 Jaar- en seizoensgemiddelden

```
# Plotten van de jaarlijkse gemiddelde windsnelheid in Koksijde en Middelkerke
plot(jaarijksgemiddelde$Group.1, jaarijksgemiddelde$x, col = "darkblue", pch = 20, xlab =
  "Jaar", ylab = "Windsnelheid (m/s)", main = "Jaarlijkse gemiddelde windsnelheid
  STATION 195X-2016")
```

#### LOWESS

```
# lowessplots
plot(SEIZOENgemSTATION$jaar, SEIZOENgemSTATION$x, pch = 20, col = "darkblue", xlab = "Jaar
  ", ylab = "Windsnelheid (m/s)", main = "LOWESS (SEIZOEN)")
lines(lowess(SEIZOENgemSTATION$jaar, SEIZOENgemSTATION$x, f = 0.75), col = "green", lwd =
  2)
lines(lowess(SEIZOENgemSTATION$jaar, SEIZOENgemSTATION$x, f = 0.3), col = "red", lwd = 2)
```

#### Lineaire regressie

```
# Parameters schatten met lineaire regressie
summary(lm(DATASET$VARIABLE~DATASET$jaar))

# Plotten van gemiddelden met daarboven trendlijn volgens de geschatte parameters (analoog
  voor alle andere dergelijke plots in de volgende secties en hoofdstukken)
plot(DATASET$jaar, DATASET$x, col = "white", pch = ".", xlab = "Jaar", ylab = "Windsnelheid
  (m/s)", main = "TITEL")
lines(DATASET$jaar, DATASET$x, col = "darkblue")
curve(beta_0 + beta_1 *(x-195X), add = T , col = "chartreuse2", lwd = 2)
```

#### Polynomiale regressie

```
# Polynomiale regressie met polynomen van graad 2, 3 en 4.
model1 <- lm(lenterm$x ~ poly(lenterm$JAAR,2))
model2 <- lm(lenterm$x ~ poly(lenterm$JAAR,3))
model3 <- lm(lenterm$x ~ poly(lenterm$JAAR,4))

# Plotten polynomen
plot(lenterm$JAAR, lenterm$x, pch = 20, col = "blue", xlab = "Jaar", ylab = "Windsnelheid (
  m/s)", main = "Polynomiale regressie")
lines(lenterm$JAAR, predict(model1), col="green3", lwd = 2)
lines(lenterm$JAAR, predict(model2), col="greenyellow", lwd = 2)
lines(lenterm$JAAR, predict(model3), col="green4", lwd = 2)
```

### Niet-parametrische lineaire regressie

```
# Permutatietest
P <- matrix(0, 1000,1)
for (i in 1:1000){
  DATASET$P <- sample(DATASET$x)}
V[i,1] <- summary((lm(DATASET$P ~ DATASET$jaar)))$coefficients[[2]]

# Significantieniveau bepalen bij permutatietest
sum(V$V < lm(GemWindKok$x~GemWindKok$jaar)$coefficient[[2]]) # aantal gegenereerde beta_1 <
  oorspronkelijke beta_1
```

### Mann-Kendalltest en Sen's slope

```
# Package nodig voor Mann-Kendall
install.packages("trend")
library("trend", lib.loc="C:/Program Files/R/R-3.4.0/library")

# Mann-Kendalltest uitvoeren op jaar- of seizoensgemiddelden
ts <- ts(DATASET$x, start = 195X, end = 2016, frequency = 1)
mk.test(ts)

# Sen's slope berekenen, met 95% betrouwbaarheidsinterval voor beta_1
sens.slope(ts, level = 0.95)
```

### C.6.3 Maandgemiddelde windsnelheden

```
# Maandgemiddelden berekenen
MAAND <- aggregate(Koksijde2017[Koksijde2017$MONTH==X,]$ws, by = list(Koksijde2017[
  Koksijde2017$MONTH==X,]$YEAR), FUN=mean, na.rm=TRUE)
```

