

Modelit
Rotterdamse Rijkweg 126
3042 AS Rotterdam
Telefoon +31 10 4623621



info@modelit.nl
www.modelit.nl

in opdracht van RIKZ

Validatie golfparameters Amelanders zeegebied voor stormseizoen 2005/2006 met Wavix IV

Datum 04 oktober 2006

Modelit
KvK Rotterdam 24290229

1 Inleiding

Tijdens het stormseizoen 2005/2006 hebben meetboeien op een vijftal locaties in het Ameland zeeget elke 10 minuten golfgegevens verzameld. Vanwege de gevoeligheid van de meetboeien zijn veel van deze metingen echter onbetrouwbaar of ontbreken zelfs helemaal. Hierdoor is het noodzakelijk om de meetgegevens aan te vullen en te controleren op juistheid.

Voor het aanvullen en controleren van meetgegevens is in 2005 in opdracht van RIKZ de applicatie WavixIV ontwikkeld. Met behulp van deze applicatie is het mogelijk om met behulp van een model van neurale netwerken meetgegevens te controleren en waar nodig aan te vullen.

In dit document worden de stappen beschreven die doorlopen zijn om de golfgegevens in het Ameland zeeget tijdens het stormseizoen 2005/2006 te valideren met WavixIV. De belangrijkste resultaten zijn tevens beschikbaar op deze Cd-rom en zijn:

- Een regressiemodel en een model van neurale netwerken voor het bijstellen van golfgegevens met behulp van WavixIV. Deze modellen zijn te vinden op de meegeleverde Cd-rom in de directories 'Regressiemodel' en 'Netwerken'.
- Een set van gevalideerde golfparameters op alle locaties in het Ameland zeeget in .dia formaat. De gevalideerde data zijn te vinden op de meegeleverde Cd-rom in de directory 'Validatie/Gevalideerd'.
- Een beschrijving van de kwaliteit van de door het model bijgeschatte golfparameters met speciale aandacht voor 'zwaar' weer.

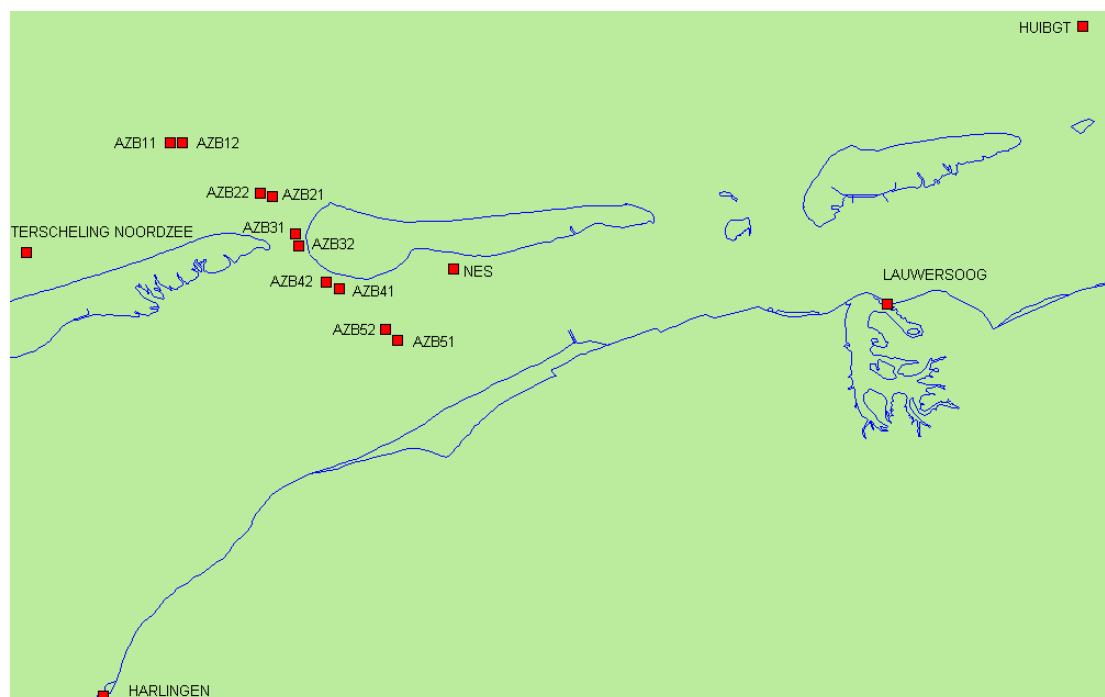
Op de meegeleverde Cd-rom zijn bovendien de tussenresultaten voor elke stap beschikbaar. Op deze manier kan gemakkelijk het gehele validatieproces worden doorlopen en gecontroleerd worden op juistheid. Bovendien kan op deze manier dit project dienen als leidraad voor toekomstige projecten aangezien het de eerste keer is dat WavixIV gebruikt wordt voor de validatie van golfgegevens.

2 De beschikbare data

Ten behoeve van de validatie van de golfparameters in het Amelander zeegat voor het stormseizoen 2005/2006 zijn twee .dia bestanden beschikbaar met de gemeten waterhoogte en de golfparameters. De benodigde informatie over de windrichting en windsnelheid op het meetstation Huibertgat is verkregen uit de Matroos database. Een overzicht van de inhoud van deze bestanden is beschreven in Tabel 1.

Tabel 1: De beschikbare tijdreeksen voor de validatie van golfparameters.

