

# Inhoudsopgave

<b>Summary .....</b>	<b>1</b>
<b>1 Inleiding.....</b>	<b>3</b>
1.1 Wind is meer dan wind alleen.....	3
1.2 Doelstelling .....	5
1.3 Opzet.....	5
1.4 Opmaak data .....	8
<b>2 Methodologie .....</b>	<b>11</b>
2.1 Lineaire regressie .....	11
2.1.1 Theorie achter lineaire regressie .....	11
2.1.2 Waarom lineaire regressie?.....	12
2.2 Dataproblemen.....	13
2.2.1 Heteroskedasticiteit.....	13
2.2.2 Autocorrelatie.....	14
2.2.3 Non-normaliteit.....	15
2.2.4 Multicollineariteit.....	16
<b>3 Analyse .....</b>	<b>17</b>
3.1 Inleiding .....	17
3.2 Toetsing en eventuele correctie dataproblemen.....	17
3.2.1 Heteroskedasticiteit.....	17
3.2.2 Autocorrelatie.....	22
3.2.3 Non-normaliteit.....	23
3.2.4 Multicollineariteit.....	24
3.3 Van back-up model naar meervoudig lineair regressie-model .....	25
3.3.1 Het back-up model.....	25
3.3.2 Het enkelvoudig lineaire regressiemodel.....	26
3.3.3 Het meervoudige lineaire regressiemodel.....	28
3.3.4 Nadere beschouwing volgorde van toelating .....	31
3.4 Het betrekken van lokale beschuttingsfactoren in de modelbouw.....	33
3.4.1 Waarom kan modellering van lokale beschuttingsfactoren van belang zijn? .....	33
3.4.2 Clusteren van windrichtingen .....	34
3.4.3 Windrichting modelleren met Fourier analyse.....	41
3.4.4 Mate van vermindering autocorrelatie .....	46
3.4.5 Gevolgen van beschuttingsverschillen tussen windmasten na correctie.....	47
3.5 Het opstellen van betrouwbaarheidsintervallen .....	52
<b>4 Conclusies.....</b>	<b>55</b>
<b>5 Suggesties .....</b>	<b>57</b>
<b>6 Aanbevelingen .....</b>	<b>59</b>

<b>Dankwoord .....</b>	<b>61</b>
<b>Referenties .....</b>	<b>63</b>
<b>Bijlagen .....</b>	<b>65</b>
Bijlage 1: Quotiëntplots van windsnelheden .....	66
Bijlage 2: Coëfficiëntplots van hellingshoek bij clusteren van windrichting .....	67
Bijlage 3: Coëfficiëntplots van hellingshoek bij Fourier modellering.....	68
Bijlage 4: Coëfficiëntplots van hellingshoek (combinatie bijlage 2 en 3).....	69

# Summary

Information about the wind speed and wind direction is of great importance for an airplane during take-off or touch-down. When pilots have this information at their disposal, they can use this to take off or touch down safely. In other cases can be decided to wait for a while before landing or taking off. When we speak about landing, in certain situations even can be decided to flight to another airport. For this reason windmasts are placed at airports nearby the several take-off and touch-down zones, so that the information required can be obtained.

This paper handles a statistical procedure that makes a prediction about the 2 minute average wind speed at a certain windmast. When this windmast temporally doesn't function (for example due to a defect or to maintenance) through which he isn't able to generate measurement data, the developed procedure can be used for approximating the wind speed at this windmast.

For this same goal, nowadays airports often make use of back-up schemes by which values of wind speed and wind direction that are measured at a certain windmast are used in substitution for the wind speed and wind direction at the windmast which is out of order. This method has three important disadvantages. Firstly, due to differences in terrain roughness, a substitution can be severely biased. Secondly, making use of merely one other windmast, this method doesn't take account for the capricious pattern that characterizes wind speed and wind direction and at last the back-up scheme doesn't say anything about the reliability of a substitution.

The procedure discussed in this paper is based on a regression model that measurements, obtained from more then one other potential windmast at the airport, uses to make a prediction about the wind speed at a windmast which is out of order. Hereby the capricious pattern of the wind speed has been smoothed out.

Furthermore the model deals with the several differences in terrain roughness at the airport. This has been done in two different manners. At one side by clustering to wind direction (discrete manner of modeling) and at the other side by making use of Fourier analysis (continued manner of modeling). Both methods create coefficients that are dependent of the wind direction. It has been shown that modeling with the help of Fourier analysis satisfies the most in favour of clustering to wind direction, by which is corrected fully for deviations in wind speeds between windmasts caused by differences in terrain roughness.

The procedure is also provided with an error margin that stands for the extent of the reliability of a certain prediction. This reliability appears to be strongly dependent of the state of the weather at the moment of making the prediction. The length of the error margin has been adjusted to this, so that the procedure takes into account phenomena as heavy thunderstorms but also calmer weathertypes with small changes in wind speed.

Because of modeling terrain roughness and giving error margins the procedure is useful for application during extreme weather conditions. This is an important advantage of the procedure in favour of operational back-up schemes, because of the great importance of adequate information about the wind speed during extreme weather conditions.

The latter is the main reason to advice replacing the operational back-up schemes at airports by the method discussed in this paper.



# 1 Inleiding

## 1.1 Wind is meer dan wind alleen

Iedereen heeft wel op de een of andere manier met het weer te maken. Of het nu beroepsmatig is of gewoon in ons dagelijks leven: niemand ontkomt aan de vele facetten die gepaard kunnen gaan met een bepaald soort weertype.

Kennis over de weersituatie kan van groot belang zijn bij het waarborgen van de veiligheid van mensen. Zo is dat ook het geval wanneer men spreekt over vliegverkeer. Bepaalde weersomstandigheden kunnen immers voor gevaarlijke situaties zorgen; denk hierbij aan mist of laaghangende bewolking met als gevolg beperkt zicht aan de grond, noodweer met zware regen of hagel met opnieuw als gevolg beperkt zicht en harde wind, die het moeilijk maakt om een vliegtuig onder controle te houden.

Het zal niet de eerste keer zijn dat een vliegtuig als gevolg van extreme weersomstandigheden neerstort in een zwaar bevolkt gebied met alle gevolgen van dien. Andere, kleinere ongelukken komen veelvuldig voor en dienen met alle mogelijke moeite voorkomen te worden.

Wanneer een vliegtuig gaat opstijgen of gaat landen zijn gevaarlijke weersituaties extra risicovol. Immers, de afstand van het vliegtuig tot het aardoppervlak is dan klein, waardoor eventuele noodzakelijke stuurcorrecties vaak niet meer tijdig kunnen worden uitgevoerd om desastreuze gevolgen te voorkomen.

De wind speelt bij dit alles een zeer belangrijke rol. Vooral bij hoge dwarswindsnelheid (het deel van de windsnelheid dat loodrecht staat op de vliegrichting) is het zeer moeilijk om een vliegtuig onder controle te houden. Dit effect zal variëren per type vliegtuig: een zware Boeing zal een bepaalde dwarswind beter trotseren dan een kleiner en lichter vliegtuigtype.

Bovenstaande maakt duidelijk dat het van onmisbaar belang is om op de hoogte te zijn van de actuele windsituatie (windsnelheid en windrichting) en met name van de dwarswinden bij de verschillende plaatsen van landen en opstijgen (de zogenaamde 'touch-down' en 'take-off' zones).

Voor een beeld van de toekomstige ontwikkeling van de windsituatie is bij de grotere luchthavens een meteorologische dienst aanwezig. Met behulp van de door deze dienst verstrekte informatie blijven de verkeersleiding en piloten op de hoogte van veranderingen in de windsituatie.

Om kennis te hebben van de actuele windsituatie zijn op de meeste luchthavens windmeetmasten geplaatst. Deze meten de windsnelheid en windrichting op tien meter van de grond. In Figuur 1-1 op de volgende pagina staat een dergelijke windmeetmast afgebeeld zoals die op Amsterdam Schiphol Airport wordt gebruikt.

Omdat bij eenzijdig baangebruik (d.w.z. een vliegtuig stijgt altijd op in één bepaalde richting terwijl hij landt in de tegenovergestelde richting) de touch-down zone en de take-off zone gedeeltelijk overlappend dan wel aansluitend zijn wordt er in het algemeen volstaan met het plaatsen van één windmast in de buurt van iedere touch-down zone. Deze metingen worden door de verkeersleiding gebruikt om te beslissen of het verantwoord is een vliegtuig te laten opstijgen dan wel te laten landen. Afhankelijk van de hoogte van de dwarswindsnelheid (op basis van de windsnelheid en windrichting) worden piloten geadviseerd een andere landingsbaan te gebruiken of nog even te

wachten met de landing. In extreme gevallen wordt er zelfs uitgeweken naar een andere luchthaven.

Bij het doen van windmetingen wordt wel eens tegen problemen aangelopen. Zo kan het gebeuren dat een windmeetmast door een bepaald defect of door het plegen van onderhoud niet functioneert. Er kunnen dan geen gegevens van de windmast verkregen worden en geen directe uitspraak over de actuele windsituatie bij de desbetreffende touch-down of take-off zone.

Om deze reden maken veel luchthavens gebruik van een back-up schema. Hierbij worden de metingen van een andere mast op de luchthaven (indien aanwezig) gebruikt als vervanging voor de metingen van de niet-functionerende windmast. Deze aanpak blijkt niet altijd geschikt te zijn vanwege de soms grote afstanden tussen de masten en omdat de omgevingen van de twee windmasten behoorlijk kunnen verschillen. Hierdoor zegt in bepaalde gevallen een meting gedaan bij de back-up mast niet zoveel over de windsnelheid bij de niet-functionerende windmast.

In dit onderzoek is bekeken of het mogelijk is om een betere uitspraak te doen over de windsnelheid bij een niet-functionerende windmast, gebruik makende van windgegevens afkomstig van de andere windmeetmasten op de luchthaven. Hierbij is onderzocht of modellering van lokale ruwheden de uitspraak kan verbeteren.



**Figuur 1-1:** Afbeelding van een windmeetmast zoals die gebruikt wordt te Schiphol.

## 1.2 Doelstelling

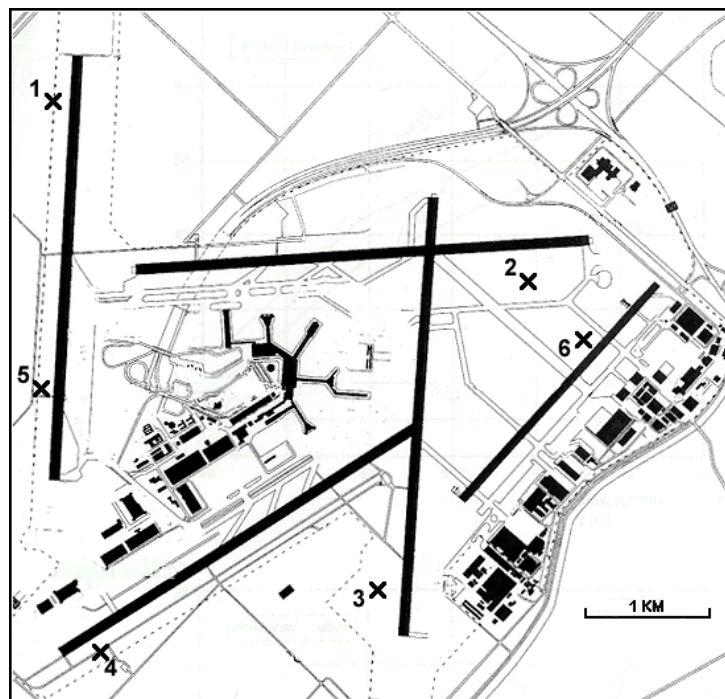
In dit onderzoek is gekeken naar methoden die een voorspelling kunnen doen over windsnelheden die optreden bij een bepaalde windmast op een luchthaven. Dit is gedaan door een antwoord te zoeken op de volgende vragen:

- Hoe en in welke mate kunnen gegevens van meerdere windmasten gebruikt worden om een uitspraak te verkrijgen over de windsnelheid van een niet-functionerende windmast?
- Hoe en in welke mate kan modellering van lokale ruwheden van het aardoppervlak de methoden verbeteren?
- Kunnen er weersituatie-afhankelijke foutenmarges worden aangebracht aan de uitspraken?
- In welke mate zijn de uitspraken ontwikkeld op basis van de modelbouw beter dan de huidige back-up procedures?

Bij deze vragen gaat de uitspraak over de windsnelheid steeds over de gemiddelde windsnelheid per twee minuten, aangezien deze waarde naast de grootste windstoot van de laatste tien minuten bij de meeste luchthavens gebruikt worden om beslissingen te nemen over het wel dan niet kunnen opstijgen of landen van een vliegtuig.

## 1.3 Opzet

Bij het analyseren van de in de vorige paragraaf gestelde vragen wordt gebruik gemaakt van windmeetgegevens afkomstig van Amsterdam Airport Schiphol. Schiphol heeft op dit moment de beschikking over een vijftal start- en landingsbanen. De ligging van deze banen staat aangegeven in Figuur 1-2.



**Figuur 1-2:** Plattegrond Amsterdam Airport Schiphol.

Omdat één van deze banen (de Zwanenburgbaan) dubbelzijdig wordt gebruikt zijn er bij deze baan twee touch-down zones; in totaal zijn er op Schiphol zes touch-down zones. Om deze reden zijn er zes windmasten geplaatst: bij elke touch-down zone één. De officiële benamingen van deze masten zijn resp. 19R, 27, 01R, 06, 01L en 22. De masten zullen voor de eenvoud in het vervolg benoemd worden met resp. 1, 2, 3, 4, 5 en 6. De situering van deze masten staat eveneens aangegeven in Figuur 1-2.

De windmasten staan vanwege veiligheidsvoorschriften niet al te dicht bij de landingsbanen opgesteld. De afstand van een windmast tot het hart van de baan is normaal gesproken circa 110 meter (bij metalen masten is deze afstand gelijk aan 200 meter). Het plaatsen en beheren van de masten wordt verzorgd door het Koninklijk Nederland Meteorologisch Instituut (KNMI). Het KNMI is namelijk verantwoordelijk voor het verstrekken van windgegevens aan de verkeersleiding van Schiphol.

De windmasten op Schiphol meten de windsnelheid in meters per seconde en de windrichting in graden per tijdsinterval van  $\frac{1}{4}$  seconde op een hoogte van 10 meter (conform internationale regelgeving). Vervolgens worden van de laatste 12 metingen (tijdsinterval 3 seconden) de gemiddelden bepaald. Op basis van deze 3-seconden-gemiddelden worden een aantal variabelen in een bestand opgeslagen. Elk bestand bevat een aantal windvariabelen van één etmaal (in Nederland tijdens zomertijd betekent dit van 02-02 uur en tijdens wintertijd van 01-01 uur) en één mast. Hieronder volgt een overzicht van de verschillende variabelen:

- Windsnelheid:
  - 1) *FF*: maximale waarde van alle 48 ( $12 \times 4$ ) 3-seconden-gemiddelden in de afgelopen 12 seconden,
  - 2) *FI*: rekenkundig gemiddelde van alle niet-overlappende 3-seconden-gemiddelden in de afgelopen minuut,
  - 3) *FA*: rekenkundig gemiddelde van alle niet-overlappende 3-seconden-gemiddelden in de afgelopen 10 minuten,
  - 4) *FX*: maximale waarde van alle 2400 ( $48 \times 5 \times 10$ ) 3-seconden-gemiddelden in de afgelopen 10 minuten (gust-waarden),
  - 5) *F<sub>s</sub>F*: standaardafwijking van alle niet-overlappende 3-seconden-gemiddelden in de afgelopen 10 minuten.
  
- Windrichting:
  - 1) *DD*: gemiddelde waarde in de afgelopen 12 seconden,
  - 2) *DI*: vectorieel gemiddelde waarde van de 5 12-seconden-gemiddelden in de afgelopen minuut,
  - 3) *DX*: meest gekrompen windrichting (d.w.z. tegen de klok in), extreem uit afgelopen 10 minuten,
  - 4) *DN*: meest geruimde windrichting (d.w.z. met de klok mee), extreem uit afgelopen 10 minuten,
  - 5) *D<sub>s</sub>D*: standaardafwijking van alle 12-seconden-gemiddelden in de laatste 10 minuten.

Om de 12 seconden wordt de data van deze 10 variabelen in het bestand opgeslagen, zodat in één etmaal 7200 keer ( $5 \times 60 \times 24$ ) de 10 variabelen worden verwerkt.

De windrichting wordt weergegeven in graden. Hierbij staat een windrichting van 0 graden voor de windrichting noord en 90 graden voor oost. Zo kan met de klok

meegegaan worden tot de cyclus is voltooid en uitgekomen wordt op een windrichting van 360 graden die gelijk is aan een windrichting van 0 graden en dus weer aan noord.

De verkeersleiding van Schiphol heeft zelf ook meetgegevens van iedere windmast tot haar beschikking en deze wordt getoond op een digitale display. Deze zijn o.a. als volgt:

- ◆ Windsnelheid:
  - $F2$ : rekenkundig gemiddelde van alle niet-overlappende 3-seconden-gemiddelden in de afgelopen 2 minuten,
  - $FX$ : zie boven.
- ◆ Windrichting:
  - $D2$ : vectorieel gemiddelde van alle 10 12-seconden-gemiddelden in de afgelopen 2 minuten.

Deze data worden op het display getoond en elke minuut aangepast.

Zoals eerder vermeld gebruikt de verkeersleiding de  $F2$ - en  $FX$ -waarden om besluiten te nemen over het al dan niet tijdelijk afsluiten van een landingsbaan. Het is daarom zinvol om modellen op te zetten die over deze twee grootheden een uitspraak doen. In dit onderzoek is gekeken naar modellen voor de  $F2$ -waarden. Hierbij kan gebruik gemaakt worden van o.a. bovenstaande variabelen.

De methoden die in dit onderzoek zijn toegepast berusten op regressiemodellen. Hierbij wordt getracht betrouwbare uitspraken te genereren over de gemiddelde windsnelheid per twee minuten ( $F2$ ) bij een bepaalde windmast. Omdat Schiphol over zes windmasten beschikt kan er een zestal optimale modellen worden opgesteld: voor iedere windmast één.

Voordat de modellen kunnen worden opgesteld dienen de data op de juiste manier opgemaakt worden. Dit is terug te vinden in de volgende paragraaf.

Omdat lineaire regressie in dit onderzoek een belangrijke rol speelt, zal in Hoofdstuk 2 de theorie hierachter in het kort uitgelegd worden alsmede waarom hiervoor gekozen is. Ook wordt in dit hoofdstuk een aantal dataproblemen besproken die gepaard kunnen gaan met lineaire regressie.

Hoofdstuk 3 omslaat de eigenlijke analyse. Allereerst wordt er getoetst op de dataproblemen en vervolgens begonnen met de modelbouw. Hierbij wordt als uitgangspunt de huidige methode van back-up beschouwd en wordt dit uitgebreid door gebruik te maken van enkelvoudige lineaire regressie en vervolgens meervoudige lineaire regressie door meer dan één windmast in de modelbouw te betrekken.

Doordat de omgeving (bebouwing/begroeiing) van een object een rol speelt in de aldaar optredende windsnelheid zal in Hoofdstuk 3 ook gekeken worden naar het effect van de windrichting in de modellen. Hierbij zullen verschillende methoden besproken worden om dit effect te modelleren.

Tenslotte wordt in dit hoofdstuk besproken hoe betrouwbaarheidsintervallen rondom een bepaalde uitspraak kunnen worden aangegeven.

In Hoofdstuk 4 staan de conclusies uit dit onderzoek beschreven, in Hoofdstuk 5 suggesties voor eventueel vervolgonderzoek en in Hoofdstuk 6 aanbevelingen ten behoeve van de implementatie van de besproken procedure.

## 1.4 Opmaak data

In de vorige paragraaf werd vermeld dat de windmasten niet al te dicht bij de landingsbanen geplaatst zijn vanwege strikte veiligheidsnormen. Ondanks deze betrekkelijk ruime afstanden hebben passerende vliegtuigen in een aantal gevallen een te groot effect op de windmetingen (met name op de *FX*-waarden) door het veroorzaken van bepaalde luchtwervelingen, ook wel vortices genoemd. Bij twee windmastlocaties (1 en 4) is dit effect dermate ernstig dat de gegevens afkomstig van deze masten in bepaalde situaties te veel afwijken van de werkelijke windsituatie. Om deze reden wordt sinds enige tijd een algoritme toegepast (zie van der Meulen [1998]) waarbij de metingen bij de desbetreffende windmasten meteen worden gecorrigeerd en vervolgens worden opgeslagen. Dit gebeurt zodanig dat in de opgeslagen data nauwelijks nog invloed merkbaar is van de opgetreden vortices. Het algoritme is voor windmast 1 operationeel vanaf 3 juli 1997 en voor windmast 4 vanaf 5 september 1997. Voor dit onderzoek wordt daarom gebruik gemaakt van data die lopen vanaf 5 september 1997. Op 19 oktober 1998 is windmast 3 verplaatst naar de andere kant van de landingsbaan. Data na 19 oktober 1998 worden daarom niet meegenomen in dit onderzoek.

De aangeleverde data zijn in de vorm van 12-seconden data. Hierdoor zijn bijvoorbeeld de *F1*-waarden niet disjunct, omdat deze een groter interval (één minuut) bestrijken. Ze overlappen elkaar dus voor een gedeelte. Evenzo is dit het geval bij de overige variabelen. Bij het opstellen van regressiemodellen is het echter wel beter te beschikken over disjuncte data. De data zijn daarom op dusdanige wijze gefilterd dat er disjuncte, maar aaneengesloten *F1*-waarden overblijven. Uit de disjuncte *F1*-waarden zijn vervolgens disjuncte *F2*-waarden berekend.

Om te voorkomen dat effecten van bepaalde artefacten doorwerken in de modellen voor alle masten, is gekozen om bij het bouwen van de modellen niet identieke data te gebruiken. De procedure die hierbij gevolgd is staat aangegeven in Tabel 1-1.

afhankelijke windmast	metingen
1	1,7,13,...
2	2,8,14,...
3	3,9,15,...
4	4,10,16,...
5	5,11,17,...
6	6,12,18,...

**Tabel 1-1:** Selectieprocedure van de metingen voor de verschillende modellen.

Uitleg Tabel 1-1: Voor het model waarbij een uitspraak gegenereerd wordt voor windmast 1 worden de metingen 1,7,13,enz. van alle masten genomen. Per twaalf minuten wordt er dan dus één meting per variabele gebruikt. Deze procedure geldt evenzo voor de andere modellen.

Verder is de data opgesplitst in twee onafhankelijke delen: een deel voor modelbouw en het andere deel voor validatie van de modeloutput. Hierbij behoren metingen met een even dagnummer tot de dataset voor modelbouw en metingen met een oneven dagnummer tot de dataset voor validatie. Deze twee delen zullen resp. de modeldata en validatiedata genoemd worden.

Bij de analyse zal een optimaal model opgesteld worden waarbij een uitspraak wordt gegenereerd over de gemiddelde 2-minuut windsnelheid bij windmast 1. De overige modellen behorende bij de andere masten (2 t/m 6) kunnen met behulp van dezelfde procedure worden opgesteld.

Bij het opstellen van deze modellen is besloten geen gebruik te maken van zeer lage windsnelheden aangezien lage windsnelheden moeilijk te interpreteren zijn en aangezien lage windsnelheden door technische oorzaken minder nauwkeurig gemeten kunnen worden. Om deze reden worden in dit onderzoek windsnelheden van minder dan 2 meter per seconde niet gebruikt bij het opstellen van de regressievergelijkingen.



## 2 Methodologie

### 2.1 Lineaire regressie

In dit onderzoek is het lineaire regressiemodel de grondslag. Daarom zal in het kort de theorie van lineaire regressie uitgelegd worden en zal worden ingegaan op de vraag waarom lineaire regressie in dit onderzoek is gebruikt.

#### 2.1.1 Theorie achter lineaire regressie

Wanneer men verbanden tussen twee grootheden  $x$  en  $y$  wil onderzoeken, waarbij  $x$  als verklarende variabele kan worden beschouwd en  $y$  als afhankelijke variabele, kan men dit verband goed vastleggen in een lineair regressiemodel als het verband rechtlijnig van aard is. Dit kan men makkelijk nagaan door een spreidingsdiagram te maken. Ook wanneer er bijvoorbeeld een kwadratisch verband is dan kan men lineaire regressie toepassen, waarbij dan een kwadratische term als extra factor wordt toegevoegd aan het model.