### C.6.4 Dagen met krachtige wind

```
# Aantal dagen met windsnelheden boven 11 m/s berekenen
boven11 <- Koksijde2017[which(Koksijde2017$ws > 11),]
boven11$Date <- with(boven11, sprintf("%d-%02d-%02d", boven11$YEAR, boven11$MONTH,
  boven11$DAY))
C <- data.frame(unique(boven11$Date))
write.csv(C, file = "boven11.csv") # dataset aangepast in Excel
boven11 <- read_csv("~/data/boven11.csv")
```

```

D <- data.frame(table(boven11$jaar))
D$Freq <- as.numeric(D$Freq)
D$jaar <- c(1952:2017)
D <- D[-c(66),]

# plot aantal dagen (4 tresholds)
plot(dagen$jaar, dagen$dagen, col = "darkblue", pch = ".", xlab = "Jaar", ylab = "Aantal
metingen", ylim = c(0,220))
lines(dagen$jaar, dagen$dagen, col = "darkblue", lwd = 2) # aantal dagen boven 10 m/s
lines(dagen$jaar, dagen$V4, col = "red", lwd = 2) # aantal dagen boven 12 m/s
lines(dagen$jaar, dagen$V6, col = "chartreuse3", lwd = 2) # aantal dagen boven 14 m/s
lines(dagen$jaar, dagen$V8, col = "orange", lwd = 2) # aantal dagen boven 16 m/s
legend("topright", c("> 10 m/s", "> 12 m/s", "> 14 m/s", "> 16 m/s"), lwd = c(2,2,2,2), col
= c("darkblue", "red",
"chartreuse3", "orange"), cex = 0.90)
curve(y = 0, col = "gray76", add=T)

```

### C.6.5 Dagen met zwakke tot matige wind

```

# Plots met verschillende thresholds
plot(N$jaar, N$Freq, col = "white", pch = ".", xlab = "Jaar", ylab = "Aantal dagen", ylim =
c(0,100))
lines(N$jaar, N$Freq, col = "darkblue")
lines(R$jaar, R$Freq, col = "red")
lines(V$jaar, V$Freq, col = "limegreen")
lines(WW$jaar, WW$Freq, col = "orange")
lines(abline(h = 0, col = "gray76"))
# De dataset N, R, V en W bevatten het aantal dagen per jaar waarop de windsnelheid een
bepaalde waarde overschrijdt.

```

### C.6.7 Trendanalyse op de verdelingsparameters

#### Visuele inspectie

```

# Plots met geschatte Weibullparameters per jaar
DATASETMETPARAMETERS <- for(YEAR in c(1952:2016)){print(fit_weibull(na.omit(DATASET$ws)))}
plot(DATASETMETPARAMETERS$JAAR, DATASETMETPARAMETERS$c1, pch=".", axes=F, xlab=NA, ylab=NA,
cex=1, ylim = c(3.6,7.4), col = "cadetblue4", lwd = 1)
lines(DATASETMETPARAMETERS$JAAR, DATASETMETPARAMETERS$VARIABLE, pch=".", axes=F, xlab=NA,
ylab=NA, cex=1, ylim = c(3,8), col = "burlywood4", lwd = 1)
axis(side = 4)
mtext(side = 4, line = 3, 'Weibullparameter c')

# Weibullparameterplot voor Koksijde
par(mfrow = c(2,2))
matplot(parameters$JAAR, cbind(parameters$kl, parameters$c1), ylab="Weibullparameter k",
col=c("red", "green"), type="l", lwd = 2, xlab = "Jaar", main = "Lente Koksijde")
matplot(parameters$JAAR, cbind(parameters$kg, parameters$c2), ylab="Weibullparameter k",
col=c("red", "green"), type="l", lwd = 2, xlab = "Jaar", main = "Zomer Koksijde")
matplot(parameters$JAAR, cbind(parameters$kh, parameters$ch), ylab="Weibullparameter k",
col=c("red", "green"), type="l", lwd = 2, xlab = "Jaar", main = "Herfst Koksijde")
matplot(parameters$JAAR, cbind(parameters$kw, parameters$cw), ylab="Weibullparameter k",
col=c("red", "green"), type="l", lwd = 2, xlab = "Jaar", main = "Winter Koksijde")