Bestand	Inhoud			Begin	Eind	Tijdstap
	Locatie	Variabele				
wa_amzgt.dia	TERSLNZE	WATHTE		01-10-2005	31-05-2006	10 min
	NES	WATHTE		01-10-2005	31-05-2006	10 min
	HARLGN	WATHTE		01-10-2005	31-05-2006	10 min
	LAUWOG	WATHTE		01-10-2005	31-05-2006	10 min
wind_direction.dat	HUIBGT	WINDRTG		01-10-2005	01-06-2006	10 min
wind_speed.dat	HUIBGT	WINDSHD		01-10-2005	01-06-2006	10 min
golf_p.dia	AMLDZGTB11	H1/3,HTE3,Hm0,TH1/3,Th0,Tm02		31-10-2005	16-05-2006	10 min
	AMLDZGTB12	H1/3,HTE3,Hm0,TH1/3,Th0,Tm02		02-11-2005	19-05-2006	10 min
	AMLDZGTB21	H1/3,HTE3,Hm0,TH1/3,Th0,Tm02		31-10-2005	19-05-2006	10 min
	AMLDZGTB22	H1/3,HTE3,Hm0,TH1/3,Th0,Tm02		31-10-2005	19-05-2006	10 min
	AMLDZGTB31	H1/3,HTE3,Hm0,TH1/3,Th0,Tm02		31-10-2005	19-05-2006	10 min
	AMLDZGTB32	H1/3,HTE3,Hm0,TH1/3,Tm02		31-10-2005	19-05-2006	10 min
	AMLDZGTB41	H1/3,HTE3,Hm0,TH1/3,Th0,Tm02		31-10-2005	19-05-2006	10 min
	AMLDZGTB42	H1/3,HTE3,Hm0,TH1/3,Tm02		31-10-2005	19-05-2006	10 min
	AMLDZGTB51	H1/3,HTE3,Hm0,TH1/3,Tm02		31-10-2005	19-05-2006	10 min
	AMLDZGTB52	H1/3,HTE3,Hm0,TH1/3,Tm02		31-10-2005	19-05-2006	10 min



Figuur 1: Locaties van de stations met meetgegevens.

Naast de golfparameters die opgesomd zijn in Tabel 1 zijn ook de volgende parameters aanwezig in het aangeleverde .dia bestand met de golfparameters:

- Fp3
de frequentie waar het energiespectrum maximaal is
- Hmax
de maximaal gemeten golfhoogte

- Sobh
de gemiddelde spreiding van de richting uit het richtingspectrum
- Th3
de gemiddelde richting uit het golfrichtingspectrum 30-100 mHz
- Tmax
de maximaal gemeten golfperiode
- Tmmin_10
de gemiddelde golfperiode uit de spectrale momenten $m_{-1} + m_0$ van 30-1000 mHz

Deze parameters kunnen echter (nog) niet gevalideerd worden met WavixIV. De validatie is dus beperkt tot de volgende golfparameters:

- H1/3
de gemiddelde golfhoogte uit het hoogste 1/3 deel van de golven
- HTE3
de laagfrequente golfhoogte uit het spectrum van 30-100mHz
- Hm0
de significante golfhoogte uit het energiespectrum van 30-500 mHz
- TH1/3
de gemiddelde golfperiode uit het hoogste 1/3 deel van de golven
- Th0
de gemiddelde richting uit het golfrichtingspectrum 30-500 mHz ten opzichte van het kaartnoorden.
- Tm02
de gemiddelde golfperiode uit de spectrale momenten m_0 en m_2 van 30-500 mHz

Waarbij bovendien gebruikt gemaakt wordt van:

- De windrichting en windsnelheid op het station Huibertgat
- De gemeten waterhoogte op de stations: Terschelling Noordzee en Nes.

Uit Tabel 1 is af te leiden dat in verband met de databeschikbaarheid de maximale mogelijke studieperiode loopt van 2 november 2005 tot en met 16 mei 2006.

N.B. De vier bronbestanden uit Tabel 1 zijn te vinden in directory 'Bronbestanden' van de meegeleverde CD-rom. In de header van de twee bestanden wind_speed.dat en wind_direction.dat is beschreven hoe de data is verkregen uit de Matroos database.

N.B. In het bestand Amelander_zeegat.dat in de directory 'Basisdata' op de meegeleverde CD-rom zijn alle reeksen opgenomen die als basis dienen voor het verdere validatieproces.

3 Het validatie model

De kern van de validatieprocedure in WavixIV wordt gevormd door een set van neurale netwerken die schattingen kunnen maken van de te valideren golfparameters door verbanden te leggen tussen deze golfparameters en parameters op andere locaties en andere tijdstippen. Aan de hand van deze schattingen kunnen incorrecte data (automatisch) opgespoord en vervangen worden.

Om de automatische schattings/validatie procedure van WavixIV te kunnen gebruiken moeten een aantal stappen doorlopen worden die in de hierna volgende secties verder uitgewerkt worden:

1. Selecteren van de data waarop het validatiemodel gekalibreerd moet worden zie sectie 3.1
2. Schatten van een regressiemodel voor het maken van initiële schattingen zie sectie 3.2
3. Specificeren en trainen van neurale netwerken zie sectie 3.3
4. Maken van een 'grove' initiële schatting van de golfparameters zie sectie 4.1
5. Toepassen van de automatische schattings/validatie procedure van WavixIV zie sectie 4.2

Aangezien de backup-boeien vanaf het komende stormseizoen 2006/2007 weggehaald zullen worden is er voor gekozen om het studiegebied op te delen in twee gedeelten die onafhankelijk van elkaar gevalideerd worden:

1. De hoofdsensoren (AMZB11, AMZB21, AMZB31, AMZB41, AMZB51) dit zijn de meetboeien die blijven liggen
2. De nevensensoren (AMZB12, AMZB22, AMZB32, AMZB42, AMZB52) dit zijn de meetboeien die verwijderd zullen worden

Op deze manier kan het regressiemodel en het model van neurale netwerken voor de hoofdsensoren in de toekomst gebruikt blijven worden.

N.B. De opdeling van het studiegebied in hoofdsensoren en nevensensoren is gespecificeerd in de bestanden hoofdsensoren.txt en nevensensoren.txt en zijn te vinden in de directory 'Stuurfiles' van de meegeleverde Cd-rom.