Ook kan men eventuele niet-lineaire verbanden terug brengen tot een lineair verband door het uitvoeren van een of meerdere transformaties (logaritme, wortel, etc.) op de variabelen.

Het model ziet er bij enkelvoudige lineaire regressie als volgt uit:

$$y = \alpha + \beta x + \varepsilon, \quad (2-1)$$

met  $\alpha$  de constante,  $\beta$  de hellingshoek en  $\varepsilon$  de term voor de storingen. Deze storingen zijn onafhankelijk van elkaar en normaal verdeeld met verwachting 0 en variantie  $\sigma^2$ . De parameters worden geschat met de zogenaamde 'gewone kleinste kwadraten' methode (Eng. Ordinary Least Squares (OLS)). Hierbij wordt

$$Q(c, d) = \sum_{i=1}^n [y_i - (c + dx_i)]^2 \quad (2-2)$$

geminimaliseerd met  $n$  het aantal waarnemingen en  $c$  en  $d$  variabel. De optimale waarden voor  $c$  en  $d$  waarbij (2.2) minimaal is worden berekend door de eerste orde condities op te lossen:

$$\partial Q / \partial c = 0, \quad (2-3)$$

$$\partial Q / \partial d = 0. \quad (2-4)$$

Uit deze twee condities volgen de optimale waarden van  $c$  en  $d$  (resp.  $a$  en  $b$ ):

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (2-5)$$

$$a = \bar{y} - b\bar{x}. \quad (2-6)$$

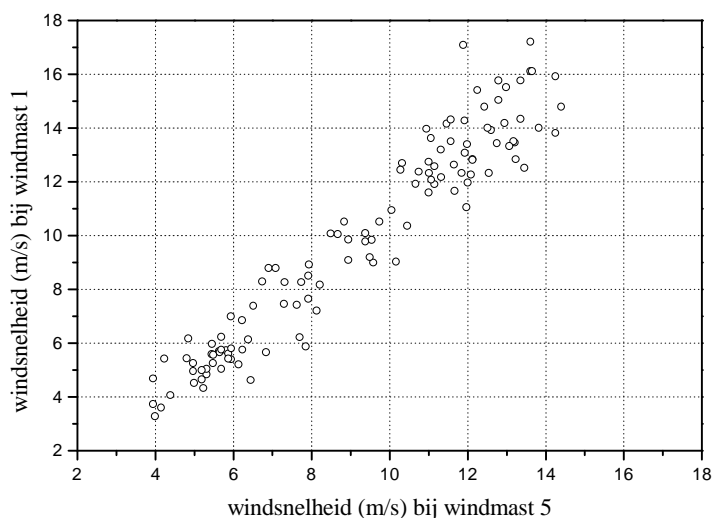
Deze waarden ( $a$  en  $b$ ) zijn de OLS schattingen van resp.  $\alpha$  en  $\beta$  en hiermee kan de regressielijn van  $y$  op  $x$  opgesteld worden:

$$\hat{y} = a + bx. \quad (2-7)$$

Dit model kan vrij eenvoudig uitgebreid worden tot een meervoudig lineair regressiemodel wanneer de grootheid  $y$  afhangt van meerdere verklarende variabelen. Dit standaard meervoudig lineair regressiemodel kan verder stelselmatig uitgebreid en aangepast worden totdat het meest geschikte model is gebouwd. In het verloop van dit onderzoek zal dit gebeuren.

### 2.1.2 Waarom lineaire regressie?

In dit onderzoek wordt veelal gebruik gemaakt van lineaire regressiemodellen om een uitspraak te genereren over de windsnelheid bij een niet-functionerende windmast. Hiervoor is gekozen, omdat als logische verklarende variabele de windsnelheid van een andere windmast gebruikt kan worden welke lineair afhangt van de windsnelheid van de niet-functionerende windmast zoals te zien is in Figuur 2-1. Hierin staan de windsnelheden gemeten bij windmast 1 en 5 op 26 december 1997 tegen elkaar uitgezet. De “ruis” die in dit figuur waarneembaar is zal in dit onderzoek worden getracht te verminderen door het aanpassen van het regressiemodel.



**Figuur 2-1:** Scatterplot van windsnelheid bij windmast 1 tegen windsnelheid bij windmast 5.

Op veel luchthavens wordt op dit moment al gebruik gemaakt van lineaire regressiemodellen, maar dan van zeer eenvoudige. Immers, bij een back-up schema worden waarden die bij een bepaalde mast gemeten worden gebruikt als vervanging voor de niet-functionerende windmast (zie Paragraaf 1.1). Dit kan beschouwd worden als een enkelvoudig lineair regressiemodel dat in zijn geheel is vastgelegd door het opleggen van twee restricties. De waarde van  $a$  is hierbij op 0 is gesteld en de waarde

van  $b$  op 1. Of dit de beste methode is om uitspraken te genereren is één van de punten die in dit onderzoek aan de orde komen.

## 2.2 Dataproblemen

Lineaire regressie kan niet zonder meer toegepast worden. De data kunnen namelijk bepaalde eigenschappen hebben waardoor OLS resultaten een vertekend beeld geven. Eigenschappen waarbij dit het geval is en die in het vervolg van dit onderzoek veelvuldig aan de orde komen zijn heteroskedasticiteit en autocorrelatie. Een andere eigenschap die vertekende OLS resultaten kan opleveren is non-normaliteit. Tenslotte kan bij meervoudige lineaire regressie sprake zijn van multicollineariteit. Dit levert geen onjuiste regressieresultaten, maar zorgt er wel voor dat de coëfficiënten van individuele verklarende variabelen minder goed zijn te interpreteren.

Deze vier eigenschappen worden hieronder besproken en er wordt ingegaan op de vragen of en hoe hiervoor gecorrigeerd kan worden.

### 2.2.1 Heteroskedasticiteit

Het verschijnsel waarbij de variantie van de storingsterm niet constant is, maar varieert over de waarnemingen wordt ook wel heteroskedasticiteit genoemd (zie Greene [1997, pp. 540-576]). In formulevorm:

$$\text{Var}(\varepsilon_t) = \sigma_t^2, \quad t = 1, 2, \dots, T, \quad (2-8)$$

waarbij

$$\sigma_t^2 \neq \sigma_s^2, \quad t \neq s, \quad (2-9)$$

en waarbij  $T$  het totaal aantal tijdmetingen is.

Vaak is de vorm van heteroskedasticiteit dusdanig dat de variantie van de storingsterm afhankelijk is van één van de verklarende variabelen in het model of van een functie van meerdere verklarende variabelen.

Het aanwezig zijn van heteroskedasticiteit zorgt ervoor dat bij OLS aan bepaalde waarden te veel gewicht wordt toegekend. Hierdoor zijn de schattingen van de coëfficiënten die volgen uit OLS nog wel zuiver, maar niet meer efficiënt (d.w.z. grotere standaardfouten). Verder worden de standaardfouten van de coëfficiënten en de daaraan gekoppelde  $t$ -waarden en  $P$ -waarden verkeerd geschat. Ook kunnen geen conclusies meer verbonden worden aan uitkomsten van  $F$ -toetsen.

De standaardfouten kunnen op de juiste manier worden geschat door gebruik te maken van White's schatter (zie White [1980b]). Deze schatter ziet er in matrixnotatie als volgt uit:

$$\text{geschatte Cov}(\mathbf{b}) = \frac{T}{T-k} (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_{t=1}^T e_t^2 \mathbf{x}_t \mathbf{x}_t' \right) (\mathbf{X}'\mathbf{X})^{-1}, \quad (2-10)$$

waar  $\mathbf{b}$  staat voor de vector met de coëfficiënten (inclusief constante) uit het model,  $k$  voor het aantal parameters in het model,  $\mathbf{X}$  voor de matrix met in de kolommen de

verklarende variabelen (inclusief constante) en in de rijen de verschillende tijdmetingen,  $e_t$  voor de modelfout (residu, realisatie van  $\varepsilon_t$ ) op tijdstip  $t$  en  $\mathbf{x}_t$  voor de vector met verklarende variabelen op tijdstip  $t$ .

De standaardfouten van de coëfficiënten kunnen vervolgens berekend worden door de wortels van de diagonaalelementen van (2-10) te nemen. Deze standaardfouten zijn een juiste afspiegeling van de echte standaardfouten, zodat conclusies op basis van de bijbehorende  $t$ - en  $P$ -waarden geen vertekend beeld geven. Het econometrisch softwarepakket EViews is voorzien van White's schatter.

Verder kan ook getracht worden de coëfficiënten efficiënter te maken door een techniek toe te passen die corrigeert voor de aanwezige heteroskedasticiteit. Hierbij worden alle variabelen (afhankelijke en verklarende) vermenigvuldigd met een gewicht, zodanig dat waarden met een hoge residuele variantie minder zwaar meetellen in de regressie. Deze methode wordt WLS ('Weighted Least Squares') genoemd (zie Greene [1997, pp. 555-556]). WLS resulteert in efficiënte coëfficiënten.

Omdat echter vaak het precieze gewicht niet bekend is, zal deze waarde geschat moeten worden. In veel gevallen kan hiervoor een functie op de verklarende variabelen gebruikt worden, welke ook met OLS kan worden geschat. Vervolgens kan WLS toegepast worden met de geschatte waarden van de gewichten. Dit wordt FGLS ('Feasible Generalized Least Squares') genoemd (zie Greene [1997, pp. 511-517 en 558-567]).

Wanneer WLS of eventueel FGLS wordt toegepast is het niet meer nodig White's schatter te gebruiken bij het schatten van de covariantiematrix. Er is dan immers gecorrigeerd voor de aanwezige heteroskedasticiteit.

In Paragraaf 3.2.1 zal onderzocht worden in hoeverre er hier sprake is van heteroskedasticiteit. Tevens wordt bekeken of en hoe hiervoor gecorrigeerd kan worden.

### 2.2.2 Autocorrelatie

Naast heteroskedasticiteit kan een andere residuele eigenschap ook voor vertekeningen zorgen: autocorrelatie (zie Greene [1997, pp. 577-611]). Dit is het verschijnsel waarbij de storingsterm onderling niet onafhankelijk, maar gecorreleerd is. In formulevorm:

$$\text{Corr}[\varepsilon_t, \varepsilon_{t-k}] = \rho_k, \quad k = 1, 2, 3, \dots, \quad (2-11)$$

waarbij

$$\rho_k \neq 0. \quad (2-12)$$

Wanneer (2-11) een waarde van +1 heeft duidt dit op perfecte positieve correlatie, een waarde van -1 duidt op perfecte negatieve correlatie en een waarde van 0 duidt op totaal geen correlatie.

De gevolgen van autocorrelatie zijn dezelfde als bij heteroskedasticiteit. De berekende coëfficiënten zijn nog wel zuiver, maar niet efficiënt. De standaardfouten van de coëfficiënten worden verkeerd geschat en dus ook de daaraan gekoppelde  $t$ -waarden en  $P$ -waarden. Tenslotte kunnen geen conclusies verbonden worden aan de  $F$ -toetsen.

In welke mate de standaardfouten verkeerd worden geschat en in welke richting, hangt af van de mate en van de soort van de aanwezige autocorrelatie. Hoe sterker de autocorrelatie des te meer wijken de standaardfouten van de coëfficiënten af van de werkelijke standaardfouten. Wanneer de autocorrelatie positief is (wat meestal het geval is bij tijdreeksen) zullen de standaardfouten te laag worden geschat.

De standaardfouten kunnen echter op een alternatieve wijze berekend worden. Zoals White's schatter rekening houdt met de aanwezige heteroskedasticiteit, zo doet de Newey-West schatter (zie Newey en West [1987a]) dat voor heteroskedasticiteit én voor autocorrelatie. In matrixnotatie ziet deze schatter er als volgt uit:

$$\text{geschatte Cov}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{\Omega} (\mathbf{X}'\mathbf{X})^{-1}, \quad (2-13)$$

waarbij

$$\mathbf{\Omega} = \sum_{t=1}^T e_t^2 \mathbf{x}_t \mathbf{x}_t' + \sum_{v=1}^q \left[1 - \frac{v}{q+1}\right] \sum_{t=v+1}^T (\mathbf{x}_t e_t e_{t-v} \mathbf{x}_{t-v}' + \mathbf{x}_{t-v} e_{t-v} e_t \mathbf{x}_t'). \quad (2-14)$$

In (2-14) staat  $q$  voor een parameter die staat voor het aantal waarnemingen in de tijd terug dat wordt gebruikt om de dynamica in  $e_t$  te modelleren.

De standaardfouten van de coëfficiënten kunnen vervolgens, zoals bij White's schatter, berekend worden door de wortels van de diagonaalelementen van (2-14) te nemen. Deze standaardfouten zijn opnieuw een juiste afspiegeling van de echte standaardfouten, zodat conclusies op basis van de bijbehorende  $t$ - en  $P$ -waarden geen vertekend beeld geven. Het econometrisch softwarepakket EViews is voorzien van de Newey-West schatter.

### 2.2.3 Non-normaliteit

Veel toetsen die gebaseerd zijn op de regressieresultaten zijn gebaseerd op de aanname dat de storingsterm normaal verdeeld is, met verwachting 0 en constante variantie. In formule vorm:

$$\varepsilon_t \sim N(0, \sigma^2), \quad t = 1, 2, \dots, T. \quad (2-15)$$

De variantie is in het geval van heteroskedasticiteit niet constant. Middelen om hiervoor te corrigeren zijn besproken in Deelparagraaf 2.2.1.

Wanneer de storingsterm niet uit de normale verdeling komt kunnen er geen conclusies worden getrokken uit de verschillende  $t$ -waarden van de coëfficiënten en dus ook niet uit de  $P$ -waarden. Deze waarden zijn namelijk gebaseerd op de veronderstelling van normaliteit.

Wanneer de storingsterm wel uit de normale verdeling komt kan op basis van de residuele standaardfout een betrouwbaarheidsinterval voor een bepaalde voorspelling

$$\hat{y}_h, \quad h = T, T+1, \dots \quad (2-16)$$

geconstrueerd worden en wel op de volgende manier (zie Greene [1997, pp. 369]):

$$P\left(\hat{y}_h - t_{\alpha/2}(v) \cdot s(\hat{y}_h) \leq y \leq \hat{y}_h + t_{\alpha/2}(v) \cdot s(\hat{y}_h)\right) = 1 - \alpha, \quad (2-17)$$

waarbij

$$s(\hat{y}_h) = \sqrt{s^2 + \mathbf{x}_h' s^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_h}, \quad (2-18)$$

en waarbij  $1-\alpha$  staat voor het betrouwbaarheidsniveau en  $t_{\alpha/2}(v)$  voor het  $100 \cdot (1-\alpha/2)\%$ -percentiel van de Student's  $t$  verdeling bij  $v$  vrijheidsgraden. In (2-18) staat  $s$  voor de residuele standaardfout. Doorgaans wordt het betrouwbaarheidsniveau op 95% gesteld door  $\alpha$  een waarde van 0.05 mee te geven. Op deze manier ontstaat een 95%-betrouwbaarheidsinterval waarvan gesteld kan worden dat het voor 95% zeker is dat de werkelijke waarde  $y$  zich in dit interval bevindt.

#### 2.2.4 Multicollineariteit

Wanneer de verklarende variabelen in het meervoudige regressiemodel sterk gecorreleerd zijn met elkaar is er sprake van multicollineariteit (zie Greene [1997, pp. 418-427]).

Multicollineariteit impliceert geen vertekeningen maar zorgt ervoor dat coëfficiënten relatief klein zijn en hoge standaardfouten hebben, waardoor individuele coëfficiënten minder goed zijn te interpreteren. In sommige gevallen kan het teken van de coëfficiënt zelfs tegenovergesteld zijn in vergelijking met het verwachte teken (+ wordt – of – wordt +).

Dit wordt veroorzaakt doordat een groot deel van de verklarende kracht die individuele verklarende variabelen hebben kunnen overeenkomen met elkaar, waardoor wanneer de verklarende variabelen gezamenlijk worden opgenomen in het regressiemodel deze verklarende kracht wordt opgedeeld tussen de verklarende variabelen onderling. Het gevolg hiervan is dat de verklarende kracht per verklarende variabele in het meervoudige regressiemodel een stuk kleiner is dan het geval is bij de verschillende enkelvoudige regressiemodellen. Het kan hierdoor gebeuren dat een bepaalde verklarende variabele geen significante coëfficiënt heeft. Zijn toegevoegde waarde in het model is dan gering terwijl zijn individuele kracht toch behoorlijk kan zijn.

Het belangrijkste gevolg van multicollineariteit is dus dat de coëfficiënten in het meervoudige regressiemodel relatief klein zijn en de bijbehorende standaardfouten relatief groot, zodat de individuele effecten van de verklarende variabelen minder goed zijn te interpreteren. Wanneer bij de modelbouw echter het doel is om de totale voorspelkracht te maximaliseren dan is het niet erg dat de individuele effecten minder goed zijn te interpreteren. De interesse gaat dan immers niet uit naar de individuele effecten. Wel kan eventueel de toegevoegde waarde van een bepaalde verklarende variabele worden getoetst.

# 3 Analyse

## 3.1 Inleiding

In dit hoofdstuk zal in Paragraaf 3.2 allereerst onderzocht worden of de dataproblemen die in Paragraaf 2.2 zijn behandeld daadwerkelijk optreden bij de modellen die in dit onderzoek aan de orde komen. Tevens wordt bekeken of hiervoor gecorrigeerd kan en/of dient te worden.

Hierna kan aan de werkelijke modelbouw begonnen worden waarbij in Paragraaf 3.3 de stap wordt gezet van de back-up methode naar het enkelvoudige regressiemodel en vervolgens het enkelvoudige regressiemodel wordt uitgebreid tot het meervoudige regressiemodel door het opnemen van meer dan één mast in het model.

In Paragraaf 3.4 wordt dit meervoudige regressiemodel uitgebreid door de windrichting als extra factor mee te nemen in het model. Dit wordt gedaan op twee manieren. Allereerst wordt er geclusterd naar windrichting en de tweede manier betreft de windrichting in het model door gebruik te maken van Fourier analyse.

## 3.2 Toetsing en eventuele correctie dataproblemen

De dataproblemen die in Paragraaf 2.2 aan de orde zijn geweest zullen in deze paragraaf getoetst worden met behulp van een meervoudig regressiemodel dat in het verdere verloop van dit onderzoek vaker terug zal komen in aangepaste vorm, gecorrigeerd voor de dataproblemen. Dit model doet een uitspraak over de gemiddelde 2-minuut windsnelheid bij windmast 1 aan de hand van de gemiddelde 2-minuut windsnelheden gemeten bij de overige vijf windmasten (2 t/m 6). Dit model ziet er als volgt uit:

$$F2_{1,t} = \alpha + \sum_{i=2}^6 \beta_i F2_{i,t} + \varepsilon_t, \quad (3-1)$$

waarbij  $F2_{i,t}$  staat voor de gemiddelde windsnelheid bij windmast  $i$  over de laatste twee minuten gezien vanaf tijdstip  $t$ .

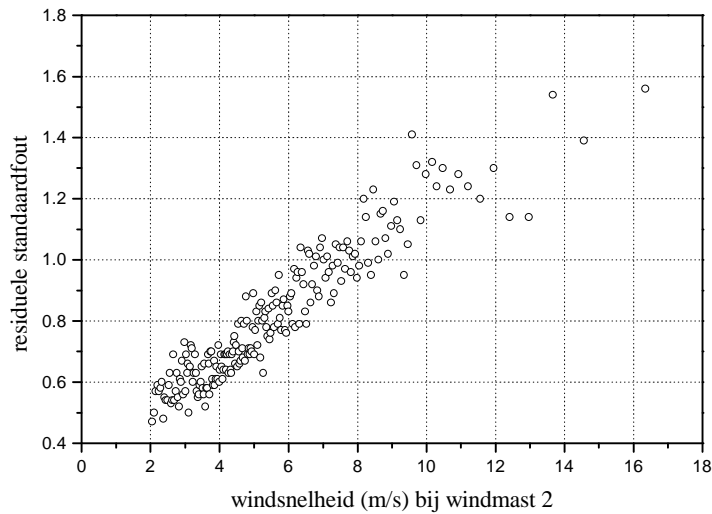
Als blijkt dat er sprake is van één of meerdere van de besproken dataproblemen zal indien mogelijk hiervoor gecorrigeerd worden.

### 3.2.1 Heteroskedasticiteit

Bij model (3-1) bestaat op basis van fysische gronden het vermoeden dat er sprake is van heteroskedasticiteit. Immers, naarmate de wind hogere snelheden bereikt worden onderlinge verschillen tussen de windsnelheden bij de verschillende windmasten in absolute zin vaak groter worden wat zich mogelijk uit in een grotere residuele variantie bij hogere windsnelheden.

Dit is onderzocht door de uit model (3-1) voortgekomen residuen te sorteren naar windsnelheid bij windmast 2 van klein naar groot. In principe zou hetzelfde kunnen worden gedaan met de windsnelheden bij de windmasten 3 t/m 6, maar windmast 2 is als voorbeeld genomen. Vervolgens is per 100 metingen de wortel van de gemiddelde

kwadratische waarde van de residuen (residuele standaardfout) berekend. Figuur 3-1 laat het resultaat zien en laat eveneens het vermoeden bevestigen.



**Figuur 3-1:** Scatterplot windsnelheid bij windmast 2 tegen de gemiddelde residuele standaardfout uit model (3-1).

De trend die in Figuur 3-1 zichtbaar is, is bij benadering lineair. De residuele variantie lijkt dus niet constant, maar lineair afhankelijk van het kwadraat van de windsnelheid.

Een formelere methode voor de detectie van heteroskedasticiteit is de Goldfeld-Quandt test (zie Greene [1997, pp. 551-552]). Hierbij worden alle metingen gesorteerd naar windsnelheid bij windmast 2 van klein naar groot, waarna de datareeksen in twee delen (aantal metingen per deel resp.  $n_1$  en  $n_2$ ,  $n_1 + n_2 = T$ ) worden gesplitst. Vervolgens wordt model (3.1) op beide delen geschat met OLS. De residuele varianties in beide delen kunnen met elkaar worden vergeleken met een  $F$ -toets:

$$F[n_2 - k, n_1 - k] = \frac{\sum_{j=1}^{n_1} e_{2,j}^2 / (n_2 - k)}{\sum_{j=n_1+1}^T e_{1,j}^2 / (n_1 - k)}, \quad (3-2)$$

waarbij  $k$  gelijk is aan het aantal parameters in model (3-1). Onder de nulhypothese van geen heteroskedasticiteit is (3-2)  $F$  verdeeld met  $n_2 - k$  en  $n_1 - k$  vrijheidsgraden. De beide delen zijn hier even groot gekozen waarna de  $F$ -toets de volgende waarde krijgt:

$$F[10900 - 6, 10900 - 6] = \frac{10000.620 / (10900 - 6)}{4337.015 / (10900 - 6)} = 2.31.$$

De kritieke  $F$ -waarde bij een betrouwbaarheidsniveau van 95% is 1.03. De nulhypothese kan dus stellig verworpen worden, zodat er goede redenen zijn om aan te nemen dat er sprake is van heteroskedasticiteit.

Nu bekend is dat er sprake blijkt te zijn van heteroskedasticiteit is het verstandig om de in Deelparagraaf 2.2.1 besproken methoden toe te passen op de nader te analyseren regressiemodellen om vertekeningen te voorkomen of om efficiëntere coëfficiënten te verkrijgen.

Het voorkomen van vertekeningen kan gedaan worden door gebruik te maken van White's schatter. Onderstaande tabel geeft aan in hoeverre de standaardfouten in model (3-1) die via de standaardmethode worden geschat afwijken ten opzichte van de standaardfouten die worden geschat met behulp van de schatter van White:

parameter	coëfficiënt	standaardfout coëfficiënt	
		OLS	White
$\alpha$	-0.027	$1.42 \cdot 10^{-2}$	$1.51 \cdot 10^{-2}$
$\beta_2$	0.209	$0.78 \cdot 10^{-2}$	$0.92 \cdot 10^{-2}$
$\beta_3$	0.217	$0.75 \cdot 10^{-2}$	$0.90 \cdot 10^{-2}$
$\beta_4$	0.137	$0.69 \cdot 10^{-2}$	$0.84 \cdot 10^{-2}$
$\beta_5$	0.404	$0.65 \cdot 10^{-2}$	$0.82 \cdot 10^{-2}$
$\beta_6$	0.063	$0.74 \cdot 10^{-2}$	$0.88 \cdot 10^{-2}$

**Tabel 3-1:** Regressieresultaten van model (3-1) met OLS en White schatters.

De standaardfout volgens White is bij de constante met 6% en bij de hellingshoeken zijn de standaardfouten met resp. 18, 20, 22, 26 en 19% toegenomen. De aanwezige heteroskedasticiteit zorgt dus voor duidelijke vertekening en dus is het zeer gewenst om de standaardfouten volgens White te berekenen.