```

## C.6.8 Discussie

```
# Goldfeld-Quandttest om homoscedasticiteit na te gaan
gqtest(MAAND$x~MAAND$jaar)

# Gepaarde t-test om na te gaan of het in de winter significant meer waait
t.test(winterrm1$x,herfstrm1$x,paired=TRUE)
t.test(winterrm1$x,zomerrm1$x,paired=TRUE)
t.test(winterrm1$x,lenterrm1$x,paired=TRUE)

# Correlatiecoefficient van opeenvolgende residuen bepalen
m1 <- lm(GemWindKok$x~GemWindKok$jaar)
Residuen <- data.frame(resid(m1))
residue2 <- rbind(Residuen, NA)
residue2["res2"] <- rbind(NA, Residuen)
cor(residue2$res, residue2$res2)

# Durbin-Watsonstest
dwtest(DATAFRAME$x~DATAFRAME$jaar)
```

## Variatie van de windsnelheid

```
# Variantie per jaar berekenen
variantiejaarKok <- aggregate(Koksijde2017$ws, by = list(Koksijde2017$YEAR), FUN=var, na.rm
=TRUE)
variantiejaarMid <- aggregate(Middelkerke2017$ws, by = list(Middelkerke2017$YEAR), FUN=var,
na.rm=TRUE)

# Varianties per jaar plotten
plot(varantiejaarKok$Group.1, variantiejaarKok$x, col = "white",xlab = "Jaar", ylab =
expression(Variantie~(m^2/s^2)), pch = ".", main = "Variantie Koksijde (jaar)")
plot(varantiejaarMid$Group.1, variantiejaarMid$x, col = "white",xlab = "Jaar", ylab =
expression(Variantie~(m^2/s^2)), pch = ".", main = "Variantie Jaar Middelkerke")

# Variabiliteit wind
c <- abs(belgo[2:(nrow(belgo)-1),"ws"]-belgo[1:(nrow(belgo)-2),"ws"])
belgo$windvar <- c(NA,c$ws, NA)

# Gemiddelde variantie per maand
varpermaand <- aggregate(belgo$windvar, by = list(belgo$maand), FUN=mean, na.rm=TRUE)
varpermaand$getal <- c(1:704)
plot(varpermaand$getal, varpermaand$x)

# Min-max plot
maxpermaand <- aggregate(belgo$windvar, by = list(belgo$maand), FUN=max, na.rm=TRUE)
maxpermaand$min <- aggregate(belgo$windvar, by = list(belgo$maand), FUN=min, na.rm=TRUE)
plot(maxpermaand$getal, maxpermaand$x, ylim = c(3,17), pch = 20)
lines(maxpermaand$getal, maxpermaand$x)

# Datum maken in R
DATASET$Date <- with(DATASET, sprintf("%d-%02d", DATASET$YEAR, DATASET$MONTH))
```

```
DATASET$Date <- with(DATASET, sprintf("%d-%02d-%02d", DATASET$YEAR, DATASET$MONTH,
  DATASET$DAY))
```

## C.7 Kusttemperatuur

### C.7.1 Beschrijvende analyse

```
# Boxplots voor de temperaturen in Koksijde en Middelkerke
boxplot(Koksijde2017$temp, Middelkerke2017$temp, horizontal = TRUE, names=c("Koksijde", "
  Middelkerke"), xlab = "Temperatuur (C)")
```

### C.7.2 Trendanalyse op de temperatuur

```
# Gemiddelde jaartemperatuur in Koksijde en Middelkerke
GemTempSTATION <- aggregate(DATASET$`Temperature (C)`, by = list(DATASET$YEAR), FUN=mean,
  na.rm=TRUE)

# Plotten van drie reeksen jaargemiddelde temperaturen (Ukkel, Middelkerke, Koksijde)
matplot(GemTemp$Jaar, cbind(GemTemp$Temp, GemTemp$Kok, GemTemp$Mid), ylab="Temperatuur (C)
  ", col=c("blue", "red", "green"), type="l", lwd = 2, xlab = "Jaar")
legend("topleft", c("Ukkel", "Koksijde", "Middelkerke"), lwd = c(2,2,2), col = c("blue", "
  red", "green"), cex = 0.85)

# Legende toevoegen aan de plot
legend("bottomright", c("Ukkel", "Koksijde", "Middelkerke"), lwd = c(2,2,2), col = c("blue
  ", "red", "green"), cex = 0.65)
```