3.1 De trainingsdata

De neurale netwerken die gebruikt worden in het validatie proces en het regressie model voor de initiële grove schatting moeten gekalibreerd worden op historische (en gevalideerde) data zodat ze in staat zijn betrouwbare voorspellingen te doen voor de golfparameters.

Voor het stormseizoen 2005/2006 zijn er echter nog geen gevalideerde data beschikbaar voor de meetlocaties van het Amelanders zeevat. Het ontbreken van historische en gevalideerde data brengt twee problemen met zich mee:

Probleem 1:

In WavixIV worden alleen de golfparameters bijgeschat door het validatiemodel, daarbij worden naast andere golfparameters ook waterhoogte, windsnelheid en

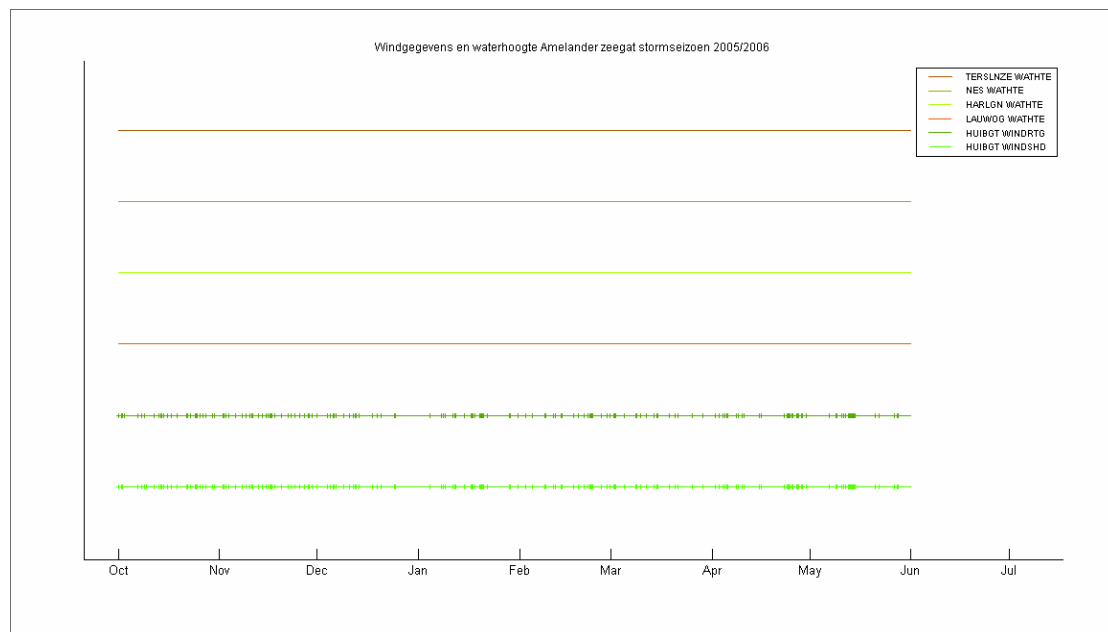
windrichting gebruikt. Omdat deze reeksen dus niet bijgeschat kunnen worden moeten ze volledig zijn. Eventuele hiaten in de reeksen van waterhoogte, windrichting en windsnelheid worden bij het importeren in WavixIV automatisch bijgeschat door middel van lineaire interpolatie.

In Figuur 2 is te zien dat de reeksen met waterhoogten volledig zijn maar dat er windgegevens ontbreken. In sommige gevallen zelfs voor zulke lange periodes (zie Figuur 3) dat de standaard lineaire interpolatie onzinnig wordt.

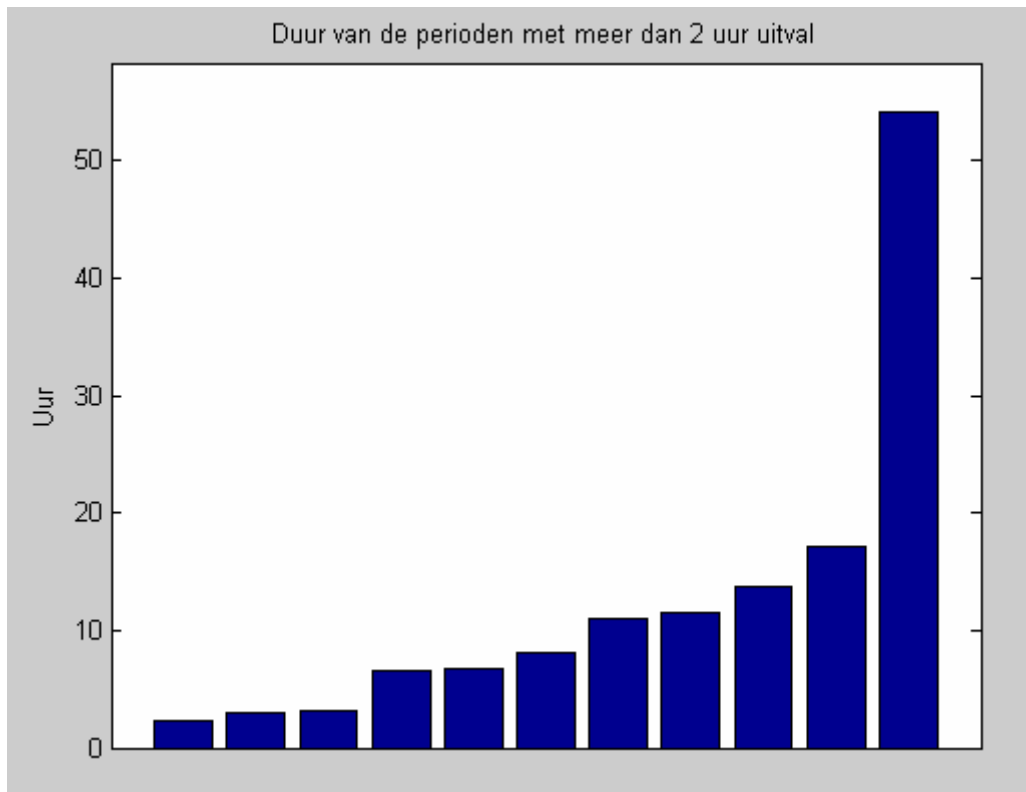
Om precies te zijn komen er 1151 hiaten (windrichting + windsnelheid) voor in de windgegevens waarbij in 17 gevallen de uitval zelfs langer dan een uur is.