Verder kunnen efficiëntere coëfficiënten verkregen worden door FGLS toe te passen. Een bijkomend zeer belangrijk voordeel van het toepassen van FGLS is dat op basis van de hieruit voorkomende resultaten een procedure gegeven wordt waarmee de breedte van het betrouwbaarheidsinterval voor de voorspelde waarde van de windsnelheid bij windmast 1 kan worden berekend. In Paragraaf 3.5 zal hierop worden teruggekomen. Bij FGLS moet het gewicht waarmee alle variabelen vermenigvuldigd worden geschat worden. Dit kan door een model op te stellen waarmee  $e_t^2$  verklaard wordt door de variabele die verantwoordelijk is voor de heteroskedasticiteit. Hiervoor wordt de gemiddelde windsnelheid van alle zes de windmasten in het kwadraat genomen. Zoals Figuur 3-1 laat zien is de windsnelheid namelijk redelijk proportioneel met de residuele standaardfout en dus de windsnelheid in het kwadraat met de residuele variantie. Verder is de gemiddelde windsnelheid van de zes windmasten een goede indicatie van de gemiddelde windsnelheid op Schiphol. In formulevorm ziet het model er dan als volgt uit:

$$e_t^2 = \gamma + \delta F2_t^2 + v_t, \quad (3-3)$$

waar  $v_t$  staat voor de storingsterm en

$$F2_t = \frac{F2_{1,t} + F2_{2,t} + F2_{3,t} + F2_{4,t} + F2_{5,t} + F2_{6,t}}{6}. \quad (3-4)$$

Model (3-3) wordt met OLS geschat door  $e_t^2$  als afhankelijke variabele te gebruiken en  $F2_t^2$  als verklarende variabele, waarbij de standaardfouten van de coëfficiënten enerzijds

geschat worden volgens de gewone OLS schatter en anderzijds met behulp van White's schatter, aangezien de storingsterm  $v_t$  waarschijnlijk heteroskedastisch van aard is. De resultaten hiervan staan in onderstaande tabel:

parameter	coëfficiënt	standaardfout coëfficiënt		$t$ -waarde coëfficiënt	
		OLS	White	OLS	White
$\gamma$	0.330	$1.38 \cdot 10^{-2}$	$1.41 \cdot 10^{-2}$	23.97	23.45
$\delta$	$9.63 \cdot 10^{-3}$	$2.87 \cdot 10^{-4}$	$4.95 \cdot 10^{-4}$	33.55	19.45

**Tabel 3-2:** Regressieresultaten van model (3-3) met OLS en White schatters.

Gezien het grote verschil tussen de OLS standaardfouten en die van White (bij  $\delta$ ) in Tabel 3-2 en daardoor het verschil tussen de  $t$ -waarden is het inderdaad duidelijk dat de storingsterm  $v_t$  heteroskedastisch is. Derhalve is het beter om naar de standaardfouten en  $t$ -waarden van White te kijken om te voorkomen dat er verkeerde conclusies worden getrokken op basis van de standaardfouten. De regressielijn is ook op basis van deze schatters duidelijk significant en dus kan op basis van  $F2_t$  een schatting berekend worden voor  $\sigma_t^2$  (en dus ook voor de breedte van het betrouwbaarheidsinterval).

Mogelijk kan op basis van een extra tweede factor een betere uitspraak verkregen worden over  $\sigma_t^2$ . Immers, naast de hoogte van de windsnelheid kan ook de variatie in de windsnelheid een rol spelen.

Wanneer het verloop van de windsnelheid zeer grillig is dan valt dit op te maken uit de waarde van  $FsF$ . Aangezien  $F2$  en  $FsF$  een lineair verband hebben kan bekeken worden of model (3-3) uitgebreid met  $FsF$  beter voldoet dan model (3-3) zonder deze term. Het volgende model wordt daardoor verder geanalyseerd:

$$e_t^2 = \gamma + \delta F2_t^2 + \kappa FsF_t^2 + v_t, \quad (3-5)$$

waarbij

$$FsF_t = \frac{FsF_{1,t} + FsF_{2,t} + FsF_{3,t} + FsF_{4,t} + FsF_{5,t} + FsF_{6,t}}{6} \quad (3-6)$$

en waar  $FsF_{i,t}$  staat voor de standaardafwijking van alle 3-seconden-gemiddelde windsnelheden bij windmast  $i$  in de 10 minuten voorafgaand aan tijdstip  $t$ .

Model (3-5) wordt geschat met OLS waarbij de covariantiematrix van de coëfficiënten berekend wordt met de schatter van White aangezien bij model (3-1) is gebleken dat de storingsterm  $v_t$  heteroskedastisch van aard is. De resultaten hiervan staan in Tabel 3-3.

Parameter	coëfficiënt	standaardfout coëfficiënt	$t$ -waarde coëfficiënt
$\gamma$	0.278	$1.40 \cdot 10^{-2}$	19.86
$\delta$	$-5.73 \cdot 10^{-3}$	$1.35 \cdot 10^{-3}$	-4.26
$\kappa$	0.764	$6.55 \cdot 10^{-2}$	11.66

**Tabel 3-3:** Regressieresultaten van model (3-5) met White schatters.

Gezien de negatieve coëfficiënt van  $F2$  (veroorzaakt door de sterk aanwezige multicollineariteit) is het beter deze variabele niet in het model op te nemen. Deze conclusie komt ook tot stand wanneer gebruik gemaakt wordt van de partiële correlatiecoëfficiënt van  $F2$ . Deze waarde is namelijk  $3.32 \cdot 10^{-3}$ , wat betekent dat slechts 0.332% van de door  $FsF$  onbeschreven variatie alsnog beschreven wordt door toevoeging van  $F2$ .

Het uiteindelijke model wordt nu als volgt:

$$e_t^2 = \gamma + \kappa F_s F_t^2 + v_t, \quad (3-7)$$

met in onderstaande tabel de regressieresultaten:

parameter	coëfficiënt	standaardfout coëfficiënt	t-waarde coëfficiënt
$\gamma$	0.275	$1.41 \cdot 10^{-2}$	19.53
$\kappa$	0.509	$2.34 \cdot 10^{-2}$	21.77

**Tabel 3-4:** Regressieresultaten van model (3.7) met White schatters.

De  $R^2$  van 0.079 in model (3-3) is verhoogd tot 0.100 in model (3-7). Dit is een redelijke verbetering en dus blijkt  $FsF$  meer verantwoordelijk te zijn voor de aanwezige heteroskedasticiteit dan  $F2$ . Aan de hand van de regressieresultaten kan nu een schatting gemaakt worden voor  $\sigma_t^2$ :

$$s_t^2 = 0.275 + 0.509 F_s F_t^2. \quad (3-8)$$

Het te gebruiken gewicht wordt nu als volgt bepaald:

$$w_t = \frac{1}{\sqrt{s_t^2}}. \quad (3-9)$$

Alle variabelen (inclusief constante) in model (3-1) kunnen nu met dit gewicht worden vermenigvuldigd. Vervolgens kan op deze variabelen OLS toegepast worden. De regressieresultaten die hieruit voortkomen zijn efficiënter dan wanneer OLS toegepast wordt zonder deze wegingsfactor. In welke mate dit is kan onderzocht worden met behulp van Tabel 3-5.

parameter	zonder weging		met weging	
	coëfficiënt	standaardfout coëfficiënt	coëfficiënt	standaardfout coëfficiënt
$\alpha$	-0.027	$1.51 \cdot 10^{-2}$	-0.059	$1.29 \cdot 10^{-2}$
$\beta_2$	0.209	$0.92 \cdot 10^{-2}$	0.207	$0.82 \cdot 10^{-2}$
$\beta_3$	0.217	$0.90 \cdot 10^{-2}$	0.206	$0.81 \cdot 10^{-2}$
$\beta_4$	0.137	$0.84 \cdot 10^{-2}$	0.120	$0.74 \cdot 10^{-2}$
$\beta_5$	0.404	$0.82 \cdot 10^{-2}$	0.445	$0.72 \cdot 10^{-2}$
$\beta_6$	0.063	$0.88 \cdot 10^{-2}$	0.055	$0.77 \cdot 10^{-2}$

**Tabel 3-5:** Regressieresultaten van model (3-1) zonder weging en met weging.

De standaardfouten in Tabel 3-5 zijn geschat met behulp van White's schatter en blijken met resp. 15, 11, 10, 12, 12 en 13% verminderd te zijn. Een redelijke verbetering treedt dus op wanneer de wegingsfactor gehanteerd wordt.

Deze procedure zal voor alle voorspellingsmodellen die in de loop van dit onderzoek aan de orde komen herhaald worden, zodat meest efficiënte schatters worden verkregen. Voorts kan met behulp van model (3-7) voor elk voorspellingsmodel een betrouwbaarheidsinterval worden gegenereerd dat afhankelijk is van de hoogte van  $FsF$ . In Paragraaf 3.5 zal daar verder op worden ingegaan.

### 3.2.2 Autocorrelatie

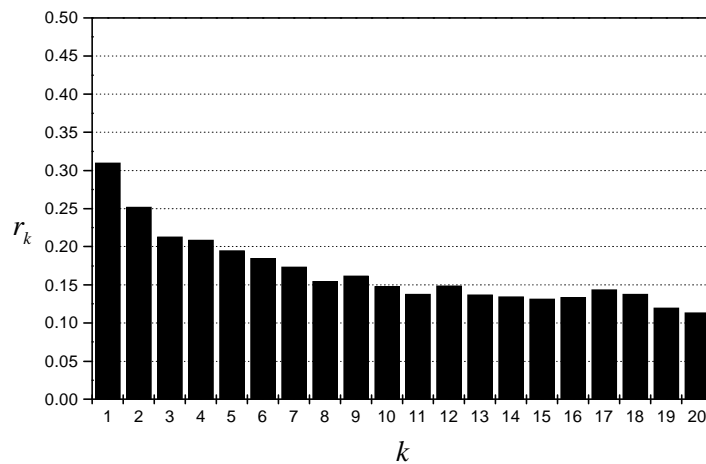
Het toetsen van autocorrelatie kan gedaan worden door gebruik te maken van de residuen  $e_t$  die voortkomen uit model (3-1) dat met OLS geschat is.

Figuur 3-2 geeft een indicatie van de mate en soort van autocorrelatie die in dit model is opgetreden. Het figuur geeft de schattingen voor  $\rho_k$  weer met  $k = 1, 2, \dots, 20$ . Deze schattingen worden als volgt geconstrueerd:

$$\hat{\rho}_k = r_k = \frac{\sum_{t=k+1}^T e_t e_{t-k}}{\sum_{t=1}^T e_t^2}, \quad k = 1, 2, \dots, 20. \quad (3-10)$$

waarbij  $T$  staat voor het totaal aantal tijdmetingen.

(3-10) wordt ook wel de autocorrelatie functie (ACF) genoemd.



**Figuur 3-2:** ACF van de residuen uit model (3-1).

Het figuur geeft duidelijk aan dat er positieve autocorrelatie is die lang aanwezig blijft. Dit heeft tot gevolg dat de standaardfouten worden onderschat en dat de betrouwbaarheidsintervallen hierdoor te nauw zijn.

Dat er sprake is van sterke autocorrelatie is niet zo verwonderlijk. Zoals eerder is opgemerkt kan het namelijk zo zijn dat, gezien de verschillende lokale ruwheden, het bij een bepaalde windrichting harder waait bij een windmast ten opzichte van een andere windmast. Het omgekeerde kan het geval zijn wanneer de wind uit een andere hoek komt. Omdat de windrichting in de tijd geen willekeurig proces is, maar een proces

waarbij de wind voor langere tijd uit een bepaalde hoek kan komen, kunnen verschillen in windsnelheden tussen windmasten enige tijd worden vastgehouden. Hierdoor kunnen de residuen uit model (3-1) een tijdlang positief zijn, waarna vervolgens de wind draait en de residuen een tijdlang negatief kunnen zijn.

De autocorrelatie kan wellicht worden verminderd door het betrekken van de windrichting in het model. Paragraaf 3.4 gaat hier uitvoerig op in.

Een andere mogelijke fysische verklaring van de autocorrelatie in de residuen is convectie. Dit is de benaming van sterke verticale luchtbewegingen die veroorzaakt worden door onder andere de plaatselijke verwarming van het aardoppervlak door de zon. Hierdoor ontstaan temperatuurverschillen, zodat complexe luchtwervelingen kunnen ontstaan, wat zich uit in onder andere lokale toenemende windsnelheid aan het aardoppervlak. Dit gebeuren kan zich enige tijd boven een bepaalde plek bevinden, waardoor een bepaalde afwijking tussen twee windmasten wederom enige tijd kan volhouden.

Convectie is niet of nauwelijks in de regressiemodellen te verwerken waardoor deze eigenschap niet in zijn geheel weggewerkt kan worden.

De autocorrelatie zorgt er dus voor dat de betrouwbaarheidsintervallen van de coëfficiënten te nauw zijn bij OLS. Onderstaande tabel geeft aan in hoeverre de standaardfouten in model (3-1) die met White's schatter worden geschat afwijken ten opzichte van de standaardfouten die worden geschat met behulp van de Newey-West schatter:

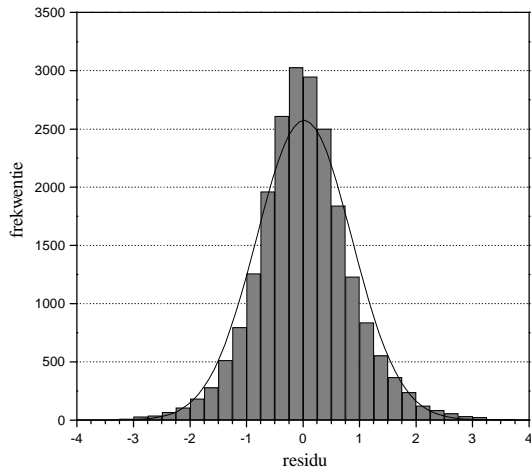
parameter	coëfficiënt	standaardfout coëfficiënt	
		White	Newey-West
$\alpha$	-0.027	$1.51 \cdot 10^{-2}$	$2.46 \cdot 10^{-2}$
$\beta_2$	0.209	$0.92 \cdot 10^{-2}$	$1.07 \cdot 10^{-2}$
$\beta_3$	0.217	$0.90 \cdot 10^{-2}$	$1.14 \cdot 10^{-2}$
$\beta_4$	0.137	$0.84 \cdot 10^{-2}$	$1.01 \cdot 10^{-2}$
$\beta_5$	0.404	$0.82 \cdot 10^{-2}$	$1.23 \cdot 10^{-2}$
$\beta_6$	0.063	$0.88 \cdot 10^{-2}$	$1.13 \cdot 10^{-2}$

**Tabel 3-6:** Regressieresultaten van model (3-1) met White en Newey-West schatters.

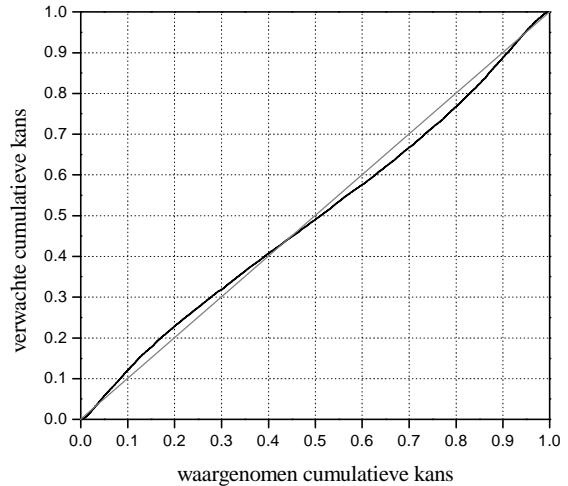
De standaardfout volgens Newey-West is bij de constante met 63% toegenomen ten opzichte van die van White en de standaardfouten bij de hellingshoeken met resp. 16, 26, 20, 51 en 29%. De aanwezige autocorrelatie zorgt dus voor duidelijke vertekening zodat het zeer gewenst is om de standaardfouten volgens Newey-West te berekenen. In het verloop van dit onderzoek zal derhalve in veel gevallen gebruik gemaakt worden van deze schatter van de covariantiematrix.

### 3.2.3 Non-normaliteit

Of er sprake is van non-normaliteit van de residuen kan getoetst worden met behulp van een histogram (Figuur 3-3) en een normal probability plot (Figuur 3-4). Opnieuw zijn hiervoor de residuen die voortkomen uit het met OLS geschatte model (3-1) gebruikt.



**Figuur 3-3:** Histogram van de residuen uit model (3-1).



**Figuur 3-4:** P-P normal probability plot van de residuen uit model (3-1).

Beide plots maken duidelijk dat de residuen redelijker wijs benaderd worden door een normale verdeling. Wel is er iets voor het eerste en iets na het derde kwantiel een relatief lage concentratie van residuen, rondom de mediaan is er juist relatief een hoge concentratie van residuen. Deze afwijkingen zijn echter betrekkelijk gering en daardoor heeft dit geen nadelige gevolgen voor de resultaten. De  $t$ - en  $P$ -waarden kunnen dus op de juiste manier worden geïnterpreteerd en de betrouwbaarheidsintervallen kunnen volgens (2-17) opgesteld worden.

### 3.2.4 Multicollineariteit

Tenslotte kan er sprake zijn van multicollineariteit. Hierop kan getoetst worden met behulp van een correlatiematrix (Tabel 3-7) van de verklarende variabelen in model (3-1). In deze tabel zijn ook de correlaties met de verklarende variabele ( $F2_1$ ) opgenomen.

	$F2_1$	$F2_2$	$F2_3$	$F2_4$	$F2_5$	$F2_6$
$F2_1$	1.000	0.914	0.912	0.905	0.921	0.899
$F2_2$	–	1.000	<b>0.921</b>	<b>0.918</b>	<b>0.906</b>	<b>0.945</b>
$F2_3$	–	–	1.000	<b>0.918</b>	<b>0.907</b>	<b>0.917</b>
$F2_4$	–	–	–	1.000	<b>0.904</b>	<b>0.918</b>
$F2_5$	–	–	–	–	1.000	<b>0.891</b>
$F2_6$	–	–	–	–	–	1.000

**Tabel 3-7:** Correlatiematrix van de verklarende variabelen en afhankelijke variabele in model (3-1).

Uit de tabel komt zeer duidelijk naar voren dat de variabelen sterk met elkaar gecorreleerd zijn. Dit is natuurlijk niet verwonderlijk. Immers, als het hard waait bij een bepaalde windmast zal het vermoedelijk ook hard waaien bij een andere windmast. Hetzelfde geldt natuurlijk bij lage windsnelheden.

Door de aanwezigheid van multicollineariteit zullen de individuele effecten van de verklarende windmasten in het meervoudige regressiemodel een minder grote rol spelen

dan wanneer er in het model slechts één verklarende windmast is opgenomen. Dit kon al afgeleid worden uit Tabel 3-6. Op basis van de hierin vermelde coëfficiënten en standaardfouten kunnen de afzonderlijke  $t$ -waarden berekend worden. Doen we dit voor bijvoorbeeld windmast 2 dan is hier de  $t$ -waarde  $0.209 / 1.07 \cdot 10^{-2} = 19.61$ . Als we deze waarde berekenen in het model waarbij windmast 2 als enige verklarende variabele is opgenomen, dan zien we een  $t$ -waarde van  $0.914 / 0.61 \cdot 10^{-2} = 148.86$ . De individuele kracht van de variabele in het enkelvoudige regressiemodel is dus veel groter dan de toegevoegde waarde van deze variabele in het meervoudige regressiemodel. Zolang de toegevoegde waarde echter significant is, is het verstandig om toch gebruik te maken van deze variabele wanneer het doel is om een maximale voorspelkracht te verkrijgen. In dit onderzoek is dat het geval en daarom wordt gekozen om ondanks de multicollineariteit net zoveel variabelen toe te voegen aan het model zolang deze variabelen een significante bijdrage leveren.

### 3.3 Van back-up model naar meervoudig lineair regressie-model

Nu de dataproblemen op hun aanwezigheid zijn getoetst en eventueel hiervoor is gecorrigeerd kan begonnen worden met de modelbouw en de analyse hiervan.

Hierbij worden eerst de verschillende back-up modellen met elkaar vergeleken waarna de restricties losgelaten worden door het model uit te breiden tot verschillende enkelvoudige regressiemodellen. Tenslotte kan dit uitgebreid worden door het stapsgewijs opnemen van meerdere windmasten tot een meervoudig lineair regressiemodel.

#### 3.3.1 Het back-up model

In de huidige situatie wordt bij de meeste luchthavens de back-up methode toegepast. Hierbij wordt een waarde van één bepaalde windmast gebruikt als vervanging van een niet-functionerende windmast. Op Schiphol is besloten windmast 5 als back-up te gebruiken voor windmast 1. Allereerst wordt hier onderzocht of windmast 5 inderdaad het best geschikt is als back-up. Hiertoe worden de volgende modellen met elkaar vergeleken:

$$F2_{1,t} = F2_{i,t} + \varepsilon_t^{li}, \quad i = 2,3,\dots,6. \quad Ii$$

Zoals eerder is opgemerkt zijn dit sterk gerespecteerde lineair regressiemodellen waarbij de parameters al vastliggen (zie Deelparagraaf 2.1.2). Onderzocht zal worden welk van deze modellen het best als back-up voldoet. Dit wordt gedaan door te kijken naar de afzonderlijke ‘*Root Mean Squared Error*’-waarden. Deze *RMSE* wordt als volgt gedefinieerd:

$$RMSE = \sqrt{\frac{1}{T-k} \sum_{t=1}^T e_t^2}, \quad (3-11)$$

waar  $k$  staat voor het aantal parameters in het model en  $e_t$  voor het residu op tijdstip  $t$ . In het geval van het back-up model is  $k$  gelijk aan nul, aangezien dit model volledig vastgelegd is.

Hoe kleiner de  $RMSE$  des te krachtiger is het model. Daarom zal voor het back-up model met de kleinste  $RMSE$ -waarde gekozen worden. Een ander argument dat voor de  $RMSE$ -waarde gekozen is, is dat deze waarde gebruikt kan worden voor het construeren van een betrouwbaarheidsinterval voor de uitspraak volgens (2-17). Immers, de  $RMSE$ -waarde kan ook geschreven worden als de wortel van de som van het kwadraat van het gemiddelde residu en de residuele variantie. In formulevorm:

$$RMSE = \sqrt{(\bar{e}_t)^2 + s^2}. \quad (3-12)$$

Vanaf de volgende deelparagraaf kan het betrouwbaarheidsinterval rechtstreeks afgeleid worden uit (3-12) aangezien het gemiddelde residu bij lineaire regressie gelijk is aan nul.

Hieronder staat een overzicht van de afzonderlijke  $RMSE$ -waarden bij de verschillende mogelijke back-up modellen:

modeldata	
verklarende windmast	$RMSE$
2	1.084
3	1.107
4	1.163
5	<b>1.023</b>
6	1.172

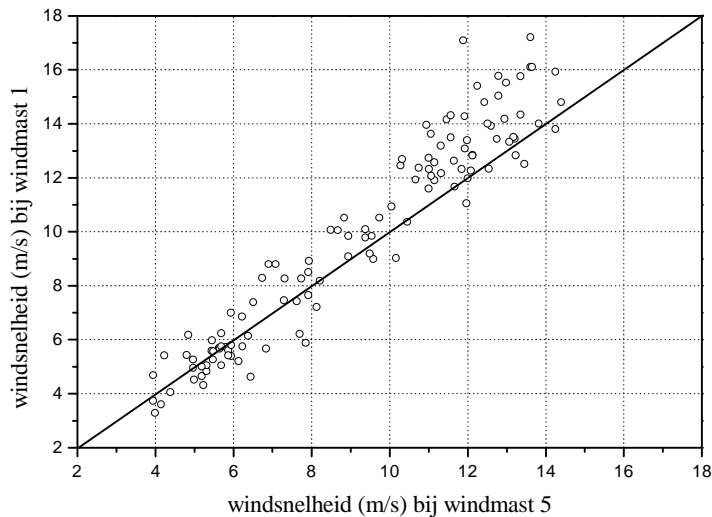
**Tabel 3-8:**  $RMSE$ -waarden bij de verschillende mogelijke back-up modellen.