### Niet-parametrische lineaire regressie

```
# Histogrammen schatters betal
P <- data.frame(matrix(0, 100000,1))
for (i in 1:100000){
  GemTemp$P <- sample(GemTemp$Kok)
  P[i,1] <- summary((lm(GemTemp$P ~ GemTemp$Jaar)))$coefficients[[2]]}
hist(P$P, breaks = 70, main = "", xlab = expression(hat(beta)[1]), ylab = "Frequentie")
```

## C.8 De zeebries

### C.8.1 Beschrijvende analyse van windrichtingen

```
# Windrozen
library("openair", lib.loc="C:/Program Files/R/R-3.4.0/library")
windRose(DATASET, ws = "ws", wd = "wd", ws.int = 1, angle = 10, type = "default", cols= c("
  green", "yellow", "orange", "red", "deeppink3"), breaks = c(0,3.3,7.9,13.8,20), grid.
  line = list(value = 1, lty = 2, col ="gray85"), width = 2, auto.text = TRUE, offset =
  10,key.position = "bottom", annotate = FALSE, angle.scale = -45, normal = FALSE)
```

### C.8.4 Zeebries identificeren in de dataset: zeebriesdagen

```

# Zeebries uit uurlijkse dataset halen
MaartSept <- belgo[which(belgo$Month == 3 | belgo$Month == 4 | belgo$Month == 5 |
belgo$Month == 6 | belgo$Month == 7 | belgo$Month == 8),] # Enkel de maanden maart t.e.m.
augustus worden gebruikt.

# Voorwaarde waaraan de wind 's morgens moet voldoen
vroeg <- MaartSept[which((MaartSept$Hour == 4 | MaartSept$Hour == 5 | MaartSept$Hour == 6 |
MaartSept$Hour == 7) & (MaartSept$ws > 1.5 & MaartSept$ws < 5.5) &
(MaartSept$wd > 60 & MaartSept$wd < 230)),]

& (MaartSept$wd > 269 | MaartSept$wd < 35)),] # Voorwaarde waaraan de wind 's middags moet
voldoen

laat <- MaartSept[which((MaartSept$Hour==11 | MaartSept$Hour==12 | MaartSept$Hour==13 |
MaartSept$Hour==14 | MaartSept$Hour==15) & (MaartSept$ws >3.3 & MaartSept$ws < 10.8)

# Doorsnede van de twee subdatasets
y<-merge(vroeg,laat,by=c("moment"))
z <- data.frame(y)

# Enkel de dagen met weinig regen worden in rekening gebracht
weinigregen <- Middelkerke2017[which(Middelkerke2017$HOUR==18 & Middelkerke2017$prec==0.0)
,]
weinigregen$moment <- paste(weinigregen$YEAR,"-",weinigregen$MONTH,"-",weinigregen$DAY)

```

### C.8.5 Resultaat van de analyse

#### Aantal zeebriesdagen

```

dagen <- aggregate(y$ws.y, by = list(y$moment), FUN=mean, na.rm=TRUE)
zeebriesdagen <- merge(dagen,weinigregen, by = c("moment")) # Aantalzeebriesdagen per jaar

GemZeebries <- aggregate(zeebriesdagen$x, by = list(zeebriesdagen$YEAR), FUN=mean, na.rm=
TRUE) # Gemiddelde sterkte van de zeebries per jaar

# Aantal zeebriesdagen plotten
plot(aantalzeebriesdagen$jaar, aantalzeebriesdagen$Freq, col = "darkblue", pch = 20, xlab =
"Jaar", ylab = "Aantal zeebriesdagen")
lines(aantalzeebriesdagen$jaar, aantalzeebriesdagen$Freq, col = "darkblue")
curve(35.1667+0.0833*(x-1955),col = "chartreuse2", lwd = 2, add = TRUE)

```

### C.9 Conclusies

```

# Plotten van drie reeksen jaargemiddelde windsnelheden (Ukkel, Middelkerke, Koksijde)
matplot(GemWindUkkel$Jaar, cbind(GemWindUkkel$Kok, GemWindUkkel$Mid), ylab="Windsnelheid (m
/s)", col=c("red", "green"), type="l", lwd = 2, xlab = "Jaar")
legend("topright", c("Middelkerke","Koksijde"), lwd = c(2,2,2), col = c("green", "red"),
cex = 0.85)