Oplossing 1:

Er wordt een aparte dataset voor de hoofd- en nevensoren aangemaakt waaruit de tijdstippen waarop er geen windgegevens zijn in Huibertgat zijn weggelaten. Op deze manier zullen de fouten die optreden bij het bijschatten van metingen die afhankelijk zijn van deze windgegevens alleen veroorzaakt worden door de foutieve windrichting en windsnelheid en niet door de modellen zelf.



Figuur 2: *Hiaten in de reeksen met waterhoogte en wind.*



Figuur 3: *Duur van de perioden met meer dan 2 uur uitval.*

Probleem 2:

Normaal gesproken worden de neurale netwerken en het regressiemodel niet gekalibreerd op de te valideren dataset. De modellen worden dan immers gekalibreerd op eventuele incorrecte data. Terwijl er van de modellen verwacht wordt dat ze discrepanties aan het licht kunnen brengen doordat ze de 'ware' waarde voorspellen.

Oplossing 2:

Als aangenomen wordt dat foutieve metingen op willekeurige tijdstippen plaatsvinden en dat ze willekeurige waarden aannemen is het van groot belang dat de neurale netwerken een grote generaliserende capaciteit hebben. Met andere woorden als er een fout voorkomt in de trainingsdata mag dat maar in beperkte mate een negatief effect hebben op de prestaties van het netwerk op de testset. Om dit te bereiken wordt de parameter `max_fail` (zie sectie 3.3) op een lage waarde (5) gezet zodat afwijkingen t.o.v. de validatieset worden beperkt en wordt er voor de testset de helft van de beschikbare data gebruikt. Bovendien worden duidelijk waarneembare verdachte waarnemingen niet meegenomen in de kalibratieset.

N.B. In de directory 'Trainingsdata' van de meegeleverde Cd-rom bevinden zich de werkgebieden `hoofdsensoren_training.wv4` en `nevensensoren_training.wv4` met de data waarop de neurale netwerken en regressiemodellen gekalibreerd zijn.

3.2 Regressiemodel

Het regressiemodel is bedoeld om initiële schattingen te kunnen maken van de golfparameters. Deze 'grove' initiële schatting is noodzakelijk omdat de neurale

netwerken alleen voorspellingen kunnen doen voor alle tijdstippen als de invoerreeksen geen hiaten bevatten, deze hiaten kunnen nu dus voorlopig opgevuld worden door deze initiële schattingen.

De regressieanalyse die wordt uitgevoerd is gebaseerd op een methode van het RIKZ, waarbij gebruik wordt gemaakt van verhoudingsgetallen tussen twee golfparameters die gelden op verschillende windklassen gedifferentieerd naar windsnelheid en windrichting. Om ook onder totale uitval nog schattingen te kunnen maken zijn hieraan de verhoudingsgetallen toegevoegd tussen de betreffende golfparameter en de windsnelheid op dezelfde locatie. De gebruikte windsnelheid- en windrichtingklassen zijn gegeven in Tabel 1.

Tabel 2: *Gebruikte windrichtingklassen(kolommen) in graden en windsnelheidklassen(rijen) in m/s.*

	0-45	45-90	90-135	135-180	180-225	225-270	270-315	315-360
0-6								
6-12								
12-inf								

N.B. De geschatte regressiemodellen voor de hoofd- en nevensensoren staan in de directory 'Regressiemodel' van de meegeleverde Cd-rom in de bestanden hoofdsensoren.vhg, nevensensoren.vhg en in ASCII formaat in de bestanden hoofdsensoren_vhg.txt en nevensensoren_vhg.txt.

3.3 Specificatie van de neurale netwerken

Voor het bijschatten van golfparameters kunnen in WavixIV feedforward neurale netwerken gedefinieerd worden. Een belangrijke eigenschap die deze neurale netwerken moeten bezitten is generaliserend vermogen. De netwerken worden getraind op een set historische gegevens en de fout op deze set wordt geminimaliseerd. Het is daarbij betrekkelijk eenvoudig om precies deze gegevens te laten reproduceren door de neurale netwerken (simpelweg door voldoende neuronen en lagen in het netwerk op te nemen). Doorslaggevend is echter de fout op nieuwe gegevens waarop niet getraind is.

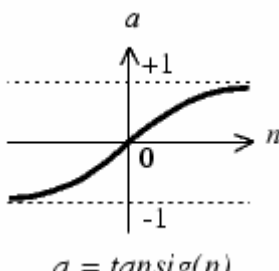
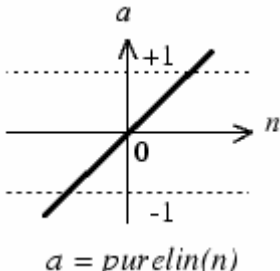
De prestaties van de neurale netwerken zijn onder andere afhankelijk van de hoeveelheid trainingsgegevens en de netwerk topologie (aantal lagen, aantal neuronen en type overgangsfuncties). Als een netwerk te groot is en er te veel parameters (gewichten) geoptimaliseerd moeten worden, kan overfitting optreden. Als een netwerk echter te klein is, zal het niet goed kunnen voorspellen. Bij overfitting zal de fout op de trainingsset heel klein worden, maar wanneer met nieuwe gegevens en voorspelling gedaan moet worden zal de fout groot zijn. Het netwerk heeft een functie 'geleerd' die specifiek is voor de trainingsset, maar niet een afspiegeling is van het onderliggende 'fysische' model en is dus niet in staat goed te generaliseren.