Duidelijk is uit deze tabel op te maken dat windmast 5 het best als back-up gebruikt kan worden, aangezien hierbij de  $RMSE$ -waarde het kleinst is. De beslissing om windmast 5 als back-up te laten fungeren voor windmast 1 is dus waarschijnlijk niet toevalligerwijs tot stand gekomen.

### 3.3.2 Het enkelvoudig lineaire regressiemodel

Het is mogelijk dat het bij een bepaalde windmast structureel harder waait dan bij een andere windmast doordat hij minder beschut ligt. Hier kan rekening mee gehouden worden door de back-up methode te vervangen door een enkelvoudig lineair regressiemodel waarbij de parameters  $\alpha$  en  $\beta$  niet gerespecteerd zijn tot resp. 0 en 1 maar geschat worden met OLS.

Figuur 3-5 laat zien dat dit model wellicht voor verbetering kan zorgen aangezien de geplote lijn (volgens back-up methode) het verband tussen de windsnelheid bij windmast 1 en de windsnelheid bij windmast 5 op 26 december 1997 (zie Figuur 2-1) niet goed lijkt te beschrijven, zodat de waarden 0 en 1 voor  $\alpha$  en  $\beta$  niet gerechtvaardigd lijken te zijn. In het vervolg van dit onderzoek zal blijken dat de windrichting hierbij een belangrijke rol speelt.



**Figuur 3-5:** Modelbouw volgens back-up methode.

Om zo efficiënt mogelijke schatters te verkrijgen zijn in de enkelvoudige lineaire regressiemodellen alle variabelen (afhankelijke en verklarende) inclusief constante met de gewichtsfactor  $w_t$  vermenigvuldigd omdat gebleken is dat er sprake is van heteroskedasticiteit (zie Deelparagraaf 3.2.1). Op deze manier ontstaan de volgende modellen:

$$F2_{1,t} w_t = \alpha w_t + \beta_i F2_{i,t} w_t + \varepsilon_t^{\text{Ii}} w_t, \quad i = 2, 3, \dots, 6. \quad \text{Ii}$$

De covariantiematrix van de coëfficiënten wordt berekend met de Newey-West schatter om verkeerde conclusies op basis van de  $t$ -waarden ten gevolge van autocorrelatie te voorkomen. De regressieresultaten staan in onderstaande tabel:

model	parameter	coëfficiënt	standaardfout coëfficiënt	$t$ -waarde coëfficiënt	95% betrouwbaar- heidsinterval coëfficiënt
II2	$\alpha$	0.409	$3.03 \cdot 10^{-2}$	13.49	$\langle 0.350, 0.468 \rangle$
	$\beta_2$	0.895	$0.59 \cdot 10^{-2}$	151.35	$\langle 0.883, 0.907 \rangle$
II3	$\alpha$	0.284	$2.87 \cdot 10^{-2}$	9.90	$\langle 0.228, 0.340 \rangle$
	$\beta_3$	0.982	$0.59 \cdot 10^{-2}$	165.78	$\langle 0.970, 0.994 \rangle$
II4	$\alpha$	0.652	$3.26 \cdot 10^{-2}$	19.97	$\langle 0.588, 0.712 \rangle$
	$\beta_4$	0.910	$0.66 \cdot 10^{-2}$	138.36	$\langle 0.897, 0.923 \rangle$
II5	$\alpha$	0.168	$2.58 \cdot 10^{-2}$	6.52	$\langle 0.117, 0.219 \rangle$
	$\beta_5$	0.979	$0.58 \cdot 10^{-2}$	169.32	$\langle 0.968, 0.990 \rangle$
II6	$\alpha$	0.568	$3.35 \cdot 10^{-2}$	16.96	$\langle 0.502, 0.634 \rangle$
	$\beta_6$	0.879	$0.63 \cdot 10^{-2}$	139.53	$\langle 0.867, 0.891 \rangle$

**Tabel 3-9:** Modeloutput van de verschillende enkelvoudige lineaire regressiemodellen.

Uit deze tabel blijkt dat de verschillende constanten en hellingshoeken significant afwijken van 0 en 1, waarmee het vermoeden bevestigd wordt dat de back-up methode niet optimaal is.

Vervolgens is nagegaan welke windmast het best als verklarende variabele in het model kan worden opgenomen. Hierbij is opnieuw de  $RMSE$ -waarde als maatstaf gebruikt.  $k$

neemt nu de waarde 2 aan, aangezien er twee parameters zijn die geschat moeten worden. Dit heeft echter nauwelijks invloed op de resultaten, omdat  $T$  in vergelijking met  $k$  behoorlijk groot is.

De modelresultaten zijn ook toegepast op de validatiedata, zodat ze op de juiste manier kunnen worden geïnterpreteerd. Immers, bij het toevoegen van parameters aan het model wordt de verklarende kracht van het model altijd beter. In theorie kan men hiermee doorgaan tot 100% wordt verklaard. Wanneer de geschatte parameters echter worden toegepast op een onafhankelijke dataset, dan zal blijken dat de verklarende kracht op een gegeven moment niet meer stijgt, wat ook niet reëel zou zijn aangezien er in de praktijk vrijwel altijd een onverklaarbaar gedeelte overblijft. De betrouwbaarheidsintervallen die geconstrueerd kunnen worden op basis van de *RMSE*-waarden die voortkomen uit de validatiedata geven dan ook een beter beeld van de marges die behoren aan een voorspelling van de windsnelheid bij windmast 1.

De resultaten volgen hieronder, ter vergelijking zijn tevens nogmaals de back-up resultaten weergegeven:

modeldata			validatiedata			
verklarende windmast	<i>RMSE</i>		verklarende windmast	<i>RMSE</i>		verschil (%)
	back-up	regressie		back-up	regressie	
2	1.084	1.061	2	1.089	1.063	-2.4
3	1.107	1.075	3	1.122	1.095	-2.7
4	1.163	1.115	4	1.188	1.140	-4.0
5	<b>1.023</b>	<b>1.017</b>	5	<b>1.011</b>	<b>1.005</b>	-0.6
6	1.172	1.144	6	1.181	1.152	-2.5

**Tabel 3-10:** *RMSE*-waarden van resp. de back-up modellen en enkelvoudige lineaire regressiemodellen.

Uit Tabel 3-10 kan opgemaakt worden dat zowel bij de back-up als bij de enkelvoudige regressiemodellen windmast 5 het meest geschikt is om een uitspraak te genereren over windmast 1. Verder is te zien dat de verbetering die optreedt wanneer het enkelvoudige regressiemodel wordt gebruikt i.p.v. het back-up model betrekkelijk gering is. Het enkelvoudige regressiemodel is echter wel een goed uitgangspunt voor modeluitbreiding.

### 3.3.3 Het meervoudige lineaire regressiemodel

Je kunt je afvragen of het meenemen van één enkele mast optimaal is bij het doen van een zo goed mogelijke uitspraak. Wellicht kunnen meerdere masten een betere uitspraak genereren, aangezien een koppeling van deze masten een soort filtering zou kunnen aanbrengen in het grillige patroon wat de gemiddelde 2-minuut windsnelheid bij een bepaalde windmast kenmerkt. Daartoe kan het enkelvoudig lineair regressiemodel stapsgewijs uitgebreid worden tot een meervoudig lineair regressiemodel middels 'stepwise selection' (zie Draper en Smith [1980]). Het statistische softwarepakket SPSS voorziet in deze selectieprocedure en is hiervoor gebruikt. Hierbij wordt het (zeer beperkte) enkelvoudige regressiemodel met enkel de constante in het model als beginpunt genomen en wordt per stap de windmast met de hoogste  $t$ -waarde opgenomen in het model als deze  $t$ -waarde voldoende groot is ( $>1.96$ ). Verder wordt telkens de windmast met de laagste  $t$ -waarde uit het model verwijderd als deze  $t$ -waarde te laag wordt ( $<1.96$ ). Hiermee wordt doorgegaan totdat geen enkele windmast meer voor opname in of verwijdering uit het model in aanmerking komt.

Deze modellen worden met OLS geschat waarbij alle variabelen (afhankelijke en verklarende) inclusief constante worden vermenigvuldigd met de wegingsfactor  $w_t$ . Verder wordt de covariantiematrix van de coëfficiënten opnieuw berekend met de Newey-West schatter om verkeerde conclusies op basis van de  $t$ -waarden te voorkomen. Na de analyse blijken uiteindelijk alle windmasten te worden opgenomen, waarbij er tussentijds geen variabelen uit het model zijn verwijderd. Hierbij is windmast 5 als eerste in het model opgenomen (zoals af te leiden is uit de  $t$ -waarden in Tabel 3-9), vervolgens windmast 2, gevolgd door resp. windmast 3, 4 en 6.

De modeluitbreiding ziet er derhalve als volgt uit (vanaf het optimale back-up model):

$$\begin{aligned}
 F2_{1,t} &= F2_{5,t} + \varepsilon_t^A, & \text{A} \\
 F2_{1,t} w_t &= \alpha w_t + \beta_5 F2_{5,t} w_t + \varepsilon_t^B w_t, & \text{B} \\
 F2_{1,t} w_t &= \alpha w_t + \sum_{i=2,5} \beta_i F2_{i,t} w_t + \varepsilon_t^C w_t, & \text{C} \\
 F2_{1,t} w_t &= \alpha w_t + \sum_{i=2,3,5} \beta_i F2_{i,t} w_t + \varepsilon_t^D w_t, & \text{D} \\
 F2_{1,t} w_t &= \alpha w_t + \sum_{i=2,3,4,5} \beta_i F2_{i,t} w_t + \varepsilon_t^E w_t, & \text{E} \\
 F2_{1,t} w_t &= \alpha w_t + \sum_{i=2}^6 \beta_i F2_{i,t} w_t + \varepsilon_t^F w_t. & \text{F}
 \end{aligned}$$

De regressieresultaten van deze modellen staan in onderstaande tabel:

model	parameter	coëfficiënt	standaardfout coëfficiënt	$t$ -waarde coëfficiënt	95% betrouwbaar- heidsinterval coëfficiënt
B	$\alpha$	0.168	$2.58 \cdot 10^{-2}$	6.52	$\langle 0.117, 0.219 \rangle$
	$\beta_5$	0.979	$0.58 \cdot 10^{-2}$	169.31	$\langle 0.968, 0.990 \rangle$
C	$\alpha$	0.007	$2.34 \cdot 10^{-2}$	0.29	$\langle -0.039, 0.053 \rangle$
	$\beta_5$	0.591	$0.92 \cdot 10^{-2}$	64.55	$\langle 0.573, 0.609 \rangle$
	$\beta_2$	0.409	$0.88 \cdot 10^{-2}$	46.44	$\langle 0.392, 0.426 \rangle$
D	$\alpha$	-0.072	$2.26 \cdot 10^{-2}$	-3.17	$\langle -0.116, -0.028 \rangle$
	$\beta_5$	0.481	$1.02 \cdot 10^{-2}$	47.31	$\langle 0.461, 0.501 \rangle$
	$\beta_2$	0.281	$0.91 \cdot 10^{-2}$	31.06	$\langle 0.263, 0.299 \rangle$
	$\beta_3$	0.268	$0.93 \cdot 10^{-2}$	28.84	$\langle 0.250, 0.286 \rangle$
E	$\alpha$	-0.056	$2.27 \cdot 10^{-2}$	-2.47	$\langle -0.100, -0.012 \rangle$
	$\beta_5$	0.446	$1.08 \cdot 10^{-2}$	41.54	$\langle 0.425, 0.467 \rangle$
	$\beta_2$	0.238	$0.92 \cdot 10^{-2}$	26.02	$\langle 0.220, 0.256 \rangle$
	$\beta_3$	0.217	$1.00 \cdot 10^{-2}$	21.77	$\langle 0.197, 0.237 \rangle$
	$\beta_4$	0.131	$0.95 \cdot 10^{-2}$	13.76	$\langle 0.112, 0.150 \rangle$
F	$\alpha$	-0.059	$2.27 \cdot 10^{-2}$	-2.62	$\langle -0.103, -0.015 \rangle$
	$\beta_5$	0.445	$1.08 \cdot 10^{-2}$	41.03	$\langle 0.424, 0.466 \rangle$
	$\beta_2$	0.207	$1.01 \cdot 10^{-2}$	20.43	$\langle 0.187, 0.227 \rangle$
	$\beta_3$	0.206	$1.04 \cdot 10^{-2}$	19.88	$\langle 0.186, 0.226 \rangle$
	$\beta_4$	0.120	$0.95 \cdot 10^{-2}$	12.59	$\langle 0.101, 0.139 \rangle$
	$\beta_6$	0.055	$1.05 \cdot 10^{-2}$	5.29	$\langle 0.034, 0.076 \rangle$

**Tabel 3-11:** Regressieresultaten van modeluitbreiding via ‘stepwise selection’.

Uit de tabel valt duidelijk af te lezen dat de  $t$ -waarden van een bepaalde coëfficiënt (b.v. die van  $\beta_5$ ) in rap tempo afnemen bij het toenemen van het aantal windmasten. Dit wordt veroorzaakt door het gezamenlijk kleiner worden van de coëfficiënten alsmede het toenemen van de standaardfouten (en dus het breder worden van de betrouwbaarheidsintervallen rondom de coëfficiënten), wat beiden het gevolg is van de sterke multicollineariteit die aanwezig is. De individuele kracht van de verschillende windmasten neemt dus snel af bij het toenemen van het aantal windmasten, maar de totale voorspelkracht blijft stijgen aangezien in het uiteindelijke model nog steeds alle variabelen een significante toegevoegde waarde hebben.

Met behulp van de *RMSE* kan nu stapsgewijs bekeken worden in hoeverre het toevoegen van de verschillende windmasten de uitspraak over de windsnelheid bij windmast 1 bevordert. Opnieuw zijn verder de regressieresultaten uit Tabel 3-11 toegepast op de validatiedata, waarvan vervolgens wederom de *RMSE*-waarden berekend kunnen worden:

modeldata			validatiedata		
model	<i>RMSE</i>	totale percentuele verandering <i>RMSE</i>	model	<i>RMSE</i>	totale percentuele verandering <i>RMSE</i>
A	1.023	–	A	1.011	–
B	1.017	-0.6	B	1.005	-0.6
C	0.890	-13.0	C	0.883	-12.7
D	0.856	-16.3	D	0.858	-15.1
E	0.847	-17.2	E	0.850	-15.9
F	0.845	-17.4	F	0.848	-16.1

**Tabel 3-12:** Stapsgewijze verandering in de *RMSE* vanaf het meest optimale back-up model tot en met het meervoudige regressiemodel waarin alle masten zijn opgenomen via de ‘stepwise selection’ procedure.

De tabel maakt duidelijk dat de stapsgewijze verbetering (van C t/m F) steeds kleiner wordt. Dit is niet verwonderlijk gezien de door de multicollineariteit veroorzaakte steeds kleiner worden van de  $t$ -waarden in Tabel 3-11.

De grootste winst wordt behaald door het toevoegen van een tweede windmast (windmast 2) naast windmast 5 en (in mindere mate) een derde windmast (windmast 3). Het toevoegen van een vierde windmast (windmast 4) en zeker de laatste (windmast 6) levert wel een verbetering op, maar deze is zeer gering. De toegevoegde individuele waarden van deze twee masten is dus niet groot, waardoor met het oog op het eventuele niet functioneren van meerdere masten naast windmast 1 volstaan kan worden met model D.

Tabel 3-12 gaf een overzicht van de *RMSE*-waarden bij de verschillende modellen. Hierbij ging het steeds om de gemiddelde waarde over alle windsnelheden. Eerder werd echter vastgesteld dat de residuele standaardfout (en dus *RMSE*) toeneemt naarmate de windsnelheid hoger wordt. Het is daarom interessant om te kijken naar de kracht van de modellen A t/m F bij oplopende windsnelheden, temeer omdat vooral hoge windsnelheden met het oog op de veiligheid van belang zijn. Daarom is gekeken naar de berekende waarde van de *RMSE* in ieder interval van 2 m/s. Ook is gekeken naar de totale winst van het model ten opzichte van het back-up model per interval. De resultaten hiervan staan in Tabel 3-13.

validatiedata							
windsnelheid (m/s)	RMSE						verandering (%) F t.o.v. A
	A	B	C	D	E	F	
2-4	0.669	0.676	0.618	0.608	0.604	0.604	9.7
4-6	0.858	0.848	0.772	0.754	0.753	0.753	12.2
6-8	1.171	1.156	1.020	0.990	0.982	0.981	16.2
8-10	1.434	1.422	1.203	1.160	1.146	1.141	20.4
10-12	1.615	1.607	1.367	1.315	1.283	1.276	21.0
>12	2.036	2.054	1.674	1.601	1.551	1.540	24.4

**Tabel 3-13:** RMSE-waarden bij de verschillende modellen per interval van 2 m/s.

Uit deze tabel komt inderdaad naar voren dat de residuele standaardfout groter wordt naarmate de gemiddelde windsnelheid hogere waarden aanneemt. Dit is geen verrassende conclusie. Wat wel opvalt is dat ook de behaalde winst geleidelijk groter wordt naarmate de windsnelheid toeneemt. De kracht van de modellen lijkt dus te groeien bij toenemende snelheid. Gezien het belang bij hoge windsnelheden is dit een belangrijke reden om modelbouw te prefereren boven de back-up methode.

### 3.3.4 Nadere beschouwing volgorde van toelating

In de vorige deelparagraaf is besproken wat de meest optimale methode is bij het opnemen van de verschillende windmasten. Ook werd aangegeven dat in het geval van het niet functioneren van één of meerdere verklarende windmasten volstaan kan worden met model D waarin drie masten zijn opgenomen als verklarende variabelen. Bij dit laatste is het natuurlijk niet vanzelfsprekend dat de windmasten 3 en 6 (die als laatste zijn toegevoegd) niet functioneren. De andere windmasten komen hier natuurlijk ook voor in aanmerking.

De vraag die in bovenstaande al verborgen zit is dan ook wat voor invloed de volgorde van toelating heeft op kracht van de modellen. Bekend is reeds wat de meest optimale volgorde is, maar er staat een vraagteken bij het verschil in voorspellende kracht tussen de meest optimale volgorde en een andere volgorde.

Om hier meer over te kunnen zeggen is de meest optimale volgorde vergeleken met de minst optimale volgorde. Deze wordt bepaald via dezelfde ‘stepwise selection’ procedure zoals die ook gebruikt werd bij het bepalen van de meest optimale volgorde, met het verschil dat nu telkens de windmast wordt opgenomen met de laagste  $t$ -waarde i.p.v. de hoogste  $t$ -waarde. Verder worden er geen windmasten tussentijds uit het model verwijderd, wat dus neerkomt op een soort ‘forward selection’ (zie Draper en Smith [1980]). Opnieuw worden de modellen geschat met OLS waarbij alle variabelen vermenigvuldigd worden met de gewichtfactor  $w_t$  en de covariantiematrix van de coëfficiënten met de Newey-West schatter geschat wordt. Na analyse blijkt de volgorde van toelating nu resp. windmast 4, 6, 2, 3 en 5 te zijn. Als beginstap wordt het back-up model genomen met de hoogste RMSE-waarde (windmast 6, zie Tabel 3-8).

De modelbouw wordt dan als volgt:

$$\begin{aligned}
 F2_{1,t} &= F2_{5,t} + \varepsilon_t^{A-}, & \text{A-} \\
 F2_{1,t} w_t &= \alpha w_t + \beta_4 F2_{4,t} w_t + \varepsilon_t^{B-} w_t, & \text{B-} \\
 F2_{1,t} w_t &= \alpha w_t + \sum_{i=4,6} \beta_i F2_{i,t} w_t + \varepsilon_t^{C-} w_t, & \text{C-}
 \end{aligned}$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2,4,6} \beta_i F2_{i,t} w_t + \varepsilon_t^{D-} w_t, \quad \text{D-}$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2,3,4,6} \beta_i F2_{i,t} w_t + \varepsilon_t^{E-} w_t, \quad \text{E-}$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2}^6 \beta_i F2_{i,t} w_t + \varepsilon_t^F w_t. \quad \text{F}$$

De regressieresultaten staan in onderstaande tabel:

model	parameter	coëfficiënt	standaardfout coëfficiënt	t-waarde coëfficiënt	95% betrouwbaar- heidsinterval coëfficiënt
B-	$\alpha$	0.652	$3.26 \cdot 10^{-2}$	19.97	$\langle 0.588, 0.716 \rangle$
	$\beta_4$	0.910	$0.66 \cdot 10^{-2}$	138.36	$\langle 0.897, 0.923 \rangle$
C-	$\alpha$	0.384	$3.10 \cdot 10^{-2}$	12.36	$\langle 0.323, 0.445 \rangle$
	$\beta_4$	0.514	$1.08 \cdot 10^{-2}$	47.50	$\langle 0.493, 0.535 \rangle$
	$\beta_6$	0.433	$0.98 \cdot 10^{-2}$	19.97	$\langle 0.414, 0.452 \rangle$
D-	$\alpha$	0.264	$2.86 \cdot 10^{-2}$	9.22	$\langle 0.208, 0.320 \rangle$
	$\beta_4$	0.375	$1.04 \cdot 10^{-2}$	35.93	$\langle 0.355, 0.395 \rangle$
	$\beta_6$	0.148	$1.30 \cdot 10^{-2}$	11.40	$\langle 0.123, 0.173 \rangle$
	$\beta_2$	0.433	$1.35 \cdot 10^{-2}$	32.10	$\langle 0.407, 0.459 \rangle$
E-	$\alpha$	0.118	$2.66 \cdot 10^{-2}$	4.44	$\langle 0.066, 0.170 \rangle$
	$\beta_4$	0.243	$1.07 \cdot 10^{-2}$	22.67	$\langle 0.222, 0.264 \rangle$
	$\beta_6$	0.069	$1.29 \cdot 10^{-2}$	5.34	$\langle 0.044, 0.094 \rangle$
	$\beta_2$	0.330	$1.22 \cdot 10^{-2}$	27.16	$\langle 0.306, 0.354 \rangle$
	$\beta_3$	0.357	$1.23 \cdot 10^{-2}$	29.12	$\langle 0.333, 0.381 \rangle$
F	$\alpha$	-0.059	$2.68 \cdot 10^{-2}$	-2.62	$\langle -0.112, -0.006 \rangle$
	$\beta_4$	0.120	$0.95 \cdot 10^{-2}$	12.59	$\langle 0.101, 0.139 \rangle$
	$\beta_6$	0.055	$1.05 \cdot 10^{-2}$	5.28	$\langle 0.034, 0.076 \rangle$
	$\beta_2$	0.207	$1.01 \cdot 10^{-2}$	20.43	$\langle 0.187, 0.227 \rangle$
	$\beta_3$	0.206	$1.04 \cdot 10^{-2}$	19.88	$\langle 0.186, 0.226 \rangle$
	$\beta_5$	0.445	$1.08 \cdot 10^{-2}$	41.03	$\langle 0.424, 0.466 \rangle$

**Tabel 3-14:** Regressieresultaten van modeluitbreiding in de minst optimale volgorde van het opnemen van de windmasten.

Bovenstaande coëfficiënten kunnen worden toegepast op de validatiedata, waarna hiervan de *RMSE*-waarden berekend kunnen worden. Deze staan in Tabel 3-15, waarin ze vergeleken worden met de *RMSE*-waarden die volgden uit de optimale volgorde van opname.