```

# Bibliografie

- [1] AHRENS, C. D. *Meteorology today: an introduction to weather, climate, and the environment*. Cengage Learning, 2012.
- [2] AKDAĞ, S. A., AND DINLER, A. A new method to estimate Weibull parameters for wind energy applications. *Energy conversion and management* 50, 7 (2009), 1761–1766.
- [3] BAKKER, A. M., AND VAN DEN HURK, B. J. Estimation of persistence and trends in geostrophic wind speed for the assessment of wind energy yields in Northwest Europe. *Climate dynamics* 39, 3-4 (2012), 767–782.
- [4] BAULE, W. J., AND SHULSKI, M. D. Climatology and trends of wind speed in the Beaufort/Chukchi Sea coastal region from 1979 to 2009. *International Journal of Climatology* 34, 8 (2014), 2819–2833.
- [5] BICHET, A., WILD, M., FOLINI, D., AND SCHÄR, C. Causes for decadal variations of wind speed over land: Sensitivity studies with a global climate model. *Geophysical Research Letters* 39, 11 (2012).
- [6] BORNE, K., CHEN, D., NUNEZ, M., ET AL. A method for finding sea breeze days under stable synoptic conditions and its application to the Swedish west coast. *International Journal of Climatology* 18, 8 (1998), 901–914.
- [7] BROUWERS, J., PEETERS, B., STEERTEGEM, M. V., ET AL. Klimaatrapport 2015, MIRA. Tech. rep., Vlaamse Milieumaatschappij, 2015.
- [8] BROUYAUX, F., ET AL. Oog voor het klimaat. Tech. rep., Koninklijk Meteorologisch Instituut van België, 2015.
- [9] BROWER, M., BARTON, M., LLEDÓ, L., AND DUBOIS, J. A study of wind speed variability using global reanalysis data. *AWS Truepower, New York* (2013).
- [10] CAMINADE, C., MEDLOCK, J. M., DUCHEYNE, E., MCINTYRE, K. M., LEACH, S., BAYLIS, M., AND MORSE, A. P. Suitability of European climate for the Asian tiger mosquito *Aedes albopictus*: recent trends and future scenarios. *Journal of the Royal Society Interface* (2012), rsif20120138.
- [11] DADASER-CELIK, F., AND CENGIZ, E. Wind speed trends over Turkey from 1975 to 2006. *International Journal of Climatology* 34, 6 (2014), 1913–1927.
- [12] DEBOOSERE, F. Is het weer aan zee anders dan in het binnenland? [Online; accessed 15 maart 2017].
- [13] ELIA. Elia > Projecten > Netprojecten > Noordzee. [Online; accessed 21 mei 2017].
- [14] GILBERT, R. O. *Statistical methods for environmental pollution monitoring*. John Wiley & Sons, 1987.
- [15] GOOD, P. I. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer series in statistics. Springer, 2000.
- [16] GOOD, P. I. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. Springer series in statistics. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2004.