De methode die gebruikt is om bij de neurale netwerken voor het voorspellen van golfparameters overfitting te voorkomen en de generalisatie te verbeteren is de zogenaamde early stopping methode. De beschikbare gegevens worden in een trainingsset (40%) een testset(10%) en een validatieset(50%) opgedeeld. De netwerken worden getraind met data uit de trainingsset en tijdens het trainen wordt de fout op de validatieset gecontroleerd. Tijdens de training zal de fout op de validatieset net als de fout op de trainingsset afnemen. Als het netwerk echter begint te overfitten, zal de fout op de trainingsset blijven afnemen, terwijl de fout op de

validatieset toeneemt. Als deze stijging voortzet, zal het trainen stoppen (na max_fail (zie Tabel 3) keer een stijging van de fout op de validatieset) en de gewichten gekozen worden behorende bij het minimum van de fout op de validatieset. Met de testset wordt vervolgens de prestatie van het netwerk getest (zie Tabel 6 en Tabel 7).

Uiteindelijk is voor de feedforward neurale netwerken gekozen met de configuratie gespecificeerd zoals in Tabel 3, de belangrijkste algemene instellingen zoals gespecificeerd in Tabel 4 en de veronderstelde afhankelijkheden van de data zoals beschreven in Tabel 5.

Tabel 3: De netwerkconfiguratie

Laag 1	Laag 2
Neuronen:10	Neuronen: 1
Overgangsfunctie: tansig	Overgangsfunctie: purelin
 <p>$a = tansig(n)$</p>	 <p>$a = purelin(n)$</p>

Tabel 4: De belangrijkste algemene netwerkinstellingen.

Parameter	Waarde	Commentaar
Trainingsfunctie	Trainbfg	Quasi Newton optimalisatie algoritme met linesearch (srchbac)
Doelfunctie	MSE	Mean Squared Error
Pca drempel	0	Er wordt geen dimensie reductie uitgevoerd op de trainingsdata
Testset	0.50	50 procent van de data wordt gebruikt voor het meten van de prestaties van de netwerken. De netwerken zijn geheel onafhankelijk van deze data getraind.
Validatieset	0.10	10 procent van de data wordt gebruikt om tijdens de training de prestaties van de netwerken te bepalen. Als de prestaties van de netwerken verbeteren op de trainingsset maar verslechteren op de validatieset wordt gestopt.
Max_fail	5	Aantal keren dat de fout op de validatieset mag stijgen t.o.v. de trainingset.

Tabel 5: Veronderstelde afhankelijkheden per golfparameter.

Variabele	Afhankelijkheden	Relatieve tijdstippen (min)
H1/3	H1/3 op het station zelf	-6 -3 -1 1 3
	Hm0 op het station zelf	-6 -3 -1 0 1 3
	HTE3 op het station zelf	0
	Wind op het station Huibertgat	0
	Wathte op de stations Nes en Terslnze	0
HTE3	HTE3 op het station zelf	-6 -3 -1 1 3
	Tm02 op het station zelf	-6 -3 0 3
	Hm0 op het station zelf	-6 -3 0 3
	HTE3 op de twee dichtstbijzijnde stations	-3 0 3
	Tm02 op de twee dichtstbijzijnde stations	-3 0 3
	Hm0 op de twee dichtstbijzijnde stations	-3 0 3
	Wind op het station Huibertgat	-18 -12 -6 0
	Wathte op dichtstbijzijnde station	-6 -3 0 3 6
Hm0	Hm0 op het station zelf	-6 -3 -1 1 3
	Hm0 op de twee dichtstbijzijnde stations	-6 -3 0 3
	Wind op het station Huibertgat	-18 -12 -6 0
	Wathte op dichtstbijzijnde station	-6 -3 -1 0 1 3 6
TH1/3	TH1/3 op het station zelf	-3 -1 1 3
	Hm0 op het station zelf	-3 -1 0 1 3
	Tm02 op het station zelf	-3 -1 0 1 3
	H1/3 op het station zelf	-3 -1 0 1 3
	TH1/3 op de twee dichtstbijzijnde stations	-3 0 3
	Wind op het station Huibertgat	-18 -12 -6 0
	Wathte op de stations Nes en Terslnze	-6 -3 0 3 6
Th0	Th0 op het station zelf	-6 -3 -1 1 3
	Th0 op de twee dichtstbijzijnde stations	-6 -3 0 3
	Wind op het station Huibertgat	-18 -12 -6 -3 0
	Wathte op de stations Nes en Terslnze	-6 -3 -1 0 1 3 6
Tm02	Tm02 op het station zelf	-6 -3 -1 1 3
	Hm0 op het station zelf	-6 -3 -1 0 1 3
	Tm02 op de twee dichtstbijzijnde stations	-6 -3 0 3
	Hm0 op de twee dichtstbijzijnde stations	-6 -3 0 3
	Wind op het station Huibgt	-18 -12 -6 0
	Wathte op de stations Nes en Terslnze	-6 -3 -1 0 1 3 6

Ten behoeve van de kalibratie van de neurale netwerken worden de tijdreeksen met een richtingskarakter (wind en golfrichting) ontbonden in een noordzuid en oostwest component. Deze noordzuid en oostwest representatie blijven echter onzichtbaar voor de gebruiker.

N.B. De complete specificatie van de neurale netwerken is te vinden in de directory 'Netwerken' van de meegeleverde CD-rom in de netwerkbestanden amelander_zeegat_hoofdsensoren.asc en amelander_zeegat_nevensensoren.asc

In Tabel 6 en Tabel 7 zijn de resultaten vermeld van de prestaties van de neurale netwerken op het voorspellen van de golfparameters op de locaties met de hoofd- en nevensensoren. Als norm is de (R)oot (M)ean (S)quared (E)rror gebruikt van de voorspelling van de neurale netwerken en de gemeten waarden uit de testset.

Tabel 6: Prestaties van de neurale netwerken op de hoofdsensoren.