Uit Tabel 3-15 kan een aantal interessante zaken gehaald worden:

- Vooral wanneer er weinig masten (1 of 2) zijn opgenomen in het model, is de kracht van de modellen via de meest optimale volgorde een stuk hoger dan de modellen via de minst optimale manier.
- Model C- (twee verklarende masten) scoort minder (*RMSE*: 1.044) dan het optimale back-up model (*RMSE*: 1.011).
- Bij de minst optimale volgorde is de toegevoegde waarde bij elke opname van een windmast redelijk groot, ook bij de laatste opnames. Gezien de voorspelkracht van

het model is het dan ook verstandig om hierbij zoveel mogelijk masten op te nemen in het model.

validatiedata			validatiedata		
model	<i>RMSE</i>	totale percentuele verandering <i>RMSE</i> t.o.v. model A	model	<i>RMSE</i>	totale percentuele verandering <i>RMSE</i> t.o.v. model A
A	1.011	–	A-	1.181	16.8
B	1.005	-0.6	B-	1.152	13.9
C	0.883	-12.7	C-	1.027	1.6
D	0.858	-15.1	D-	0.962	-4.8
E	0.850	-15.9	E-	0.928	-8.2
F	0.848	-16.1	F-	0.848	-16.1

**Tabel 3-15:** Vergelijking *RMSE* tussen meest optimale volgorde en minst optimale volgorde.

Het verschil tussen de twee procedures lijkt dus behoorlijk. Gezien het feit dat hier de meest extreme procedures met elkaar vergeleken zijn, zullen de verschillen wanneer andere volgorden worden gebruikt niet zo groot zijn. Toch is het verstandig om bij de modellen gebruik te maken van alle windmasten waarvan metingen beschikbaar zijn. Bovenstaande laat immers zien dat bepaalde masten een behoorlijke toegevoegde waarde hebben, zodat door het niet opnemen van deze masten de kwaliteit van de modellen danig aangetast wordt.

### 3.4 Het betrekken van lokale beschuttingsfactoren in de modelbouw

#### 3.4.1 Waarom kan modellering van lokale beschuttingsfactoren van belang zijn?

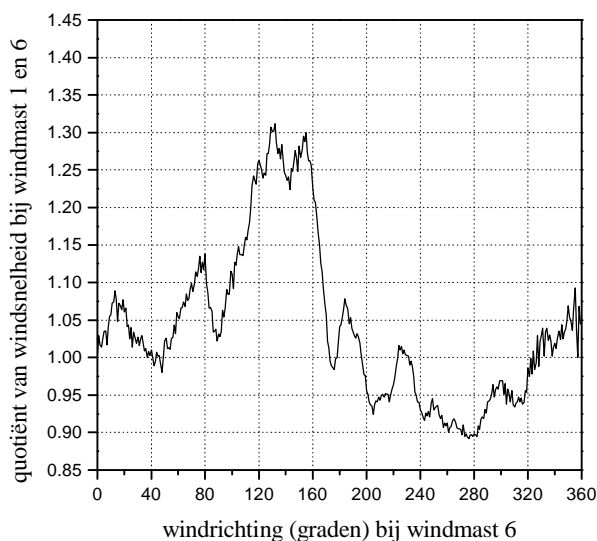
In Deelparagraaf 3.2.2 kwam al aan het licht dat het verstandig kan zijn om de windrichting mee te nemen in het model, gezien de verschillende lokale ruwheden die aanwezig zijn op een luchthaven. Wanneer hier rekening mee gehouden wordt zal wellicht de autocorrelatie afnemen maar kan ook de kracht van de modellen toenemen. De verschillende ruwheden worden voornamelijk veroorzaakt door bebouwing en in mindere mate begroeiing. Deze ruwheden hebben invloed op de windsnelheid omdat wanneer wind zich over een bepaald traject voortbeweegt wordt afgeremd als er veel obstakels in het landschap zijn. Vandaar dat het aan de kust vrijwel altijd harder waait dan in het binnenland, omdat de zee nauwelijks wrijving veroorzaakt. Dit is ook de reden dat tropische cyclonen vaak beduidend in kracht afnemen zodra ze boven het continent verschijnen.

Op luchthavens zoals Schiphol Airport zijn de verschillen in ruwheid niet van die orde, maar wellicht wel dusdanig dat wanneer er rekening mee wordt gehouden de modellen krachtiger worden en beter in staat zijn een uitspraak te genereren over de windsnelheid bij een niet-functionerende windmast.

De invloed van de lokale beschutting kan in de modelbouw verwerkt worden door de windrichting in de modelbouw te betrekken. Immers, wanneer er bijvoorbeeld aan de noordzijde van een windmast veel bebouwing is ten opzichte van de noordzijde van een

andere windmast dan zal de gemiddelde windsnelheid bij noordelijke windrichtingen vermoedelijk lager zijn bij de eerste windmast ten opzichte van gemiddelde windsnelheid bij de tweede windmast doordat de bebouwing (en begroeiing) een remmend effect heeft op de windsnelheid.

Om een indicatie te krijgen van de mate waarin dit optreedt staat hieronder een figuur afgebeeld waarin per graad windrichting bij windmast 6 het quotiënt geplot is van de gemiddelde 2-minuut windsnelheid bij windmast 1 tegen de gemiddelde 2-minuut windsnelheid bij windmast 6. Een overzicht van de quotiëntplots van de gemiddelde 2-minuut windsnelheid bij windmast 1 tegen die van de overige windmasten is gegeven in Bijlage 1.



**Figuur 3-6:** Quotiëntplot van de windsnelheid bij windmast 1 en de windsnelheid bij windmast 6.

Duidelijk is in Figuur 3-6 te zien dat het nogal uitmaakt waar de wind vandaan komt. Zo valt op te maken dat het bij windrichting 130 ongeveer 30% harder waait bij windmast 1 t.o.v. windmast 6, vermoedelijk veroorzaakt door de aanwezigheid van een drietal grote hangars met een hoogte van circa 30 meter op betrekkelijk korte afstand ten zuidoosten van windmast 6. Bij windrichting 280 is het juist andersom en waait het ongeveer 10% minder hard bij windmast 1 t.o.v. windmast 6. Tussen de windrichtingen 130 en 280 bestaat dus een behoorlijk verschil. De lokale ruwheden hebben blijkbaar een behoorlijke invloed op de windsnelheid, waardoor een essentiële verbetering van het model mogelijk lijkt wanneer de windrichting wordt meegenomen in het model.

In dit onderzoek is de modellering van de windrichting gedaan met behulp van een discrete methode waarbij de windrichting geclusterd is in sectoren van 20 graden. Een andere methode die onderzocht is houdt de continuïteit intact met behulp van Fourier analyse.

### 3.4.2 Clusteren van windrichtingen

De discrete manier waarbij windrichting in het model wordt meegenomen maakt gebruik van clustering van de windrichting in sectoren van 20 graden. Elke sector ( $S_i$ ) beslaat dus een interval van 20 graden en wel op de volgende manier:

$$S_j = [(j \cdot 20) - 20, j \cdot 20], \quad j = 1, 2, \dots, 18. \quad (3-13)$$

Deze sector wordt in het model opgenomen door aan iedere sector een dummyvariabele  $d_{i,j}$  toe te kennen. Deze dummyvariabele wordt als volgt gedefinieerd:

$$d_{i,j} = \begin{cases} 1 & \text{als } D2_{i,t} \in S_j \\ 0 & \text{als } D2_{i,t} \notin S_j \end{cases} \quad (3-14)$$

waarbij  $D2_{i,t}$  staat voor de gemiddelde windrichting bij windmast  $i$  over de laatste twee minuten gezien vanaf tijdstip  $t$ .

Deze dummyvariabelen worden gebruikt om aan iedere sector een afzonderlijke hellingsparameter mee te geven.

Model F wordt op deze manier uitgebouwd tot:

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2}^6 \sum_{j=1}^{18} d_{i,j} \beta_{i,j} F2_{i,t} w_t + \varepsilon_t^{F1} w_t. \quad F1$$

De modellen B t/m E kunnen nog niet op dezelfde manier uitgebouwd worden tot B1 t/m E1, aangezien het mogelijk is dat de volgorde van toelating van de windmasten is veranderend nu de windrichting als factor is meegenomen in het model.

Voordat dit onderzocht wordt is het allereerst verstandig te onderzoeken in welke mate het betrekken van de windrichting in de modelbouw tot verbetering leidt per individuele verklarende windmast. Hiertoe kunnen de enkelvoudige lineaire regressiemodellen

$$F2_{1,t} w_t = \alpha w_t + \beta_i F2_{i,t} w_t + \varepsilon_t^{IIi} w_t, \quad i = 2, 3, \dots, 6 \quad IIi$$

uit Deelparagraaf 3.3.2 vergeleken worden met de modellen

$$F2_{1,t} w_t = \alpha w_t + \sum_{j=1}^{18} d_{i,j} \beta_{i,j} F2_{i,t} w_t + \varepsilon_t^{IIIi} w_t, \quad i = 2, 3, \dots, 6, \quad IIIi$$

waarin de dummyvariabelen zijn opgenomen en derhalve achttien verschillende coëfficiënten voor de hellingshoek kunnen worden verkregen. In hoeverre deze coëfficiënten ( $b_{i,j}$ ) afwijken van de coëfficiënt ( $b_i$ ) kan worden afgelezen in Tabel 3-16 waar de OLS resultaten staan afgebeeld van de modellen II6 en III6. Hierbij is de covariantiematrix wederom geschat met behulp van de Newey-West schatter.

Uit de tabel valt op te maken dat er sectoren zijn waarvan de bijbehorende coëfficiënten behoorlijk afwijken van de coëfficiënt uit model II6. Verder is in de tabel te zien dat er sectoren zijn die onderling sterk van elkaar verschillen. Het grootste verschil bestaat tussen de sectoren 7 en 14. Deze sectoren komen ongeveer overeen met de in Deelparagraaf 3.4.1 opgemerkte windrichtingen 130 en 280 op basis van de quotiëntplot. Met behulp van de 95%-betrouwbaarheidsintervallen van de coëfficiënten uit de tabel valt echter ook op te maken dat er sectoren zijn die niet significant van elkaar verschillen.

model	parameter	coëfficiënt	standaardfout coëfficiënt	t-waarde coëfficiënt	95% betrouwbaar- heidsinterval coëfficiënt
II6	$\beta_6$	0.879	$0.63 \cdot 10^{-2}$	139.53	$\langle 0.867, 0.891 \rangle$
III6	$\beta_{6,1}$	0.864	$1.56 \cdot 10^{-2}$	55.43	$\langle 0.833, 0.894 \rangle$
	$\beta_{6,2}$	0.869	$1.44 \cdot 10^{-2}$	60.52	$\langle 0.840, 0.897 \rangle$
	$\beta_{6,3}$	0.891	$1.02 \cdot 10^{-2}$	87.43	$\langle 0.871, 0.911 \rangle$
	$\beta_{6,4}$	0.948	$1.12 \cdot 10^{-2}$	84.59	$\langle 0.926, 0.970 \rangle$
	$\beta_{6,5}$	0.933	$1.22 \cdot 10^{-2}$	76.70	$\langle 0.909, 0.956 \rangle$
	$\beta_{6,6}$	1.010	$1.49 \cdot 10^{-2}$	67.99	$\langle 0.981, 1.039 \rangle$
	$\beta_{6,7}$	1.117	$1.82 \cdot 10^{-2}$	61.57	$\langle 1.082, 1.153 \rangle$
	$\beta_{6,8}$	1.082	$1.94 \cdot 10^{-2}$	55.76	$\langle 1.044, 1.120 \rangle$
	$\beta_{6,9}$	0.913	$0.75 \cdot 10^{-2}$	121.46	$\langle 0.899, 0.928 \rangle$
	$\beta_{6,10}$	0.905	$0.76 \cdot 10^{-2}$	119.41	$\langle 0.890, 0.920 \rangle$
	$\beta_{6,11}$	0.842	$0.62 \cdot 10^{-2}$	136.00	$\langle 0.829, 0.854 \rangle$
	$\beta_{6,12}$	0.893	$0.66 \cdot 10^{-2}$	134.54	$\langle 0.880, 0.906 \rangle$
	$\beta_{6,13}$	0.830	$0.58 \cdot 10^{-2}$	142.20	$\langle 0.818, 0.841 \rangle$
	$\beta_{6,14}$	0.793	$0.60 \cdot 10^{-2}$	132.02	$\langle 0.781, 0.805 \rangle$
	$\beta_{6,15}$	0.799	$0.84 \cdot 10^{-2}$	95.09	$\langle 0.782, 0.815 \rangle$
	$\beta_{6,16}$	0.819	$0.87 \cdot 10^{-2}$	94.08	$\langle 0.802, 0.836 \rangle$
	$\beta_{6,17}$	0.865	$1.15 \cdot 10^{-2}$	75.05	$\langle 0.842, 0.887 \rangle$
$\beta_{6,18}$	0.876	$1.48 \cdot 10^{-2}$	59.23	$\langle 0.847, 0.905 \rangle$	

**Tabel 3-16:** Regressieresultaten uit de modellen II6 en III6.

Om deze reden is het verstandig om een clustering aan te brengen in de sectoren. Dit kan gedaan worden door het toetsen van diverse lineaire restricties op de parameters met behulp van de Wald-statistic (zie Greene [1997, pp. 303]).

Elke set van lineaire restricties op de parameters kan in matrixnotatie als volgt als nulhypothese worden opgeschreven:

$$H_0: \mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}, \quad (3-15)$$

waarbij  $\mathbf{R}$  een bekende matrix is met het aantal restricties  $g$  als aantal kolommen en het aantal parameters  $k$  (inclusief die van de constante) als aantal rijen,  $\boldsymbol{\beta}$  de vector met de parameters en  $\mathbf{q}$  een bekende vector met  $g$  elementen.

De Wald-statistic voor het toetsen van  $H_0$  ziet er als volgt uit:

$$W = (\mathbf{R}\mathbf{b} - \mathbf{q})' [\mathbf{R} \cdot \text{geschatte Cov}(\mathbf{b}) \cdot \mathbf{R}'] (\mathbf{R}\mathbf{b} - \mathbf{q}) \xrightarrow{d} \chi^2[g], \quad (3-16)$$

waarbij  $\mathbf{b}$  staat voor de vector met coëfficiënten (inclusief constante).

Er bestaat ook een soortgelijke  $F$ -statistic voor het toetsen van  $H_0$ . Deze kan echter hier niet gebruikt worden omdat hij gebruik maakt van de covariantiematrix zoals die op de oorspronkelijke manier wordt geschat en dus niet volgens de Newey-West schatter.

Een algoritme voor het clusteren is nu als volgt:

- 1) Sorteren  $b_{6,j}$  ( $j = 1, 2, \dots, 18$ ) van groot naar klein:  $b_{6,[1]}, b_{6,[2]}, \dots, b_{6,[18]}$ ,
- 2)  $j := 1$ ,
- 3) Toetsen restrictie  $\beta_{6,[j]} = \beta_{6,[j+1]}$ ,

- 4) Als restrictie in 3) niet wordt verworpen: Sectoren  $S_{[j]}$  en  $S_{[j+1]}$  samenvoegen, vervolgens gezamenlijke parameter toetsen op gelijkheid met  $\beta_{6,[j+2]}$ . Als deze restrictie niet wordt verworpen:  $S_{[j+2]}$  samenvoegen met de zojuist samengevoegde sectoren en toetsen op gelijkheid met  $\beta_{6,[j+3]}$  enz, anders:  $j:=j+2$  en ga naar 3),
- 5) Als restrictie in 3) wordt verworpen:  $j:=j+1$  en ga naar 3).

Bij de eerste keer dat 3) wordt aangedaan ziet  $H_0$  er als volgt uit:

$$H_0: \begin{bmatrix} \alpha \\ \beta_{[1]} \\ \beta_{[2]} \\ \vdots \\ \beta_{[18]} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (3-17)$$

Op dezelfde wijze kunnen de andere nulhypotesen worden opgesteld.

Nadat het algoritme is doorlopen blijken de volgende clusters te ontstaan:

$(S_7, S_8)$ ,  $(S_6)$ ,  $(S_4, S_5)$ ,  $(S_9, S_{10}, S_{12}, S_3)$ ,  $(S_{18}, S_2, S_{17}, S_1)$ ,  $(S_{11}, S_{13}, S_{16})$  en  $(S_{15}, S_{14})$ .

De dummyvariabelen kunnen nu als volgt gedefinieerd worden:

$$d_{6,7} = d_{6,8} = \begin{cases} 1 & \text{als } D2_{6,t} \in S_7 \cup S_8 \\ 0 & \text{als } D2_{6,t} \notin S_7 \cup S_8 \end{cases},$$

$$d_{6,6} = \begin{cases} 1 & \text{als } D2_{6,t} \in S_6 \\ 0 & \text{als } D2_{6,t} \notin S_6 \end{cases},$$

$$d_{6,4} = d_{6,5} = \begin{cases} 1 & \text{als } D2_{6,t} \in S_4 \cup S_5 \\ 0 & \text{als } D2_{6,t} \notin S_4 \cup S_5 \end{cases},$$

$$d_{6,9} = d_{6,10} = d_{6,12} = d_{6,3} = \begin{cases} 1 & \text{als } D2_{6,t} \in S_9 \cup S_{10} \cup S_{12} \cup S_3 \\ 0 & \text{als } D2_{6,t} \notin S_9 \cup S_{10} \cup S_{12} \cup S_3 \end{cases}, \quad (3-18)$$

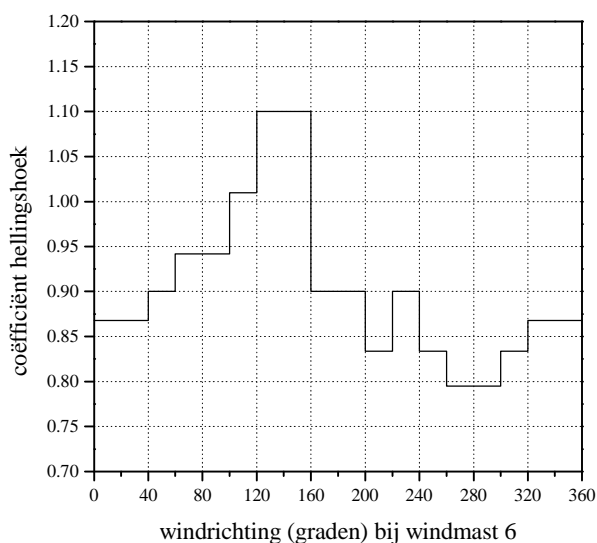
$$d_{6,18} = d_{6,2} = d_{6,17} = d_{6,1} = \begin{cases} 1 & \text{als } D2_{6,t} \in S_{18} \cup S_2 \cup S_{17} \cup S_1 \\ 0 & \text{als } D2_{6,t} \notin S_{18} \cup S_2 \cup S_{17} \cup S_1 \end{cases},$$

$$d_{6,11} = d_{6,13} = d_{6,16} = \begin{cases} 1 & \text{als } D2_{6,t} \in S_{11} \cup S_{13} \cup S_{16} \\ 0 & \text{als } D2_{6,t} \notin S_{11} \cup S_{13} \cup S_{16} \end{cases},$$

$$d_{6,15} = d_{6,14} = \begin{cases} 1 & \text{als } D2_{6,t} \in S_{15} \cup S_{14} \\ 0 & \text{als } D2_{6,t} \notin S_{15} \cup S_{14} \end{cases}.$$

Het beschreven algoritme is ook doorlopen voor de andere windmasten (2 t/m 5).

Een grafische presentatie van het verloop van  $b_{i,j}$  voor  $i = 6$  (windmast 6) is weergegeven in onderstaand figuur:



**Figuur 3-7:** Coëfficiëntplot van hellingshoek bij clusteren van windrichting voor windmast 6.

Een overzicht van de coëfficiëntplots voor de andere windmasten (2 t/m 5) is te zien in Bijlage 2. Hierin is duidelijk te zien dat de coëfficiënten in grote lijnen de quotiëntplots van Bijlage 1 volgen.

Nu de dummyvariabelen gheherdefinieerd zijn kunnen de modellen  $II_i$  en  $III_i$  ( $i = 2, 3, \dots, 6$ ) met elkaar vergeleken worden. Dit wordt gedaan middels opnieuw de *RMSE*-waarde van de modellen. Voor een betere interpretatie van de opgetreden percentuele verbeteringen in de *RMSE*-waarde zijn opnieuw de regressieresultaten toegepast op de validatiedata.

De resultaten hiervan staan in Tabel 3-17:

modeldata				validatiedata			
$i$	<i>RMSE</i> $II_i$	<i>RMSE</i> $III_i$	verandering (%) $III_i$ t.o.v. $II_i$	$i$	<i>RMSE</i> $II_i$	<i>RMSE</i> $III_i$	verandering (%) $III_i$ t.o.v. $II_i$
2	1.061	1.014	-4.4	2	1.063	1.028	-3.3
3	1.075	1.044	-2.9	3	1.095	1.072	-2.1
4	1.115	1.006	-9.8	4	1.140	1.047	-8.2
5	1.017	0.931	-8.5	5	1.005	0.932	-7.3
6	1.144	1.062	-7.2	6	1.152	1.077	-6.5

**Tabel 3-17:** Vergelijking enkelvoudige regressiemodellen zonder windrichting opgenomen ( $II_i$ ) en met windrichting opgenomen ( $III_i$ ).

Uit de tabel valt allereerst op te maken dat de percentuele verbeteringen bij de modelbouw behoorlijk wat groter zijn dan wanneer de modelresultaten worden toegepast op de validatiedata. Dit is niet verwonderlijk gezien het kleine aantal parameters bij de modellen  $II_i$  (telkens 2) ten opzichte van het aantal parameters bij de modellen  $III_i$  (resp. 8, 7, 12, 8 en nogmaals 8). Hierdoor is de overfit in de modellen  $III_i$  relatief groot ten opzichte van die in de modellen  $II_i$  en zijn de werkelijke verbeteringen kleiner dan aangegeven staat in het deel van Tabel 3-17 bij de modeldata.

Verder zien we dat bij de windmasten 4 en 5 de grootste verbetering plaats vindt, gevolgd door windmast 6. Bij de windmasten 2 en 3 vindt de kleinste verandering plaats.

Zoals eerder is aangegeven kunnen deze resultaten tot gevolg hebben dat de volgorde van toevoeging van de masten is veranderd. De optimale volgorde van toevoeging zonder rekening te houden met de windrichting was resp. 5, 2, 3, 4 en 6 (zie Deelparagraaf 3.3.3). Wellicht is deze optimale volgorde veranderd nu er wel rekening gehouden wordt met de windrichting. Of dit werkelijk het geval is, is nagegaan en hieruit blijkt dat de windmasten 3 en 4 van positie zijn gewisseld. Dit is niet erg verwonderlijk aangezien de verbetering bij windmast 4 een stuk hoger is dan bij windmast 3. De volgorde van toelating is derhalve resp. 5, 2, 4, 3 en 6, waardoor de volgende modellen nader worden geanalyseerd:

$$F2_{1,t} = F2_{5,t} + \varepsilon_t^A, \quad A$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=5}^{18} \sum_{j=1}^{18} d_{i,j} \beta_{i,j} F2_{i,t} w_t + \varepsilon_t^{B1} w_t, \quad B1$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2,5}^{18} \sum_{j=1}^{18} d_{i,j} \beta_{i,j} F2_{i,t} w_t + \varepsilon_t^{C1} w_t, \quad C1$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2,4,5}^{18} \sum_{j=1}^{18} d_{i,j} \beta_{i,j} F2_{i,t} w_t + \varepsilon_t^{D1} w_t, \quad D1$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2,3,4,5}^{18} \sum_{j=1}^{18} d_{i,j} \beta_{i,j} F2_{i,t} w_t + \varepsilon_t^{E1} w_t, \quad E1$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2}^6 \sum_{j=1}^{18} d_{i,j} \beta_{i,j} F2_{i,t} w_t + \varepsilon_t^{F1} w_t. \quad F1$$

De kracht van deze modellen kan opnieuw bekeken worden met behulp van de *RMSE*-waarde. Tevens kan er weer gekeken worden naar de stapsgewijze verbetering. De resultaten hiervan staan in onderstaande tabel:

modeldata			validatiedata		
model	<i>RMSE</i>	totale percentuele verandering <i>RMSE</i>	model	<i>RMSE</i>	totale percentuele verandering <i>RMSE</i>
A	1.023	–	A	1.011	–
B1	0.931	9.0	B1	0.932	7.8
C1	0.836	18.3	C1	0.836	17.3
D1	0.800	21.8	D1	0.808	20.1
E1	0.790	22.8	E1	0.801	20.8
F1	0.785	23.3	F1	0.795	21.4

**Tabel 3-18:** Stapsgewijze verandering in de *RMSE* vanaf het meest optimale back-up model tot en met het meervoudige regressiemodel waarin alle masten zijn opgenomen via de ‘stepwise selection’ procedure. De modellen zijn uitgebreid door tevens de windrichting te modelleren.