- [17] GUO, H., XU, M., AND HU, Q. Changes in near-surface wind speed in China: 1969–2005. *International Journal of Climatology* 31, 3 (2011), 349–358.
- [18] HAMED, K. H. Trend detection in hydrologic data: the Mann–Kendall trend test under the scaling hypothesis. *Journal of hydrology* 349, 3 (2008), 350–363.
- [19] HISDAL, H., STAHL, K., TALLAKSEN, L. M., AND DEMUTH, S. Have streamflow droughts in Europe become more severe or frequent? *International Journal of Climatology* 21, 3 (2001), 317–333.
- [20] HOLTON, J. R., AND HAKIM, G. J. *An introduction to dynamic meteorology*, vol. 88. Academic press, 2012.
- [21] HUANG, Z., AND CHALABI, Z. Use of time-series analysis to model and forecast wind speed. *Journal of Wind Engineering and Industrial Aerodynamics* 56, 2-3 (1995), 311–322.
- [22] INDHUMATHY, D., SESHAIHAH, C., AND SUKKIRAMATHI, K. Estimation of Weibull Parameters for Wind speed calculation at Kanyakumari in India. *International Journal of Innovative Research in Science, Engineering and Technology* 3, 1 (2014), 8340–8345.
- [23] JAAGUS, J. Climatic changes in Estonia during the second half of the 20th century in relationship with changes in large-scale atmospheric circulation. *Theoretical and Applied Climatology* 83, 1 (2006), 77–88.
- [24] JUSTUS, C., HARGRAVES, W., MIKHAIL, A., AND GRABER, D. Methods for estimating wind speed frequency distributions. *Journal of applied meteorology* 17, 3 (1978), 350–353.
- [25] KENDALL, M. *Rank correlation methods*. Griffin, London, 1948.
- [26] KIDMO, D., DANWE, R., DOKA, S., AND DJONGYANG, N. Statistical analysis of wind speed distribution based on six Weibull Methods for wind power evaluation in Garoua, Cameroon. *Revue des Energies Renouvelables* 18, 1 (2015), 105–125.
- [27] KLINK, K. Climatological mean and interannual variance of United States surface wind speed, direction and velocity. *International Journal of Climatology* 19, 5 (1999), 471–488.
- [28] KLINK, K. Trends and interannual variability of wind speed distributions in Minnesota. *Journal of Climate* 15, 22 (2002), 3311–3317.
- [29] KMI. Documentatie > WeerWoorden > Beaufortschaal (windkracht). [Online; accessed 7 mei 2017].
- [30] KMI. Documentatie > WeerWoorden > Bewolkingsgraad. [Online; accessed 17 februari 2017].
- [31] KMI. Documentatie > WeerWoorden > Temperatuur. [Online; accessed 7 mei 2017].
- [32] KMI. Documentatie > WeerWoorden > Zeebries. [Online; accessed 11 april 2017].
- [33] KMI. Klimaat > Algemeen klimaat in België > Neerslag. [Online; accessed 8 maart 2017].
- [34] KUNDZEWICZ, Z., AND ROBSON, A. *Detecting trend and other changes in hydrological data*. World Meteorological Organization, 2000.
- [35] LEI, Y. Evaluation of three methods for estimating the Weibull distribution parameters of Chinese pine (*Pinus tabulaeformis*). *Journal of Forest Science* 54, 12 (2008), 566–571.
- [36] MANN, H. B. Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society* (1945), 245–259.
- [37] MCVICAR, T. R., RODERICK, M. L., DONOHUE, R. J., LI, L. T., VAN NIEL, T. G., THOMAS, A., GRIESER, J., JHAJHARIA, D., HIMRI, Y., MAHOWALD, N. M., ET AL. Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation. *Journal of Hydrology* 416 (2012), 182–205.