Locatie	Variabele	Veldapparaat	RMSE op testset
AMLZGTB11	H1/3	WAVERDDR	1.87
AMLZGTB21	H1/3	WAVERDDR	1.00
AMLZGTB31	H1/3	WAVERDDR	0.84
AMLZGTB41	H1/3	WAVERDDR	0.57
AMLZGTB51	H1/3	WAVERDR	0.66
AMLZGTB11	Hm0	WAVERDDR	4.17
AMLZGTB21	Hm0	WAVERDDR	1.90
AMLZGTB31	Hm0	WAVERDDR	1.56
AMLZGTB41	Hm0	WAVERDDR	1.08
AMLZGTB51	Hm0	WAVERDR	0.97
AMLZGTB11	HTE3	WAVERDDR	2.66
AMLZGTB21	HTE3	WAVERDDR	1.25
AMLZGTB31	HTE3	WAVERDDR	0.81
AMLZGTB41	HTE3	WAVERDDR	0.37
AMLZGTB51	HTE3	WAVERDR	0.37
AMLZGTB11	Th0	WAVERDDR	0.36
AMLZGTB21	Th0	WAVERDDR	0.06
AMLZGTB31	Th0	WAVERDDR	0.05
AMLZGTB41	Th0	WAVERDDR	0.08
AMLZGTB11	TH1/3	WAVERDDR	0.15
AMLZGTB21	TH1/3	WAVERDDR	0.13
AMLZGTB31	TH1/3	WAVERDDR	0.12
AMLZGTB41	TH1/3	WAVERDDR	0.11
AMLZGTB51	TH1/3	WAVERDR	0.15
AMLZGTB11	Tm02	WAVERDDR	0.07
AMLZGTB21	Tm02	WAVERDDR	0.07
AMLZGTB31	Tm02	WAVERDDR	0.07
AMLZGTB41	Tm02	WAVERDDR	0.07
AMLZGTB51	Tm02	WAVERDR	0.09

Tabel 7: Prestaties van de neurale netwerken op de nevensensoren.

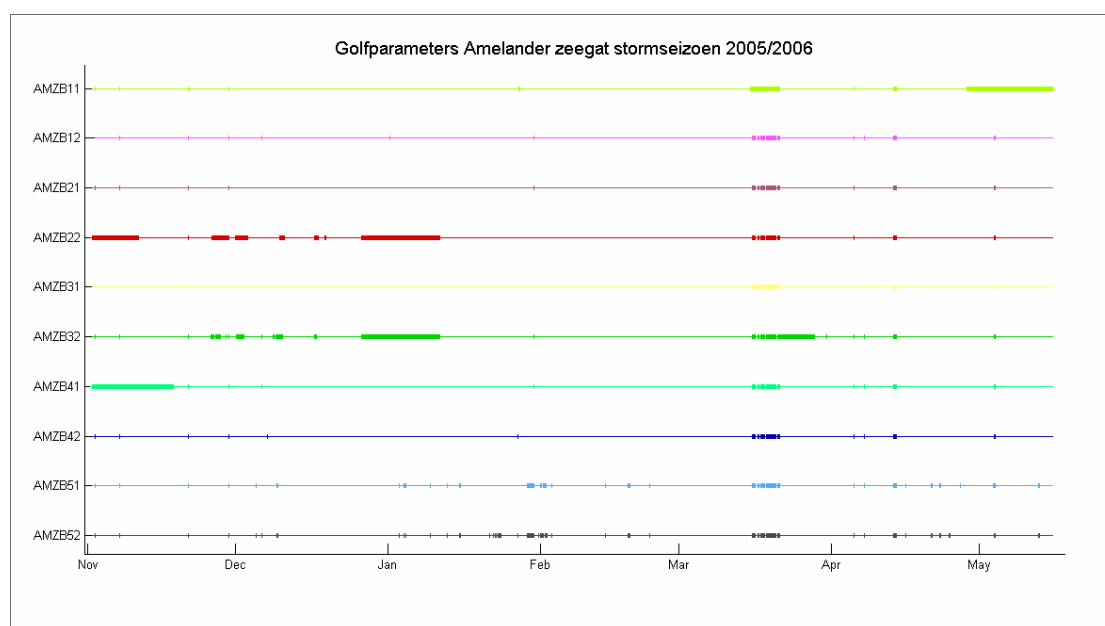
Locatie	Variabele	Veldapparaat	RMSE op testset
AMLZG12	Hm0	WAVERDDR	4.25
AMLZG22	Hm0	WAVERDDR	1.85
AMLZG32	Hm0	WAVERDR	1.53
AMLZG42	Hm0	WAVERDR	1.04
AMLZG52	Hm0	WAVERDR	0.88
AMLZG12	Tm02	WAVERDDR	0.07
AMLZG22	Tm02	WAVERDDR	0.07
AMLZG32	Tm02	WAVERDR	0.07
AMLZG42	Tm02	WAVERDR	0.11
AMLZG52	Tm02	WAVERDR	0.11
AMLZG12	HTE3	WAVERDDR	1.98
AMLZG22	HTE3	WAVERDDR	1.14
AMLZG32	HTE3	WAVERDR	0.78
AMLZG42	HTE3	WAVERDR	0.41
AMLZG52	HTE3	WAVERDR	0.32
AMLZG12	H1/3	WAVERDDR	1.79
AMLZG22	H1/3	WAVERDDR	0.98
AMLZG32	H1/3	WAVERDR	0.83
AMLZG42	H1/3	WAVERDR	0.66
AMLZG52	H1/3	WAVERDR	0.62
AMLZG12	TH1/3	WAVERDDR	0.12
AMLZG22	TH1/3	WAVERDDR	0.12
AMLZG32	TH1/3	WAVERDR	0.10
AMLZG42	TH1/3	WAVERDR	0.14
AMLZG52	TH1/3	WAVERDR	0.18
AMLZG12	Th0	WAVERDDR	0.04
AMLZG22	Th0	WAVERDDR	0.08

N.B. De complete specificatie van de neurale netwerken is te vinden in de netwerk bestanden `amelander_zeegat_hoofdsensoren.asc` en `amelander_zeegat_nevensensoren.asc` in de directory 'Netwerken' van de meegeleverde CD-rom.