Het uiteindelijke model (F1) geeft een *RMSE* van 0.795. Het uiteindelijk model zonder windrichting als factor (F) gaf een *RMSE* van 0.848. De verbetering die optreedt van model F naar F1 is derhalve 6.3%. Dit lijkt niet zo veel, maar dit komt doordat dit

percentage gemiddeld is naar windrichting. Er zullen windrichtingen zijn waarbij de verbetering meer dan 6.3% bedraagt, maar ook windrichtingen waarbij de verbetering minder dan 6.3% bedraagt. Dit is onderzocht en de resultaten staan hieronder:

validatiedata					
windrichting (graden)	RMSE			verandering (%)	
	A	F	F1	F1 t.o.v. A	F1 t.o.v. F
0-20	0.884	0.750	0.752	-14.9	0.3
20-40	0.897	0.815	0.765	-14.7	-6.1
40-60	0.947	0.853	0.804	-15.1	-5.7
60-80	1.013	0.837	0.831	-18.0	-0.7
80-100	0.949	0.747	0.702	-26.1	-6.1
100-120	0.827	0.696	0.645	-22.1	-7.4
120-140	0.935	0.889	0.726	-22.4	-18.4
140-160	0.701	0.732	0.653	-6.9	-10.8
160-180	0.664	0.649	0.607	-8.7	-6.5
180-200	0.830	0.718	0.722	-13.1	0.5
200-220	1.314	0.871	0.834	-36.5	-4.3
220-240	1.264	0.987	0.956	-24.3	-3.2
240-260	1.112	0.945	0.910	-18.2	-3.7
260-280	1.095	1.030	0.860	-21.4	-16.4
280-300	0.954	0.861	0.799	-16.3	-7.2
300-320	1.028	0.894	0.809	-21.3	-9.5
320-340	1.047	0.884	0.882	-15.7	-0.2
340-360	0.737	0.667	0.667	-9.4	0.0

**Tabel 3-19:** Verbetering in *RMSE* per sector windrichting.

Uit de laatste kolom van Tabel 3-19 kan inderdaad afgeleid worden dat de verbetering van het model zonder modellering van windrichting (model F) naar het model met modellering van windrichting (model F1) nogal varieert per windrichtingsector. Ook is dit patroon te zien in vergelijking met het optimale back-up model (model A).

Of de verbetering in modelkracht ook optreedt bij oplopende windsnelheden kan getoetst worden op basis van Tabel 3-20. Hierin is af te lezen dat per interval de toename ongeveer gelijk is. De geconstateerde verbetering blijkt dus over het gehele spectrum van windsnelheden ongeveer even groot te zijn.

validatiedata					
windsnelheid (m/s)	RMSE			verandering (%)	
	A	F	F1	F t.o.v. A	F1 t.o.v. A
2-4	0.669	0.604	0.587	-9.7	-12.3
4-6	0.858	0.753	0.710	-12.2	-17.2
6-8	1.171	0.981	0.897	-16.2	-23.4
8-10	1.434	1.141	1.073	-20.4	-25.2
10-12	1.615	1.276	1.182	-21.0	-26.8
>12	2.036	1.540	1.417	-24.4	-30.4

**Tabel 3-20:** Verbetering in *RMSE* per interval van 2 m/s.

### 3.4.3 Windrichting modelleren met Fourier analyse

Wanneer er geclusterd wordt naar windrichting kan het zijn dat twee opeenvolgende sectoren een groot verschil te zien geven in hun coëfficiënt. Stel dat dit het geval is met de sectoren  $S_1$  en  $S_2$ . Wanneer het model uitgevoerd wordt zal het derhalve bij een windrichting van 19.9 graden de coëfficiënt gebruiken behorende bij sector 1 en bij een windrichting van 20.1 de coëfficiënt van sector 2. In dit geval kan men echter niet verwachten dat het verschil tussen dermate kleine veranderingen in windrichting proporties aanneemt zoals de gemiddelde situatie in sector 1 en 2. Deze manier om de windrichting in het model te betrekken is dus tamelijk grof. Het betrekken van de windrichting tot het model kan worden verfijnd door een continue functie van de invloed van de windrichting op de windsnelheid te definiëren. Gezien het quotiëntplot van Figuur 3-6 lijkt een sinuïde/cosinuïde-fit wellicht redelijk te volstaan. Deze functies kunnen verwerkt worden in het model door een aantal sinus-/cosinusfuncties met verschillende periodelengten toe te voegen aan het regressiemodel waarna vervolgens wordt bekeken welke van deze functies significant zijn. De functies die niet significant blijken te zijn kunnen dan achterwege gelaten worden.

Het bepalen welke functies significant zijn kan worden gedaan door te kijken naar de volgende modellen:

$$F2_{1,t} w_t = \alpha w_t + \beta_i F2_{i,t} w_t + \sum_{j=1}^{\infty} \gamma_{i,j} \cos(jD2_{i,t}) F2_{i,t} w_t + \sum_{k=1}^{\infty} \delta_{i,k} \sin(kD2_{i,t}) F2_{i,t} w_t + \varepsilon_t^{IVi} w_t, \quad i = 2, 3, \dots, 6, \quad IVi$$

waarbij  $D2_{i,t}$  opnieuw staat voor de gemiddelde windrichting bij windmast  $i$  over de laatste twee minuten gezien vanaf tijdstip  $t$ .

Vervolgens kan opnieuw gebruik gemaakt worden van stepwise selection, de in Deelparagraaf 3.3.3 besproken selectieprocedure voor het opnemen en verwijderen van variabelen in een regressiemodel. Opnieuw is het hier belangrijk dat de standaardfouten van de coëfficiënten geschat worden met behulp van de Newey-West schatter, aangezien wanneer dit niet gebeurt bepaalde functies wellicht als significant worden beschouwd terwijl dit in werkelijkheid niet het geval is door de aanwezige heteroskedasticiteit en autocorrelatie.

Nadat de analyse op de modellen is uitgevoerd (functies met  $j$  groter dan 20 zijn hierbij niet opgenomen) blijken vrij veel functies significant te zijn. Hoewel significant blijkt een aantal hiervan nauwelijks enige bijdrage toe te voegen aan het model. Derhalve wordt een criterium gehanteerd om de stepwise selection op een gegeven moment te onderbreken, ook al kan op basis van significantie meerdere functies worden toegevoegd. Wanneer de *RMSE* (op drie decimalen afgerond) na toevoeging van een functie niet meer kleiner wordt, dan wordt de stepwise selection gestaakt.

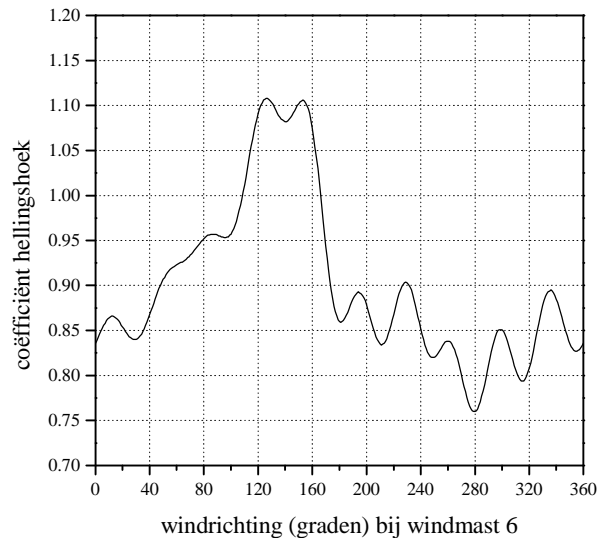
Op de volgende pagina staat een overzicht van de overgebleven cosinus-/sinusfuncties. De parameters ( $\gamma_{i,j}$  en  $\delta_{i,k}$ ) van de cosinus-/sinusfuncties die niet in dit overzicht vermeld staan worden op nul gesteld.

model	cosinusfuncties	sinusfuncties
IV2	$\cos(1D_{2,t})F_{2,t} w_t$ $\cos(2D_{2,t})F_{2,t} w_t$ $\cos(3D_{2,t})F_{2,t} w_t$ $\cos(5D_{2,t})F_{2,t} w_t$	$\sin(1D_{2,t})F_{2,t} w_t$ $\sin(2D_{2,t})F_{2,t} w_t$ $\sin(4D_{2,t})F_{2,t} w_t$ $\sin(5D_{2,t})F_{2,t} w_t$ $\sin(7D_{2,t})F_{2,t} w_t$ $\sin(9D_{2,t})F_{2,t} w_t$
IV3	$\cos(1D_{3,t})F_{3,t} w_t$ $\cos(2D_{3,t})F_{3,t} w_t$ $\cos(3D_{3,t})F_{3,t} w_t$ $\cos(5D_{3,t})F_{3,t} w_t$ $\cos(6D_{3,t})F_{3,t} w_t$	$\sin(1D_{3,t})F_{3,t} w_t$ $\sin(2D_{3,t})F_{3,t} w_t$ $\sin(4D_{3,t})F_{3,t} w_t$ $\sin(7D_{3,t})F_{3,t} w_t$ $\sin(9D_{3,t})F_{3,t} w_t$ $\sin(15D_{3,t})F_{3,t} w_t$
IV4	$\cos(1D_{4,t})F_{4,t} w_t$ $\cos(2D_{4,t})F_{4,t} w_t$ $\cos(5D_{4,t})F_{4,t} w_t$ $\cos(7D_{4,t})F_{4,t} w_t$ $\cos(10D_{4,t})F_{4,t} w_t$	$\sin(1D_{4,t})F_{4,t} w_t$ $\sin(2D_{4,t})F_{4,t} w_t$ $\sin(4D_{4,t})F_{4,t} w_t$ $\sin(5D_{4,t})F_{4,t} w_t$ $\sin(6D_{4,t})F_{4,t} w_t$ $\sin(15D_{4,t})F_{4,t} w_t$
IV5	$\cos(1D_{5,t})F_{5,t} w_t$ $\cos(3D_{5,t})F_{5,t} w_t$ $\cos(4D_{5,t})F_{5,t} w_t$ $\cos(5D_{5,t})F_{5,t} w_t$ $\cos(6D_{5,t})F_{5,t} w_t$ $\cos(12D_{5,t})F_{5,t} w_t$ $\cos(14D_{5,t})F_{5,t} w_t$ $\cos(17D_{5,t})F_{5,t} w_t$	$\sin(2D_{5,t})F_{5,t} w_t$ $\sin(3D_{5,t})F_{5,t} w_t$ $\sin(6D_{5,t})F_{5,t} w_t$ $\sin(8D_{5,t})F_{5,t} w_t$ $\sin(9D_{5,t})F_{5,t} w_t$ $\sin(20D_{5,t})F_{5,t} w_t$
IV6	$\cos(1D_{6,t})F_{6,t} w_t$ $\cos(3D_{6,t})F_{6,t} w_t$ $\cos(4D_{6,t})F_{6,t} w_t$ $\cos(5D_{6,t})F_{6,t} w_t$ $\cos(8D_{6,t})F_{6,t} w_t$ $\cos(10D_{6,t})F_{6,t} w_t$ $\cos(11D_{6,t})F_{6,t} w_t$	$\sin(1D_{6,t})F_{6,t} w_t$ $\sin(2D_{6,t})F_{6,t} w_t$ $\sin(3D_{6,t})F_{6,t} w_t$ $\sin(4D_{6,t})F_{6,t} w_t$ $\sin(10D_{6,t})F_{6,t} w_t$

Aan de hand van de geschatte waarden van  $\beta_i$ ,  $\gamma_{i,j}$  en  $\delta_{i,k}$  (resp.  $b_i$ ,  $g_{i,j}$  en  $d_{i,k}$ ) kan de windrichtingafhankelijke coëfficiënt

$$b_i + \sum_{j=1}^{\infty} g_{i,j} \cos(jD2_{i,t}) + \sum_{k=1}^{\infty} d_{i,k} \sin(kD2_{i,t})$$

berekend worden. Een grafische presentatie hiervan voor  $i = 6$  is weergegeven in onderstaand figuur:



**Figuur 3-8:** Coëfficiëntplot van hellingshoek bij Fourier modellering voor windmast 6.

Een overzicht van de coëfficiëntplots voor de andere windmasten (2 t/m 5) is te zien in Bijlage 3. Hierin is opnieuw duidelijk te zien dat de functies in grote lijnen de quotiëntplots van Bijlage 1 volgen. In Bijlage 4 zijn de coëfficiënten zoals die berekend zijn via clustering en Fourier analyse over elkaar heen gelegd, zodat ze beter met elkaar vergeleken kunnen worden.

Nu bekend is welke cosinus-/sinusfuncties een fundamentele bijdrage leveren aan het model kan bekeken worden of en in welke mate toevoeging van deze functies beter voldoet dan het clusteren naar windrichting. Hiertoe kunnen de *RMSE*-waarden van de modellen III*i* en IV*i* met elkaar vergeleken worden. Immers, III*i* betrok de windrichting in het model door te clusteren naar windrichting en IV*i* betrok de windrichting in het model door windrichtingafhankelijke cosinus-/sinusfuncties te definiëren. Een overzicht van de resultaten staat in Tabel 3-21.

Uit deze tabel volgt dat de *RMSE*-waarden in de modellen IV*i* telkens kleiner zijn dan die van de modellen III*i*. Het gebruikmaken van cosinus-/sinusfuncties lijkt dus meer geschikt om de windrichting te modelleren dan wanneer er wordt geclusterd naar windrichting.

modeldata					
$i$	$RMSE_{\text{III}i}$	$RMSE_{\text{III}i}$	$RMSE_{\text{IV}i}$	verandering (%) IIIi t.o.v. IIIi	verandering (%) IVi t.o.v. IIIi
2	1.061	1.014	1.009	4.4	4.9
3	1.075	1.044	1.041	2.9	3.2
4	1.115	1.006	0.998	9.8	10.5
5	1.017	0.931	0.919	8.5	9.6
6	1.144	1.062	1.054	7.2	7.9

validatiedata					
$i$	$RMSE_{\text{III}i}$	$RMSE_{\text{III}i}$	$RMSE_{\text{IV}i}$	verandering (%) IIIi t.o.v. IIIi	verandering (%) IVi t.o.v. IIIi
2	1.063	1.028	1.022	3.3	3.9
3	1.095	1.072	1.070	2.1	2.3
4	1.140	1.047	1.040	8.2	8.8
5	1.005	0.932	0.920	7.3	8.5
6	1.152	1.077	1.071	6.5	7.0

**Tabel 3-21:** *RMSE*-waarden van de enkelvoudige regressiemodellen zonder en met windrichting en de percentuele verschillen tussen deze modellen.

In hoeverre deze methode beter voldoet in het geval wanneer er meerdere masten in het model worden opgenomen, kan bekeken worden door de modellen B1 t/m F1 aan te passen tot:

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=5} \left( \left\{ \beta_i + \sum_{j=1}^{\infty} \gamma_{i,j} \cos(jD2_{i,t}) + \sum_{k=1}^{\infty} \delta_{i,k} \sin(kD2_{i,t}) \right\} F2_{i,t} w_t \right) + \varepsilon_t^{\text{B2}} w_t, \quad \text{B2}$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2,5} \left( \left\{ \beta_i + \sum_{j=1}^{\infty} \gamma_{i,j} \cos(jD2_{i,t}) + \sum_{k=1}^{\infty} \delta_{i,k} \sin(kD2_{i,t}) \right\} F2_{i,t} w_t \right) + \varepsilon_t^{\text{C2}} w_t, \quad \text{C2}$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2,4,5} \left( \left\{ \beta_i + \sum_{j=1}^{\infty} \gamma_{i,j} \cos(jD2_{i,t}) + \sum_{k=1}^{\infty} \delta_{i,k} \sin(kD2_{i,t}) \right\} F2_{i,t} w_t \right) + \varepsilon_t^{\text{D2}} w_t, \quad \text{D2}$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2,3,4,5} \left( \left\{ \beta_i + \sum_{j=1}^{\infty} \gamma_{i,j} \cos(jD2_{i,t}) + \sum_{k=1}^{\infty} \delta_{i,k} \sin(kD2_{i,t}) \right\} F2_{i,t} w_t \right) + \varepsilon_t^{\text{E2}} w_t, \quad \text{E2}$$

$$F2_{1,t} w_t = \alpha w_t + \sum_{i=2}^6 \left( \left\{ \beta_i + \sum_{j=1}^{\infty} \gamma_{i,j} \cos(jD2_{i,t}) + \sum_{k=1}^{\infty} \delta_{i,k} \sin(kD2_{i,t}) \right\} F2_{i,t} w_t \right) + \varepsilon_t^{\text{F2}} w_t. \quad \text{F2}$$

Deze modellen zijn geschat met OLS en hun *RMSE*-waarden kunnen vergeleken worden met die van B1 t/m F1. Dit is weergegeven in Tabel 3-22.

modeldata		
model	<i>RMSE</i>	totale percentuele verandering <i>RMSE</i>
A	1.023	–
B1	0.930	-9.1
C1	0.836	-18.3
D1	0.800	-21.8
E1	0.790	-22.8
F1	0.785	-23.3

modeldata		
model	<i>RMSE</i>	totale percentuele verandering <i>RMSE</i>
A	1.023	–
B2	0.919	-10.2
C2	0.827	-19.2
D2	0.793	-22.5
E2	0.783	-23.5
F2	0.779	-23.9

validatiedata		
model	<i>RMSE</i>	totale percentuele verandering <i>RMSE</i>
A	1.011	–
B1	0.932	-7.8
C1	0.836	-17.3
D1	0.808	-20.1
E1	0.801	-20.8
F1	0.795	-21.4

validatiedata		
model	<i>RMSE</i>	totale percentuele verandering <i>RMSE</i>
A	1.011	–
B2	0.920	-9.0
C2	0.828	-18.1
D2	0.802	-20.7
E2	0.794	-21.5
F2	0.789	-22.0

**Tabel 3-22:** Stapsgewijze verandering in de *RMSE* vanaf het meest optimale back-up model tot en met het meervoudige regressiemodel waarin alle masten zijn opgenomen via de stepwise selection procedure. De modellen zijn uitgebreid door tevens de windrichting te modelleren via clustering (B1 t/m F1) en via Fourier analyse (B2 t/m F2).

De modellen B2 t/m F2 hebben allemaal een kleinere *RMSE* dan de modellen B1 t/m F1, maar wat duidelijk naar voren komt is dat het verschil in *RMSE* kleiner wordt naarmate meer masten zijn toegevoegd aan het model. Bij het uiteindelijke model maakt het daarom niet veel uit of er geclusterd is naar windrichting (model F1) dan wel of er gebruik gemaakt is van cosinus-/sinusfuncties (model F2). De uiteindelijke verbetering van model F2 ten opzichte van model F1 is gelijk aan 0.8% (0.789 t.o.v. 0.795). Dit percentage is dus niet bijzonder groot, maar het intact houden van de continuïteit is wel een belangrijk voordeel van deze methode in vergelijking met de methode van clusteren. Of het percentage van 0.8% constant is per windrichtingsector kan geconcludeerd worden op basis van Tabel 3-23.

Uit deze tabel is op te maken dat de verschillen tussen F1 en F2 per windrichtingsector niet erg groot zijn en globaal hetzelfde beeld geven dat model F2 iets beter voldoet dan model F1.

Op wat voor manier model F2 zich verhoudt ten opzichte van model F1 bij oplopende windsnelheden kan opgemaakt worden uit Tabel 3-24. Uit deze tabel komt naar voren dat bij hogere windsnelheden het model met Fourier analyse nog wat beter voldoet dan bij lagere windsnelheden. Dit zijn echter minieme verschillen.

validatiedata			
windrichting (graden)	RMSE		verandering (%) F2 t.o.v. F1
	F1	F2	
0-20	0.756	0.745	-1.5
20-40	0.774	0.779	0.6
40-60	0.802	0.795	-0.8
60-80	0.829	0.827	-0.2
80-100	0.699	0.707	1.1
100-120	0.646	0.652	0.8
120-140	0.725	0.723	-0.2
140-160	0.657	0.656	-0.3
160-180	0.605	0.605	-0.1
180-200	0.722	0.711	-1.5
200-220	0.832	0.823	-1.1
220-240	0.948	0.942	-0.6
240-260	0.908	0.907	-0.1
260-280	0.863	0.860	-0.3
280-300	0.800	0.799	-0.0
300-320	0.808	0.799	-1.1
320-340	0.873	0.850	-2.7
340-360	0.666	0.665	-1.6

**Tabel 3-23:** Vergelijking *RMSE* tussen F1 en F2 per sector windrichting.

validatiedata			
windsnelheid (m/s)	RMSE		verandering (%) F2 t.o.v. F1
	F1	F2	
2-4	0.587	0.582	-0.9
4-6	0.710	0.708	-0.3
6-8	0.897	0.893	-0.5
8-10	1.073	1.060	-1.2
10-12	1.182	1.167	-1.3
>12	1.417	1.403	-1.0

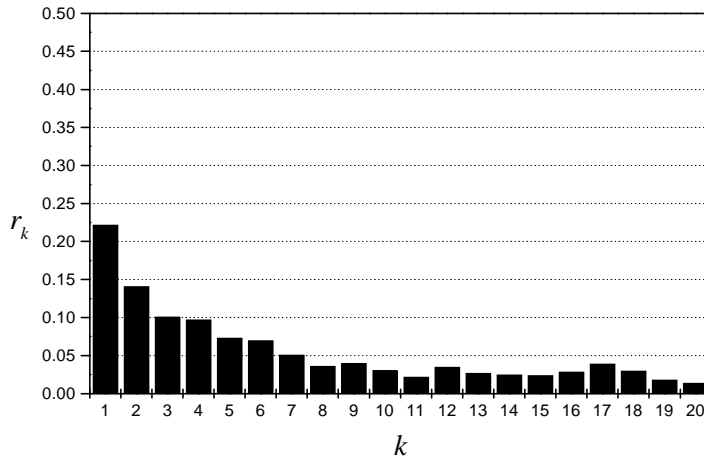
**Tabel 3-24:** Vergelijking *RMSE* tussen F1 en F2 per interval van 2 m/s.

### 3.4.4 Mate van vermindering autocorrelatie

In Deelparagraaf 3.2.2 werd vastgesteld dat er sprake was van autocorrelatie in model (3-1). Dit duidt op een gebrekkige modellering van de windsnelheid bij windmast 1 in dit model. Nu de windrichting een duidelijke toegevoegde waarde blijkt te hebben in het model kan getoetst worden of de autocorrelatie is verminderd.

Hiertoe kan opnieuw een autocorrelatie functie (ACF) opgesteld worden maar nu op basis van de residuen uit model F2. Het resultaat hiervan staat in Figuur 3-9:

Als dit figuur vergeleken wordt met Figuur 3-2 dan wordt duidelijk dat de autocorrelatie inderdaad sterk is verminderd wat erop duidt dat een groot gedeelte van tot voorheen onverklaarbare variantie in de windsnelheid bij windmast 1 wel verklaard wordt als er rekening gehouden wordt met de windrichting. Deze constatering geeft opnieuw aan dat het modelleren van windrichting een duidelijke toegevoegde waarde heeft.



**Figuur 3-9:** ACF van de residuen uit model F2.

De resterende autocorrelatie kan mogelijk verklaard worden door het in Deelparagraaf 3.2.2 besproken weerbegrip convectie. Dit verschijnsel is echter moeilijk in de modelbouw te verwerken, waardoor een verdere reductie van de autocorrelatie niet mogelijk lijkt.

### 3.4.5 Gevolgen van beschuttingsverschillen tussen windmasten na correctie

In de voorafgaande deelparagrafen is aangetoond dat het verstandig is om rekening te houden met lokale beschuttingsfactoren voor het maken van voorspellingen over de 2-minuut-gemiddelde winsnelheid bij een niet-functionerende windmast met behulp van metingen afkomstig van de overige windmasten op een luchthaven. Hierbij is rekening gehouden met lokale beschuttingsverschillen door de windrichting te modelleren. De beste manier om dit te doen bleek tot stand te komen middels Fourier analyse.

Niet aan de orde is geweest is of de modellering ook daadwerkelijk de bestaande beschuttingsverschillen in zijn geheel heeft meegenomen in de modellering. Verder kan het wellicht zo zijn dat, ook al worden de bestaande beschuttingsverschillen in zijn geheel gemodelleerd, een uitspraak over een mast minder goed is wanneer de beschuttingsverschillen met de verklarende masten groot zijn ten opzichte van een uitspraak over een mast waarbij de beschuttingsverschillen met de verklarende masten klein zijn.