- [38] MENSINK, C. Luchtvervuiling en chemische transportmodellen. Ghent University postgraduate course notes, 2016.
- [39] MONTGOMERY, D., PECK, E., AND VINING, G. *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, 2012.
- [40] MOROZOVA, A. L., AND VALENTE, M. A. Homogenization of Portuguese long-term temperature data series: Lisbon, Coimbra and Porto. *Earth System Science Data* 4, 1 (2012), 187–213.
- [41] NETER, J., KUTNER, M. H., NACHTSHEIM, C. J., AND WASSERMAN, W. *Applied linear statistical models*, vol. 4. Irwin Chicago, 1996.
- [42] NWOBI, F. N., AND UGOMMA, C. A. A comparison of methods for the estimation of Weibull distribution parameters. *Metodoloski zvezki* 11, 1 (2014), 65.
- [43] ONYUTHA, C., TABARI, H., TAYE, M. T., NYANDWARO, G. N., AND WILLEMS, P. Analyses of rainfall trends in the Nile River Basin. *Journal of Hydro-Environment Research* (2015).
- [44] PACIFIC NORTHWEST NATIONAL LABORATORY, US DEPARTMENT OF ENERGY. *Mann-Kendall Test For Monotonic Trend, in Visual Sample Plan users' guide*. [Online; accessed: 14 februari 2017].
- [45] PETERSEN, E. L., MORTENSEN, N. G., LANDBERG, L., HØJSTRUP, J., AND FRANK, H. P. Wind power meteorology. Part I: Climate and turbulence. *Wind Energy* 1, S1 (1998), 25–45.
- [46] PETERSON, E. W., AND HENNESSEY JR, J. P. On the use of power laws for estimates of wind power potential. *Journal of Applied Meteorology* 17, 3 (1978), 390–394.
- [47] PIRAZZOLI, P. A., AND TOMASIN, A. Recent near-surface wind changes in the central Mediterranean and Adriatic areas. *International Journal of Climatology* 23, 8 (2003), 963–973.
- [48] POHLERT, T. **trend**: *Non-Parametric Trend Tests and Change-Point Detection*, 2016. R package version 0.2.0.
- [49] PRYOR, S., AND BARTHELMIE, R. Long-term trends in near-surface flow over the Baltic. *International Journal of Climatology* 23, 3 (2003), 271–289.
- [50] ROMANIĆ, D., ČURIĆ, M., JOVIČIĆ, I., AND LOMPAR, M. Long-term trends of the ‘Koshava’ wind during the period 1949–2010. *International Journal of Climatology* 35, 2 (2015), 288–302.
- [51] ROUSSEEUW, P. J., AND LEROY, A. M. *Robust regression and outlier detection*, vol. 589. John Wiley & Sons, 2005.
- [52] SEN, P. K. Estimates of the regression coefficient based on Kendall’s tau. *Journal of the American Statistical Association* 63, 324 (1968), 1379–1389.
- [53] ŞEN, Z., ALTUNKAYNAK, A., AND ERDIK, T. Wind velocity vertical extrapolation by extended power law. *Advances in Meteorology 2012* (2012).
- [54] SHU, Z., LI, Q., AND CHAN, P. Investigation of offshore wind energy potential in Hong Kong based on Weibull distribution function. *Applied Energy* 156 (2015), 362–373.
- [55] STOCKER, T., QIN, D., PLATTNER, G.-K., TIGNOR, M., ALLEN, S., BOSCHUNG, J., NAUELS, A., XIA, Y., BEX, V., AND MIDGLEY, P., Eds. *IPCC 2013: Summary for Policymakers*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013, book section SPM, p. 1–30.
- [56] VAN MALDEREN, R. Fysische meteorologie. Ghent University postgraduate course notes, 2016.
- [57] VAUTARD, R., CATTIAUX, J., YIOU, P., THÉPAUT, J.-N., AND CIAIS, P. Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness. *Nature Geoscience* 3, 11 (2010), 756–761.

- 
- [58] WERAPUN, W., TIRAWANICHAKUL, Y., AND WAEWSAK, J. Comparative study of five methods to estimate Weibull parameters for wind speed on Phangan Island, Thailand. *Energy Procedia* 79 (2015), 976–981.
- [59] WIKIPEDIA. Windstreek — Wikipedia, the free encyclopedia, 2016. [Online; accessed 25 oktober 2016].
- [60] WIKIPEDIA. Anemometer — Wikipedia, the free encyclopedia, 2017. [Online; accessed 18 mei 2017].
- [61] WIKIPEDIA. Global warming — Wikipedia, the free encyclopedia, 2017. [Online; accessed 2 mei 2017].
- [62] WIKIPEDIA. Log wind profile — Wikipedia, the free encyclopedia, 2017. [Online; accessed 30 mei 2017].
- [63] WIKIPEDIA. Nonparametric statistics — Wikipedia, the free encyclopedia, 2017. [Online; accessed 10 april 2017].
- [64] WIKIPEDIA. Rayleigh distribution — Wikipedia, the free encyclopedia, 2017. [Online; accessed 27 mei 2017].
- [65] WIKIPEDIA. Waloddi Weibull — Wikipedia, the free encyclopedia, 2017. [Online; accessed 23 mei 2017].
- [66] WINDPOWER PROGRAM. Wind turbine power output variation with steady wind speed. Technical information page for the UK Wind Speed Database program [Online; accessed 4 juni 2017].
- [67] YUE, S., PILON, P., AND CAVADIAS, G. Power of the Mann–Kendall and Spearman’s rho tests for detecting monotonic trends in hydrological series. *Journal of hydrology* 259, 1 (2002), 254–271.
- [68] ZEILEIS, A., AND HOTHORN, T. Diagnostic checking in regression relationships. *R News* 2, 3 (2002), 7–10.