4 Validatie

Nu de te valideren dataset bekend is en de validatie modellen gekalibreerd zijn kan de validatie procedure gestart worden. In Figuur 4 is te zien op welke tijdstippen de golfparameters bijgeschat moeten worden (dikke lijnstukken) en in Tabel 8 en Tabel 9 is te zien hoeveel waarden voor elke locatie bijgeschat moeten worden.

De meetboeien hebben over het algemeen goed gefunctioneerd, gemiddeld over de meetboeien minder dan 10% uitval. Maar het valt wel op dat vooral de meetboeien op de locatie AZB22 (19%) en AZB32(15%) falen. Bovendien is op te merken dat uitval van meetboeien vaak simultaan gebeurt zoals bijvoorbeeld in maart 2006 voor alle meetboeien het geval is geweest.



Figuur 4: Hiaten in de reeksen met de golfparameters.

Tabel 8: Totaal aantal hiaten voor alle reeksen per locatie voor de hoofdsensoren.

Locatie	Aantal hiaten	Percentage hiaten	Aantal Waarnemingen
AMLDZGTB11	20520	12 %	168486
AMLDZGTB21	4088	2 %	168486
AMLDZGTB31	4114	2 %	168486
AMLDZGTB41	18291	11 %	168486
AMLDZGTB51	4730	3 %	140405
TOTAAL	51743	6 %	814349

Tabel 9: Totaal aantal hiaten voor alle reeksen per locatie voor de nevensensoren.

Locatie	Aantal hiaten	Percentage hiaten	Aantal Waarnemingen
AMLDZGTB12	4630	3 %	168486
AMLDZGTB22	32140	19 %	168486
AMLDZGTB32	21689	15 %	140405
AMLDZGTB42	3435	2 %	140405
AMLDZGTB52	5024	4 %	140405
TOTAAL	66918	9 %	758187

De validatie procedure bestaat nu uit twee stappen:

1. Maken van initiële schattingen en detecteren outliers
zie sectie 4.1
2. Optimaliseren van de schattingen met behulp van de conhop procedure
zie sectie 4.2

N.B. In de directory 'Validatie/Start' van de meegeleverde Cd-rom zijn twee werkgebieden en twee .dia bestanden te vinden waarin de te valideren en bij te schatten metingen op hiaat gezet zijn.

4.1 Initiële schattingen

De initiële schattingen zijn bedoeld om een startpunt te verkrijgen voor de automatische validatie procedure (conhop) en eventuele verdachte metingen alsnog af te keuren en mee te nemen in de automatische validatieprocedure. De initiële schattingen worden in drie stappen berekend:

1. Toepassen van het regressiemodel
een 'grove' initiële schatting maakt het mogelijk om de neurale netwerken te gebruiken voor het schatten van de golfparameters
2. Verbeteren initiële schatting
de neurale netwerken worden één keer toegepast, de enkelvoudige hiaten zijn nu bijgeschat.
3. Toepassen van de fast-repair procedure
de neurale netwerken worden enkele keren (drie keer in dit project) toegepast op de overgebleven hiaten om een betere startoplossing te krijgen. Zodat de laatste stap (conhop) sneller verloopt.

N.B. In de directory 'Validatie/Initiele schatting' van de meegeleverde Cd-rom zijn twee werkgebieden met daarin de initiële schattingen van de golfparameters en de uiteindelijke set hiaten die automatisch verwerkt moeten worden.

4.2 Optimaliseren van wederzijdse hiaten (conhop)

De laatste stap in het validatieproces is het automatisch bij-schatten van de overgebleven (wederzijdse) hiaten en de hiaten die ontstaan zijn in de vorige stap (initiële schattingen) door gemeten waarden alsnog af te keuren op grond van de initiële schatting.

De overgebleven hiaten kunnen normaalgesproken niet in één keer bijgeschat kunnen worden omdat ze onderling van elkaar afhankelijk zijn. WavixIV schat daarom met behulp van het model van neurale netwerken de wederzijdse hiaten. door die waarden voor de hiaten te zoeken die gegeven de gemeten waarden en geschatte betrouwbaarheid de grootste kans hebben om op te treden.

Omdat het aantal hiaten dat nog over is gebleven na het uitvoeren van de initiële schatting is het werkgebied opgedeeld in twee onafhankelijke perioden (02-11-2005 tot 05-02-2006 en 05-02-2006 tot 16-05-2006).

N.B. In de directory 'Validatie/Gevalideerd' van de meegeleverde Cd-rom zijn twee .dia bestanden met daarin de gevalideerde golfparameters voor de hoofdsensoren en voor de nevensensoren.

5 Kwaliteit van het validatiemodel

Als verondersteld wordt dat de in sectie 4 verkregen gevalideerde dataset ook de echte dataset representeert kan de kwaliteit van de schattingen van het validatiemodel bepaald worden door hiaten in de dataset aan te brengen en deze hiaten vervolgens weer met behulp van het validatiemodel weer bij te schatten.

De hiaten worden op drie verschillende manieren aangebracht in de gevalideerde dataset:

1. Hiaten worden random gekozen over alle reeksen en locaties zodanig dat ongeveer tien procent van het totale aantal metingen hiaat is, deze procedure wordt vijf keer herhaald
2. Random worden voor perioden van ongeveer één dag alle reeksen op één locatie (in totaal 20 dagen) op hiaat gezet, deze procedure wordt herhaald voor alle locaties
3. Random worden voor perioden van ongeveer één dag (in totaal 4 dagen) alle reeksen voor alle stations op hiaat gezet

In Tabel 10 en Tabel 11 worden voor elke van deze drie manieren μ (het gemiddelde verschil tussen waarneming en schatting) en σ (de standaarddeviatie van het verschil tussen de waarneming en de schatting) weergegeven.