In feite zijn we dus geïnteresseerd in de vraag of een gedeelte van de onbeschreven variatie in de *RMSE*-waarden (mate van modelkracht) van de modellen waarin de windrichting is opgenomen alsnog beschreven kan worden door de mate van beschuttingsverschillen tussen de windmasten.

Bovenstaande wordt nagegaan door allereerst een model te construeren waarin getracht wordt de *RMSE*-waarden van de modellen waarin niet de windrichting is meegenomen te verklaren. Bekeken wordt de koppeling tussen twee masten zoals de enkelvoudige lineaire regressiemodellen in Deelparagraaf 3.3.2 (model II2 t/m II6). Hierbij werd windmast 1 als afhankelijke windmast gebruikt en werden zodoende 5 waarden van de *RMSE* gegenereerd. Dit aantal is vrij klein voor het opstellen van een regressiemodel. De andere windmasten kunnen echter ook als afhankelijke windmasten worden gebruikt. Het totaal aantal *RMSE*-waarden komt hierdoor op 30. Een overzicht hiervan is gegeven in de matrix op de volgende pagina:

$$\mathbf{P} = \begin{pmatrix} 0 & 1.061 & 1.075 & 1.115 & 1.017 & 1.144 \\ 1.056 & 0 & 1.004 & 1.035 & 1.104 & 0.855 \\ 0.977 & 0.925 & 0 & 0.942 & 1.010 & 0.945 \\ 1.104 & 1.029 & 1.004 & 0 & 1.116 & 1.016 \\ 0.939 & 1.059 & 1.010 & 1.047 & 0 & 1.109 \\ 1.167 & 0.866 & 1.030 & 1.047 & 1.202 & 0 \end{pmatrix}, \quad (3-19)$$

waar  $p_{ij}$  staat voor de *RMSE*-waarde in het model zonder modellering van de windrichting met windmast  $i$  als afhankelijke windmast en windmast  $j$  als verklarende windmast.

Een logische factor die naar verwachting een deel van de variatie in *RMSE* zal verklaren is de afstand tussen twee windmasten. Immers, hoe groter de afstand tussen twee windmasten is, des te meer wijkt het windpatroon van de ene windmast af ten opzichte van het windpatroon van de andere windmast.

Een overzicht hiervan is gegeven in onderstaande afstandmatrix  $\mathbf{D}$ :

$$\mathbf{D} = \begin{pmatrix} 0 & 3920 & 4704 & 4329 & 2180 & 4513 \\ 3920 & 0 & 2696 & 4292 & 3802 & 617 \\ 4704 & 2696 & 0 & 2179 & 3197 & 2597 \\ 4329 & 4292 & 2179 & 0 & 2193 & 4435 \\ 2180 & 3802 & 3197 & 2193 & 0 & 4226 \\ 4513 & 617 & 2597 & 4435 & 4226 & 0 \end{pmatrix}, \quad (3-20)$$

waar  $d_{ij}$  staat voor de afstand in meters tussen windmast  $i$  en windmast  $j$ .

Het volgende model kan nu geanalyseerd worden:

$$p_{ij} = \alpha + \beta \cdot d_{ij} + \varepsilon. \quad (3-21)$$

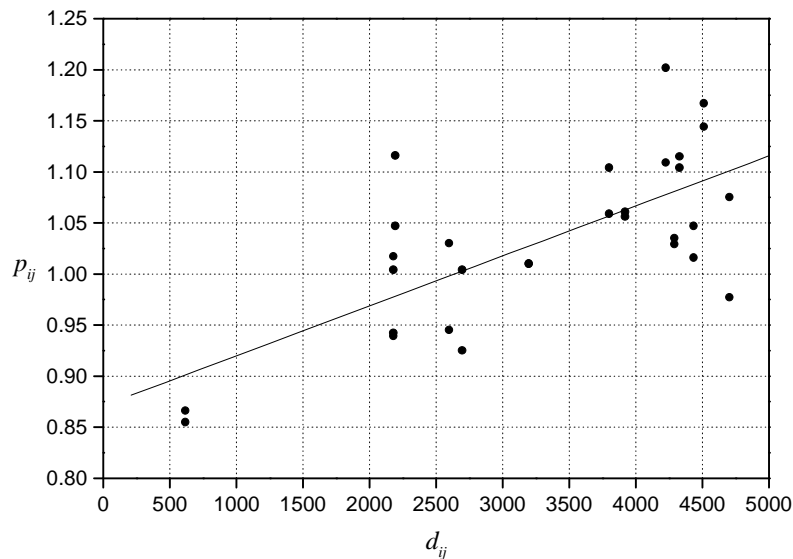
Dit model wordt met OLS geschat waarbij geen correcties aangebracht hoeven te worden in verband met de eventuele aanwezigheid van heteroskedasticiteit en autocorrelatie. De nulhypothese van geen heteroskedasticiteit wordt immers niet verworpen en aangezien er geen sprake is van een tijdreeks kan er ook van autocorrelatie geen sprake zijn. Wel zijn de *RMSE*-waarden niet geheel onafhankelijk van elkaar, omdat metingen bij één bepaalde windmast invloed hebben op meerdere *RMSE*-waarden. Hierdoor zijn de standaardfouten van de coëfficiënten wellicht niet zuiver en worden ze te laag geschat.

De numerieke resultaten van de modelschatting staan vermeld in onderstaande tabel:

parameter	coëfficiënt	standaardfout coëfficiënt	$t$ -waarde coëfficiënt
$\alpha$	0.871	$3.30 \cdot 10^{-2}$	26.37
$\beta$	$4.89 \cdot 10^{-5}$	$9.38 \cdot 10^{-6}$	5.22

**Tabel 3-25:** Regressieresultaten van model (3-21).

Een grafische weergave van de modelschatting is te zien in onderstaand figuur.



**Figuur 3-10:** Regressielijn voortkomend uit model (3-21).

Op basis van de numerieke en grafische resultaten kan gezegd worden dat de afstand een redelijke bijdrage levert aan de hoogte van de *RMSE*-waarde. De  $R^2$  is gelijk aan 0.49, wat betekent dat 49% van de variatie in de *RMSE* verklaard kan worden door de afstand tussen twee windmasten. Dit is een redelijk percentage, waarbij wel dient opgemerkt te worden dat dit voornamelijk veroorzaakt wordt door de twee gevallen met relatief kleine afstand (617 meter tussen windmast 2 en 6). Zonder deze twee punten is de waarde van  $R^2$  slechts 0.29.

Een groot deel van de variatie in de *RMSE* blijft dus vooralsnog onbeschreven. In deze paragraaf is echter gebleken dat modellering van de windrichting de *RMSE*-waarden een stuk doet verkleinen, vooral wanneer de beschuttingsverschillen tussen twee windmasten relatief groot zijn. Waarschijnlijk is het daarom zo dat wanneer de modellering van windrichting wordt meegenomen een groter deel van de *RMSE* verklaard kan worden door de afstand tussen twee windmasten.

Om deze reden zijn alle *RMSE*-waarden berekend uit de enkelvoudige lineaire regressiemodellen uit Deelparagraaf 3.4.3 (model IV2 t/m IV6). Omdat in deze modellen slechts de windsnelheid bij windmast 1 als afhankelijke variabele wordt gebruikt zijn opnieuw de *RMSE*-waarde berekend voor de modellen waarbij ook de windmasten 2 t/m 6 als afhankelijke variabelen zijn gebruikt. Een overzicht hiervan is gegeven in onderstaande matrix:

$$Q = \begin{pmatrix} 0 & 1.009 & 1.041 & 0.998 & 0.919 & 1.054 \\ 0.972 & 0 & 0.952 & 0.961 & 0.967 & 0.825 \\ 0.951 & 0.907 & 0 & 0.877 & 0.926 & 0.904 \\ 0.962 & 0.958 & 0.899 & 0 & 0.939 & 0.966 \\ 0.872 & 0.961 & 0.963 & 0.931 & 0 & 1.000 \\ 1.017 & 0.814 & 0.923 & 0.977 & 0.999 & 0 \end{pmatrix}, \quad (3-22)$$

waar  $q_{ij}$  staat voor de *RMSE*-waarde in het model met modellering van de windrichting met windmast  $i$  als afhankelijke windmast en windmast  $j$  als verklarende windmast.

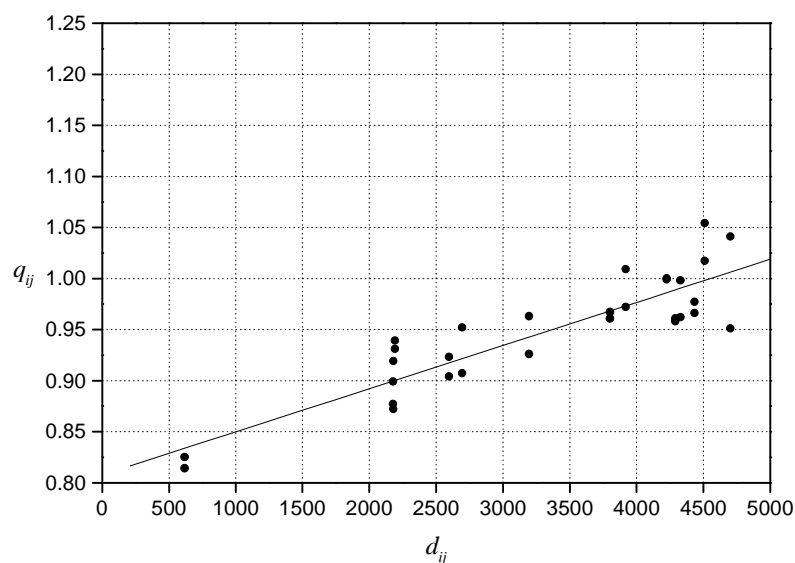
Het te analyseren model wordt daarmee als volgt:

$$q_{ij} = \alpha + \beta \cdot d_{ij} + \varepsilon. \quad (3-23)$$

Dit model wordt opnieuw met OLS geschat. De numerieke en grafische resultaten hiervan zijn als volgt:

parameter	coëfficiënt	standaardfout coëfficiënt	<i>t</i> -waarde coëfficiënt
$\alpha$	0.808	$1.48 \cdot 10^{-2}$	54.67
$\beta$	$4.22 \cdot 10^{-5}$	$4.20 \cdot 10^{-6}$	10.07

**Tabel 3-26:** Regressieresultaten van model (3-23).



**Figuur 3-11:** Regressielijn voortkomend uit model (3-23).

Visueel valt al af te leiden uit Figuur 3-11 dat de *RMSE* nu beter verklaard wordt door de afstand aangezien de punten minder afwijken van de regressielijn in vergelijking met Figuur 3-10. Meer formeel kan dit geconcludeerd worden op basis van de hogere *t*-waarden in Tabel 3-26.

De waarde van  $R^2$  is nu gelijk aan 0.78, wat een forse verhoging betekent ten opzichte van de  $R^2$  in model (3-21) van 0.49. Ook na weglating van de punten met relatief kleine afstand tussen windmast 2 en 6 is de  $R^2$  behoorlijk wat hoger: 0.66 ten opzichte van 0.29.

Duidelijk is nu dat zoals verwacht dat de gemodelleerde beschuttingsverschillen tussen windmasten een belangrijke bijdrage leveren aan de mate waarin de *RMSE* kan worden verklaard.

De vraag die nu resteert is of de overgebleven ruis in model (3-23) toch nog voor een deel wordt veroorzaakt door de bestaande beschuttingsverschillen. Als dit het geval is dan houdt dat in dat de bestaande beschuttingsverschillen onvoldoende gemodelleerd zijn of dat er door bestaande beschuttingsverschillen geen optimale correctie mogelijk is. Dit kan onderzocht worden door aan model (3-23) een verklarende variabele toe te voegen waarbij de waarden van deze variabele staan voor de percentuele verbeteringen in *RMSE* die zijn behaald door de modellen uit te breiden met de modellering van de

windrichting. Als deze variabele een significant positief coëfficiënt heeft dan duidt dit er op dat uitspraken over de windsnelheid bij een bepaalde mast aan de hand van metingen gedaan bij een andere mast minder betrouwbaar zijn bij grote beschuttingsverschillen tussen beide windmasten. Wanneer de variabele een significant negatief coëfficiënt heeft dan duidt dit er juist op dat de uitspraken bij grote beschuttingsverschillen betrouwbaarder zijn dan bij kleine beschuttingsverschillen. Dit laatste ligt echter niet in de lijn der verwachtingen.

De percentuele verbeteringen volgen direct uit de verschillen tussen  $p_{ij}$  en  $q_{ij}$ . Een overzicht hiervan staat in onderstaande matrix:

$$\mathbf{V} = \begin{pmatrix} 0 & 4.9 & 3.2 & 10.5 & 9.6 & 7.9 \\ 8.0 & 0 & 5.2 & 7.1 & 12.4 & 3.5 \\ 2.7 & 1.9 & 0 & 6.9 & 8.3 & 4.3 \\ 12.9 & 6.9 & 10.5 & 0 & 15.9 & 4.9 \\ 7.1 & 9.3 & 4.7 & 11.1 & 0 & 9.8 \\ 12.9 & 6.0 & 10.4 & 6.7 & 16.9 & 0 \end{pmatrix}, \quad (3-24)$$

waar  $v_{ij}$  staat voor de procentuele verbetering in *RMSE* van het model zonder windrichting naar het model met windrichting met windmast  $i$  als afhankelijke windmast en windmast  $j$  als verklarende windmast.

Het uiteindelijk te analyseren model is nu als volgt:

$$q_{ij} = \alpha + \beta \cdot d_{ij} + \gamma \cdot v_{ij} + \varepsilon. \quad (3-25)$$

Dit model kan opnieuw met OLS geschat worden en de numerieke resultaten hiervan staan in onderstaande tabel:

parameter	coëfficiënt	standaardfout coëfficiënt	<i>t</i> -waarde coëfficiënt
$\alpha$	0.793	$1.70 \cdot 10^{-2}$	46.75
$\beta$	$4.16 \cdot 10^{-5}$	$4.11 \cdot 10^{-6}$	10.13
$\gamma$	$2.03 \cdot 10^{-3}$	$1.28 \cdot 10^{-3}$	1.58

**Tabel 3-27:** Regressieresultaten van model (3-25).

Nagegaan is in hoeverre de verklarende variabelen in model (3-25) met elkaar correleren. Het verband hiertussen blijkt niet significant te zijn waardoor er geen sprake is van multicollineariteit. Hierdoor kunnen de coëfficiënten van de variabelen in Tabel 3-27 afzonderlijk geïnterpreteerd worden.

Aan de coëfficiënt van  $\gamma$  valt af te lezen dat de invloed van de verbetering niet significant is, zeker met het oog op de eventuele onderschatting van de standaardfouten. Dit is ook af te lezen uit de minieme toename in  $R^2$  van 0.78 naar 0.80.

Er kan nu uit bovenstaande resultaten geconcludeerd worden dat de modellering van de windrichting in zijn geheel corrigeert voor bestaande beschuttingsverschillen, zodat uitspraken over de windsnelheid bij een bepaalde mast op basis van metingen gedaan bij een andere windmast in het geval van grote beschuttingsverschillen tussen de windmasten even betrouwbaar zijn als in het geval van kleine beschuttingsverschillen.

Hieruit volgt dat men bij het plaatsen van windmasten geen rekening hoeft te houden met de beschuttingsverschillen tussen de eventuele toekomstige windmastlocaties. Hiervoor kan immers op dusdanige manier voor gecorrigeerd worden dat het feitelijk niet uitmaakt hoe groot de beschuttingsverschillen zijn.

Wel blijft natuurlijk overeind dat een locatie dient uitgezocht te worden waarbij de beschutting minimaal is. Sterke beschutting zorgt immers voor turbulentie en daardoor een grilliger windpatroon, wat uitspraken over de windsnelheid bemoeilijkt. Hierbij moet echter opgelet worden dat de beschutting niet te zeer afwijkt van de beschutting zoals die aanwezig is bij de touch-down zone van de landingsbaan. Uitspraken dienen namelijk hierop betrekking te hebben. Een locatie met minimale beschutting is derhalve niet altijd optimaal wanneer bij de touch-down zone van de landingsbaan de beschutting bij bepaalde windrichtingen wel sterk is.

Een advies over de locatie van plaatsing/herplaatsing van windmasten is daardoor afhankelijk van de volgende factoren:

- Beschuttingsverschillen tussen windmast en touch-down zone (zo klein mogelijk),
- Afstand tussen windmast en touch-down zone (zo klein mogelijk),
- Afstanden tussen windmasten onderling (zo klein mogelijk),
- Beschutting bij windmasten (minimaal).

Er hoeft dus niet meer gekeken worden naar de beschuttingsverschillen tussen de windmasten onderling.

### **3.5 Het opstellen van betrouwbaarheidsintervallen**

In Deelparagraaf 3.2.1 werd aangetoond hoe efficiënte schatters verkregen konden worden door FGLS toe te passen. Immers, door toedoen van de aanwezige heteroskedasticiteit waren de coëfficiëntschattingen niet meest efficiënt wanneer gebruik gemaakt werd van OLS. Er werd een wegingsfactor  $w_t$  bepaald waarmee alle variabelen vermenigvuldigd dienden te worden alvorens OLS toe te passen. Ook werd aangegeven dat deze methode gebruikt kon worden voor het opstellen van betrouwbaarheidsintervallen voor uitspraken over de windsnelheid. In deze paragraaf zal hier verder op worden ingegaan.

In Deelparagraaf 2.2.3 werd aangetoond hoe een betrouwbaarheidsinterval voor een bepaalde voorspelling kan worden berekend middels (2-17). Deze methode kan bij het opstellen van een betrouwbaarheidsinterval voor de gemiddelde 2-minuut windsnelheid bij windmast 1 niet zondermeer toegepast worden.

In (2-17) wordt immers de OLS schatter van de covariantiematrix gebruikt. Omdat er sprake is van autocorrelatie is dit geen zuivere schatter van de werkelijke covariantiematrix. Om deze reden dient de OLS schatter van de covariantiematrix vervangen te worden door de Newey-West schatter van de covariantiematrix zoals deze ook gebruikt werd bij het opstellen van de regressiemodellen voor de windsnelheid.

Verder wordt in (2-17) gebruikt gemaakt van een constante  $s^2$ . Aangezien gebleken is dat door de aanwezigheid van heteroskedasticiteit  $s^2$  niet constant is maar afhankelijk is van de hoogte van  $FsF$  wordt  $s^2$  daarom vervangen door een functie die opgesteld kan worden zoals gedaan is in Deelparagraaf 3.2.1. Gebeurt dit voor het optimale model F2 dan resulteert dit in

$$s_h^2 = 0.215 + 0.463FsF_h^2. \quad (3-26)$$

Wanneer deze wijzigingen worden doorgevoerd ziet het 95%-betrouwbaarheidsinterval voor een bepaalde voorspelling van de gemiddelde 2-minuut windsnelheid bij windmast 1 er als volgt uit:

$$P\left(\hat{F}2_{1,h} - 1.96 \cdot s(\hat{F}2_{1,h}) \leq F2_{1,h} \leq \hat{F}2_{1,h} + 1.96 \cdot s(\hat{F}2_{1,h})\right) = 0.95, \quad (3-27)$$

waarbij

$$s(\hat{F}2_{1,h}) = \sqrt{s_h^2 + \mathbf{x}_h' \text{cov}(\mathbf{b}) \mathbf{x}_h}. \quad (3-28)$$

In (3-28) is  $\text{cov}(\mathbf{b})$  gelijk aan (2-13).

De tweede term onder de wortel in (3-28) wordt snel kleiner wanneer er gebruik gemaakt wordt van meer data (dus wanneer  $T$  groter wordt). In dit onderzoek (en ook bij implementatie van de methode) is  $T$  behoorlijk groot (21800) waardoor de tweede term relatief klein is ten opzichte van de eerste term. Deze eerste term wordt namelijk niet kleiner bij toenemende  $T$ . Hierdoor wordt de lengte van het betrouwbaarheidsinterval voor een groot deel bepaald door het al dan niet grillige patroon in de windsnelheid ( $F_sF$ ). De waarde van  $F_sF$  bepaalt immers de hoogte van de eerste term. Slechts een zeer klein deel van de lengte van het interval wordt bepaald door de windrichting (D2). Deze waarde zit immers in de tweede term. Ook de windsnelheid ( $F2$ ), die eveneens deel uitmaakt van de tweede term, heeft een zeer geringe toegevoegde waarde bij het bepalen van de lengte van het interval. Aangezien echter  $F2$  en  $F_sF$  behoorlijk gecorreleerd zijn met elkaar, is de lengte van het betrouwbaarheidsinterval doorgaans bij hoge windsnelheden groter dan bij lage windsnelheden.

Om een indruk te krijgen van de lengte van het betrouwbaarheidsinterval staat in Tabel 3-28 voor verschillende waarden voor  $F_sF$  de bijbehorende waarden van de lengte van het 95%-betrouwbaarheidsinterval vermeld.

$F_sF_h$	cumulatief percentage $F_sF_h$	lengte van het 95%-betrouwbaarheidsinterval (m/s)
0.0	0.0	1.83
0.2	4.9	1.91
0.4	22.2	2.12
0.6	42.1	2.43
0.8	61.6	2.81
1.0	77.6	3.23
1.2	88.4	3.69
1.4	94.4	4.16
1.6	97.1	4.64
1.8	98.7	5.14
2.0	99.5	5.64

**Tabel 3-28:** Lengtes van het 95%-betrouwbaarheidsinterval bij verschillende waarden voor  $F_sF_h$ .

In Tabel 3-28 is ook een cumulatief percentage van  $F_sF$  aangeven, waaruit opgemaakt kan worden hoe vaak een bepaalde waarde van  $F_sF$  in de praktijk wordt overschreden. De verschillende waarden voor de windsnelheid en windrichting zijn bij het berekenen van de lengtes in de tabel constant gehouden. Uit het voorafgaande volgt immers dat bij veranderingen in deze waarden de lengtes van de intervallen slechts marginaal aangepast hoeft te worden.

Uit de tabel valt verder af te leiden dat in ongeveer de helft van de gevallen het interval een lengte heeft van minder dan circa 2.6 meter per seconde. In extreme situaties is de lengte van het interval meer dan 5.6 meter per seconde. Dit is echter slechts in ongeveer 0.5% van de totale tijd het geval. Dit zijn echter wel de situaties die van belang zijn, aangezien er juist dan mogelijk sprake is van gevaarlijk weer met zware windstoten.

## 4 Conclusies

In dit onderzoek is gebleken dat door gebruik te maken van een statistisch model een behoorlijke verbetering mogelijk is in de kwaliteit van een uitspraak over de gemiddelde 2-minuut windsnelheid bij een niet-functionerende windmeetmast op een luchthaven ten opzichte van de huidige operationele back-up schema's.

Dit statistisch model is een meervoudig lineair regressiemodel waarin gebruik gemaakt wordt van meetgegevens van alle overige wel-functionerende windmeetmasten op een luchthaven om een uitspraak te genereren over een bepaalde niet-functionerende windmeetmast. Tevens is dit model voorzien van modellering van de verschillende terreinruwheden nabij de windmeetmasten middels Fourier analyse.

Het opnemen van meerdere windmeetmasten zorgt ervoor dat het grillige patroon wat de gemiddelde 2-minuut windsnelheid kenmerkt voor een groot deel wordt gladgestreken. Het modelleren van de verschillende terreinruwheden zorgt ervoor dat er geen structurele afwijkingen in de uitspraak ontstaan bij bepaalde windrichtingen. Immers, wanneer de terreinruwheden niet worden gemodelleerd zal een uitspraak over de windsnelheid bij een windmeetmast met relatief veel bebouwing aan de noordzijde vaak te hoog uitvallen in het geval de wind waait uit noordelijke richtingen.

De huidige operationele back-up schema's zijn in feite ook regressiemodellen. Hierbij is echter slechts één windmast opgenomen en zijn de parameters gerestricteerd. Verder houdt een back-up schema geen rekening met verschillende terreinruwheden. Het grillige patroon wordt dus niet gladgestreken en voor structurele afwijkingen veroorzaakt door verschillende terreinruwheden wordt niet gecorrigeerd.

Verder is het statistisch model voorzien van foutenmarges omtrent de uitspraak. Hiermee wordt aangegeven hoe betrouwbaar de gegenereerde uitspraak over de gemiddelde 2-minuut windsnelheid is. Deze betrouwbaarheid blijkt sterk afhankelijk te zijn van de weersituatie. Wanneer de windsnelheid een grillig patroon volgt is het moeilijker om een betrouwbare uitspraak te genereren dan wanneer het patroon vrij stabiel is. Het model houdt hier rekening mee, zodat het ook gehanteerd kan worden bij extreme weersituaties, die juist belangrijk zijn met het oog op de veiligheid.

Het aangeven van foutenmarges is tevens een belangrijk voordeel ten opzichte van de back-up schema's. Deze zijn immers niet voorzien van foutenmarges, zodat hierbij niet aangegeven kan worden in welke mate een uitspraak over de gemiddelde 2-minuut windsnelheid bij een niet-functionerende windmeetmast representatief is voor de werkelijke gemiddelde 2-minuut windsnelheid bij deze windmeetmast.

Het statistisch model corrigeert in zijn geheel voor de aanwezige verschillen in terreinruwheden. Een bijkomend voordeel hiervan is dat bij het plaatsen/herplaatsen van windmeetmasten men niet meer hoeft te kijken naar verschillen in terreinruwheden tussen eventuele locaties van de windmeetmasten. Dit gebeurt immers wel bij het hanteren van de huidige back-up schema's.



## 5 Suggesties

In dit onderzoek is gekeken naar modellen die een uitspraak doen over de windsnelheid bij een bepaalde windmeetmast op tijdstip  $t$  aan de hand van metingen van andere windmeetmasten eveneens op tijdstip  $t$ . Omdat windvelden zich verplaatsen over een bepaald traject kan het zijn dat een bepaalde windsnelheidsverandering die zich voordoet bij een bepaalde windmeetmast zich enige tijd later pas voordoet bij een andere windmeetmast. Daarom kan de voorspelkracht van het model wellicht toenemen wanneer in bepaalde situaties windmetingen op eerdere momenten dan tijdstip  $t$  worden meegenomen in de modelbouw. Gezien de mogelijkheid van verbetering van het model is het daarom raadzaam hiernaar te kijken in een eventueel vervolgonderzoek

Verder is in dit onderzoek de aandacht gelegd op uitspraken over gemiddelde 2-minuut windsnelheden. Omdat verkeersleidingen op luchthavens ook op basis van de maximale windstoot van de laatste 10 minuten besluiten over het wel dan niet laten landen of opstijgen van een vliegtuig is het opnieuw raadzaam om in eventueel vervolgonderzoek hier meer aandacht aan te geven. Kan hierbij de maximale windstoot behandeld worden zoals de gemiddelde 2-minuut windsnelheid en kan zodoende hiervoor een soortgelijke procedure worden opgesteld of is het noodzakelijk de maximale windstoot op een geheel aparte wijze te benaderen?



## 6 Aanbevelingen

De in dit onderzoek beschreven procedure houdt rekening met verschillen in terreinruwheden. Omdat deze terreinruwheden niet constant blijven in de tijd door aanbouw, verwijdering van bebouwing of door veranderingen in de begroeiing is het daarom verstandig om de modelparameters regelmatig bij te stellen om te voorkomen dat het model rekening houdt met bepaalde beschuttingsverschillen die op het moment van toepassing niet meer aanwezig zijn.

In dit onderzoek is de beschikbare data opgedeeld in twee delen: een deel voor het bouwen van de modellen, het andere deel om de verschillende modellen met elkaar te kunnen vergelijken.

Voor operationeel gebruik is het echter niet nodig om de data in tweeën te splitsen, aangezien het hierbij niet nodig is om verschillende modellen met elkaar te vergelijken. Het optimale model is immers in dit onderzoek bepaald, zodat het slechts gaat om het schatten van de parameters in dit optimale model. Bij het bepalen van de parameters kan dus gebruik gemaakt worden van de totale data wat ten goede komt aan de betrouwbaarheid van de geschatte parameters.

Verkeersleidingen baseren beslissingen over het al dan niet verantwoord landen of opstijgen van een vliegtuig mede op de hoogte van de dwarswindsnelheid (het deel van de windsnelheid dat loodrecht staat op de vliegrichting). Naast de hoogte van de windsnelheid is kennis over de windrichting noodzakelijk om de hoogte van de dwarswindsnelheid te bepalen.

De in dit onderzoek beschreven procedure doet enkel een uitspraak over de hoogte van de windsnelheid bij een niet-functionerende windmast en niet over de windrichting. Om toch ook over de windrichting te beschikken in het geval een bepaalde windmast niet functioneert zou men de huidige operationele back-up schema's kunnen gebruiken, waarbij de windrichting van een andere windmast op de luchthaven gebruikt wordt als vervanging van de windrichting bij de niet-functionerende windmast, zoals dat ook gebeurt met de windsnelheid.

Aangezien naast de windsnelheid ook de windrichting vaak een grillig patroon volgt zijn de back-up schema's voor dit doel niet optimaal. In plaats van gebruik te maken van de waarde gemeten bij één andere windmast, is het daarom raadzaam gebruik te maken van het vectoriele gemiddelde van de windrichtingen gemeten bij alle functionerende windmasten die op een luchthaven aanwezig zijn zodat het grillige patroon grotendeels wordt gladgestreken.

In dit onderzoek is aangegeven dat bepaalde luchtwervelingen, veroorzaakt door passerende vliegtuigen, in bepaalde situaties een te groot effect hebben op de windmetingen. Dit geldt met name voor de maximale windstoot van de laatste 10 minuten en in mindere mate voor de gemiddelde 2-minuut windsnelheid. De data die in dit onderzoek gebruikt is, is hiervoor gecorrigeerd met behulp van het in Paragraaf 1.4 genoemde algoritme. Omdat dit echter een 'real time' algoritme is, kan niet altijd optimaal gecorrigeerd worden voor kunstmatige windstoten. Om deze reden is het raadzaam om voor het schatten van de modelparameters gebruik te maken van de oorspronkelijke data (dus zonder toepassing van het algoritme) die achteraf gecorrigeerd

zijn voor de aanwezige kunstmatige windstoten, aangezien het eenvoudiger is om deze achteraf te onderscheiden dan op het moment van meten.

Wanneer een bepaalde windmast niet functioneert kan gebruik gemaakt worden van de in dit onderzoek beschreven procedure waarbij alle overige windmasten op een luchthaven worden meegenomen in de modelbouw. Wanneer er echter, naast de windmast waarover een uitspraak gewenst is, meerdere windmasten niet functioneren kan niet van het optimale model gebruik gemaakt worden. Er dient daarom een aantal modellen achter de hand gehouden worden waarbij gebruik gemaakt wordt van alle windmasten die wel functioneren.

Bij Amsterdam Airport Schiphol bijvoorbeeld betekent dit vijf modellen voor de situaties waarin één extra windmast niet functioneert en tien modellen voor de situaties waarin twee windmasten niet functioneren naast de windmast waarover een uitspraak gewenst is. Deze tien modellen zullen overigens slechts sporadisch ingezet hoeven te worden, aangezien het hoogst zelden voorkomt dat er drie windmasten tegelijkertijd niet functioneren.

# Dankwoord

Hierbij wil ik de mensen bedanken die mij geholpen hebben in het vervaardigen van deze scriptie. Op de eerste plaats veel dank aan Jelke Bethlehem, mijn begeleider van de Universiteit van Amsterdam, die meerdere voorlopige versies en delen van versies van mijn scriptie heeft doorgeworsteld en zo mij veel bruikbare aanwijzingen heeft kunnen geven in het vervolmaken van de uiteindelijke scriptie.

Verder dank aan Jitze van der Meulen, mijn begeleider op Het Koninklijk Nederlands Meteorologisch Instituut, die altijd voor me klaar stond wanneer dat nodig was.

Speciale dank aan Kees Kok die een voorlopige versie van mijn scriptie met veel enthousiasme heeft gelezen en aan de basis stond van een belangrijke analysemethode welke deel uit maakt van mijn scriptie.

Tenslotte dank aan Henk Benschop, Albert Jacobs en Ben Wichers Schreur bij wie ik altijd terecht kon met vragen en die nuttige adviezen konden geven na het lezen van een voorlopige versie van mijn scriptie.



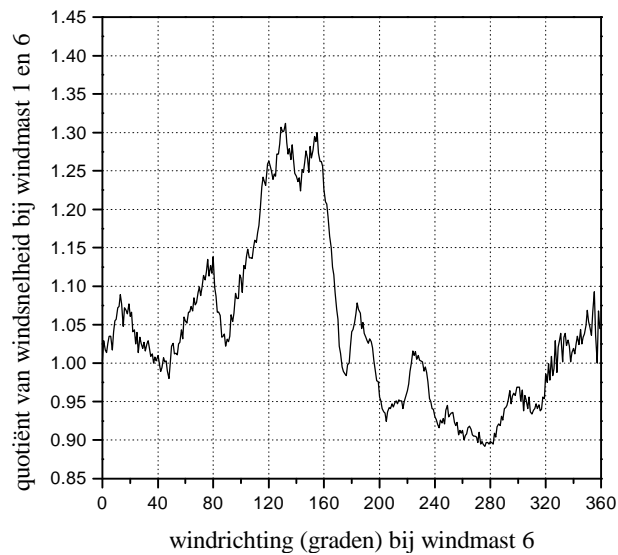
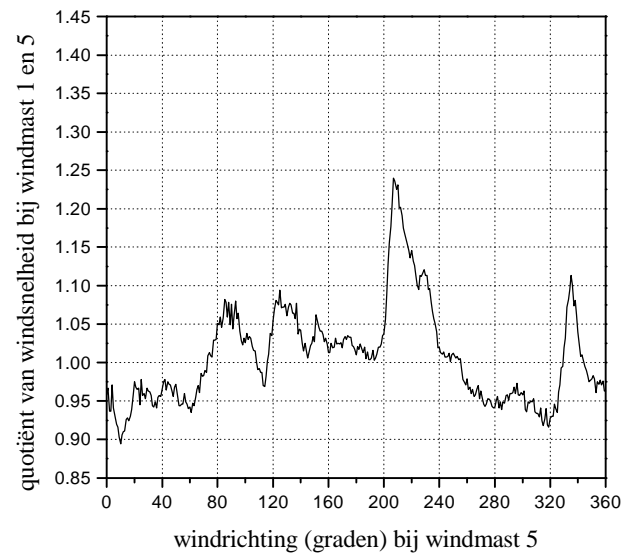
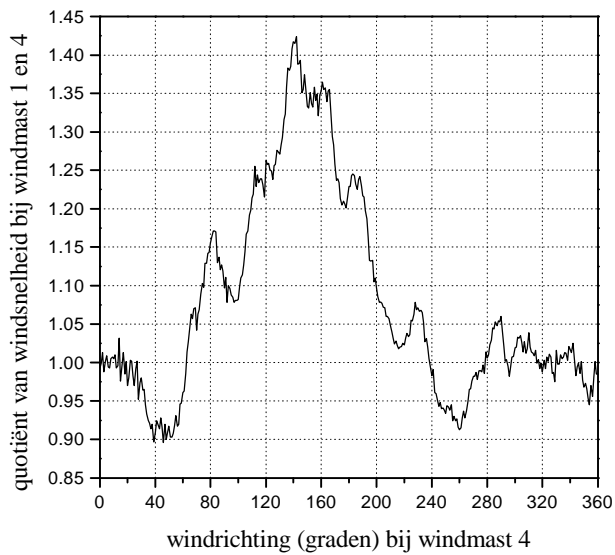
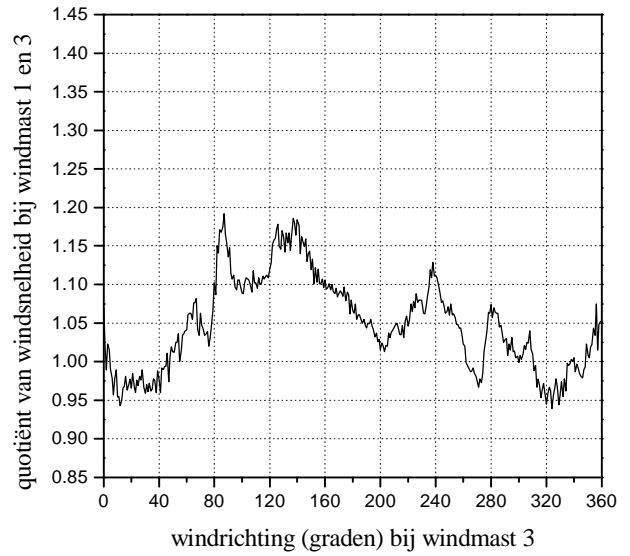
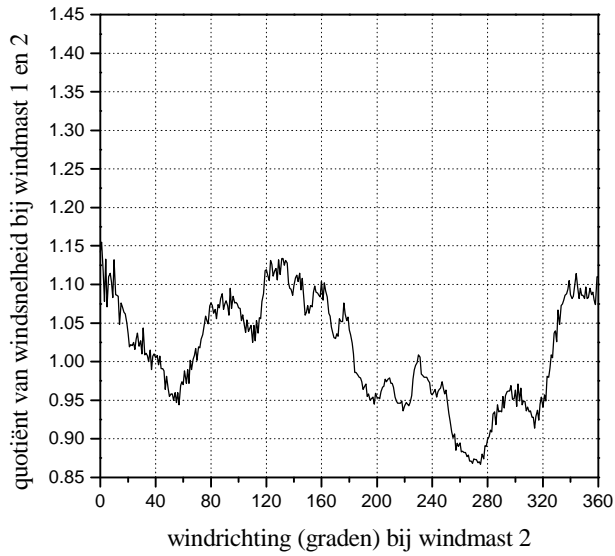
# Referenties

- Draper, N., and H. Smith. *Applied Regression Analysis*. New York: John Wiley and Sons, 1980.
- Greene, W. *Econometric Analysis*. Upper Saddle River, New Jersey: Prentice-Hall International, 1997.
- Meulen, J. van der. "Wake Vortex Induced Wind Measurement at Airfields: A Simple Algorithm to Reduce the Vortex Impact." *Instruments and Observing Methods Report No. 70 (WMO/TD – No. 877)*, 1998, pp. 139-142, uitgave WMO, Genève.
- Newey, W., en K. West. "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55, 1987a, pp. 703-708.
- White, H. "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, 48, 1980b, pp. 817-838.

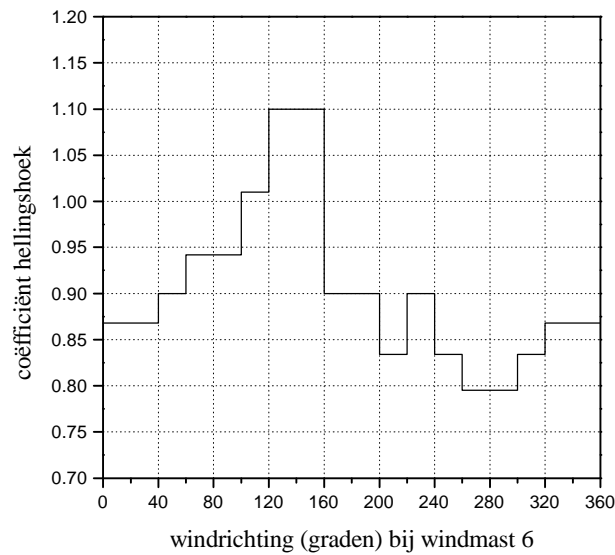
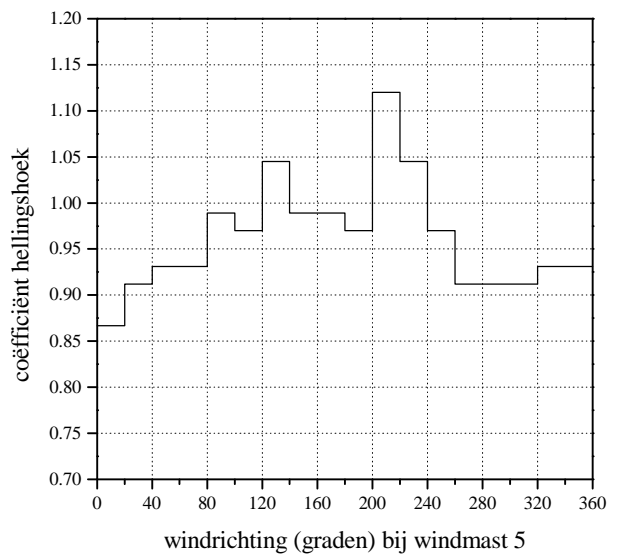
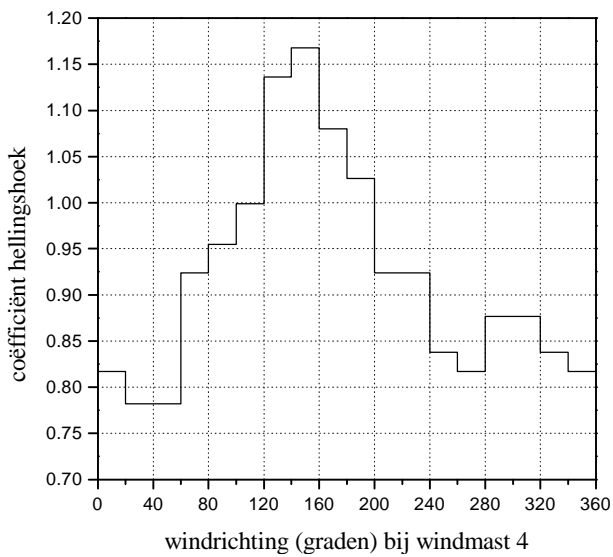
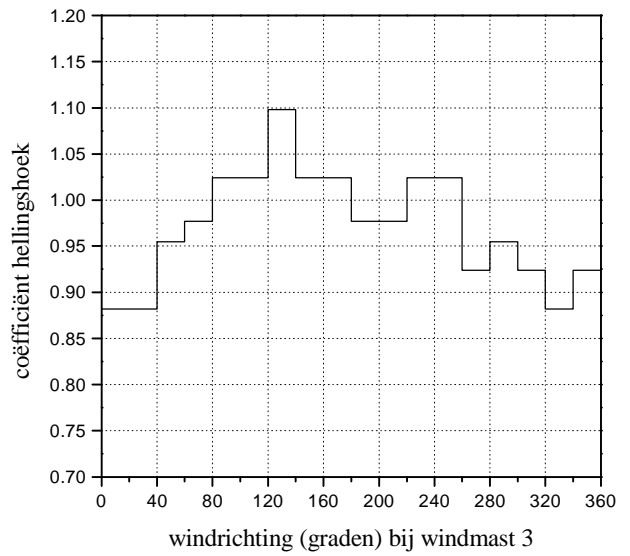
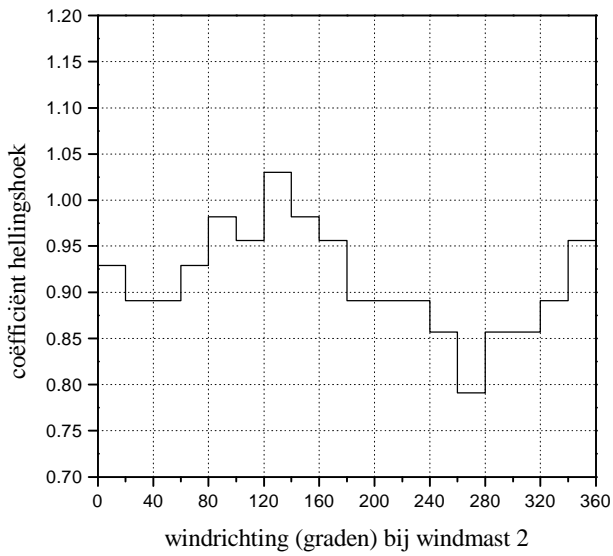


# Bijlagen

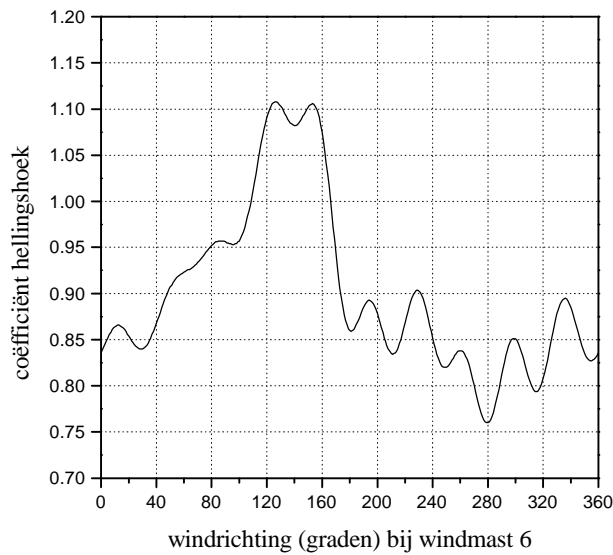
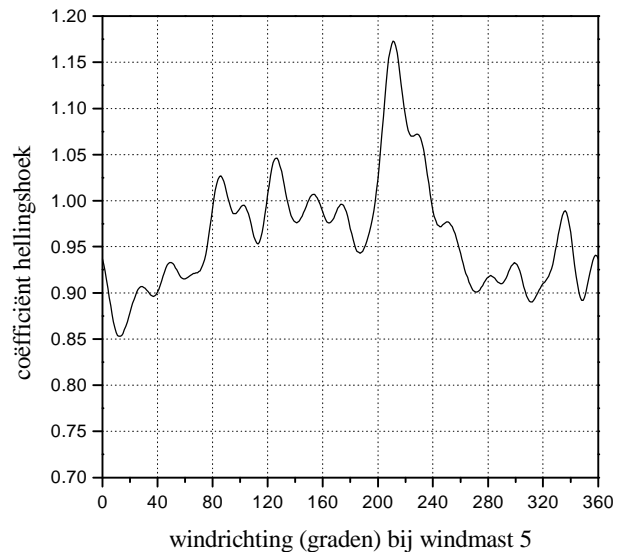
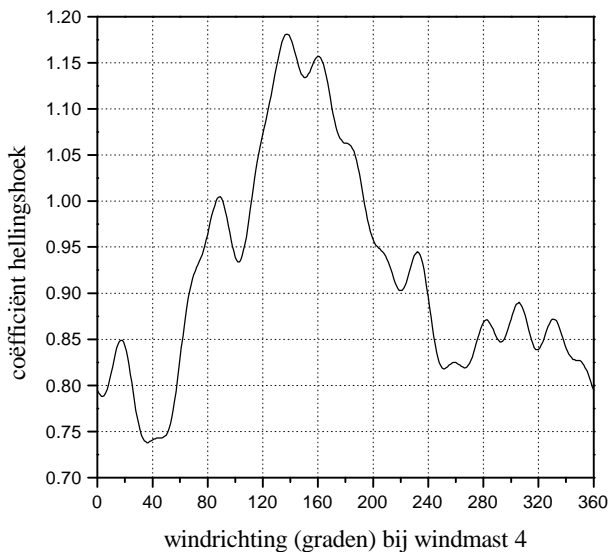
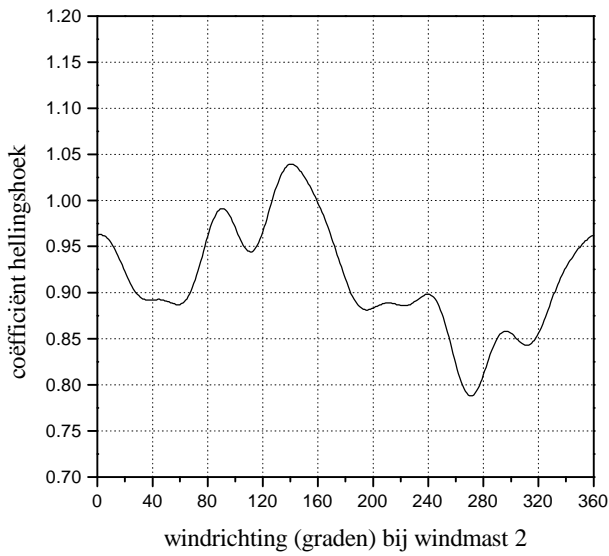
## Bijlage 1: Quotiëntplots van windsnelheden



## Bijlage 2: Coëfficiëntplots van hellingshoek bij clusteren van windrichting



### Bijlage 3: Coëfficiëntplots van hellingshoek bij Fourier modellering



## Bijlage 4: Coëfficiëntplots van hellingshoek (combinatie bijlage 2 en 3)