Om een indruk te krijgen van de kwaliteit van de door het neurale netwerk model bijgeschatte hiaten tijdens 'zwaar' weer is de analyse tevens gesplitst in een gedeelte onder 'normaal' weer (Tabel 10) en een gedeelte onder 'zwaar' weer (Tabel 11). De perioden met zwaar weer zijn:

- 15 december 2005 tot 18 december 2005
- 8 februari 2006 tot 11 februari 2006

Uit de tabellen is af te leiden dat zowel voor 'zwaar' als voor 'normaal' weer de validatiemodellen in staat zijn om nauwkeurig hiaten bij te schatten voor het geval dat de hiaten gekozen zijn volgens de eerste manier. Dit komt doordat een groot deel van de op deze manier gekozen hiaten onafhankelijk zijn van de overige hiaten en dus rechtstreeks kunnen worden bijgeschat.

De tweede manier van hiaatselectie is echter reëler, met deze manier wordt nagebootst dat een station voor een langere periode compleet uitvalt. Het validatiemodel presteert over de hele linie slechter.

Bij de laatste manier van hiaatselectie wordt nagebootst dat voor een langere periode alle stations tegelijkertijd uitvallen. Te zien is dat de geschatte waarden een steeds grotere systematische fout en spreiding laten zien. De kwaliteit van het validatiemodel rust nu voor een groot deel op de kwaliteit en aanwezigheid van de wind- en waterhoogte gegevens.

Tabel 10: *Kwaliteit van de bijgeschatte golfparameters.*

Locatie	Variabele	Case 1		Case 2		Case 3	
		μ	σ	μ	σ	μ	σ
AMZB11	H1/3	0.69	4.06	3.65	5.16	-1.32	5.69
AMZB11	HTE3	0.50	2.16	-3.20	2.34	-1.18	2.98
AMZB11	Hm0	0.97	6.21	1.71	5.95	-1.11	7.48
AMZB11	TH1/3	0.00	0.15	-0.33	0.93	-0.15	1.50
AMZB11	Th0	0.78	39.02	-10.72	9.56	-3.24	5.18
AMZB11	Tm02	0.00	0.09	-0.71	1.02	-0.75	0.99
AMZB21	H1/3	0.68	2.93	4.17	3.68	3.22	5.46
AMZB21	HTE3	0.61	1.91	1.73	5.96	0.28	11.53
AMZB21	Hm0	0.87	3.66	5.31	4.27	-7.02	11.20
AMZB21	TH1/3	0.00	0.16	0.04	0.69	0.19	1.74
AMZB21	Th0	0.75	7.48	-9.92	7.57	6.76	6.62
AMZB21	Tm02	0.00	0.09	0.03	0.44	-0.09	1.05
AMZB31	H1/3	0.55	2.01	0.26	1.86	-9.10	13.03
AMZB31	HTE3	0.54	1.16	0.74	3.41	-0.33	6.32
AMZB31	Hm0	0.66	2.71	0.88	12.71	-9.45	14.90
AMZB31	TH1/3	0.00	0.16	0.04	0.60	0.26	1.63
AMZB31	Th0	1.21	8.23	6.12	9.47	5.61	5.09
AMZB31	Tm02	0.00	0.09	0.03	0.42	0.35	1.04
AMZB41	H1/3	0.47	0.93	1.58	6.04	-2.30	8.15
AMZB41	HTE3	0.48	0.56	0.82	0.82	0.29	0.83
AMZB41	Hm0	0.50	1.55	1.99	6.28	-2.39	9.04
AMZB41	TH1/3	0.00	0.12	0.08	0.57	0.06	1.28
AMZB41	Th0	0.86	8.57	3.11	13.47	4.25	13.52
AMZB41	Tm02	0.00	0.09	0.08	0.48	0.24	0.94
AMZB51	H1/3	0.43	1.17	0.57	5.84	-0.33	7.27
AMZB51	HTE3	0.37	0.50	0.47	0.73	0.38	0.62
AMZB51	Hm0	0.38	1.12	0.03	5.99	-0.40	7.85
AMZB51	TH1/3	0.00	0.16	0.13	0.69	0.29	1.00
AMZB51	Tm02	0.00	0.11	0.06	0.53	0.04	0.69

Tabel 11: *Kwaliteit van de bijgeschatte golfparameters tijdens 'zwaar' weer.*

Locatie	Variabele	Case 1		Case 2		Case 3	
		μ	σ	μ	σ	μ	σ
AMZB11	H1/3	1.28	4.23	5.76	7.30	7.45	9.31
AMZB11	HTE3	1.33	5.15	6.21	9.66	8.86	9.90
AMZB11	Hm0	3.37	7.32	7.05	7.12	9.85	10.75
AMZB11	TH1/3	0.02	0.18	0.77	1.38	0.62	1.91
AMZB11	Th0	0.72	43.01	14.25	15.17	15.56	16.61
AMZB11	Tm02	0.01	0.12	0.42	1.01	0.80	1.13
AMZB21	H1/3	0.28	2.33	-1.19	2.15	2.39	3.31
AMZB21	HTE3	0.61	2.47	1.52	9.72	9.25	19.60
AMZB21	Hm0	0.74	4.03	0.43	4.29	1.30	4.90
AMZB21	TH1/3	0.01	0.15	-0.40	0.99	-0.20	1.33
AMZB21	Th0	0.49	2.11	-6.04	6.51	-6.38	7.15
AMZB21	Tm02	0.00	0.08	-0.02	0.34	0.14	0.77
AMZB31	H1/3	0.55	1.83	3.20	7.28	8.09	12.57
AMZB31	HTE3	0.65	2.00	4.92	6.65	6.87	7.39
AMZB31	Hm0	0.41	3.65	3.92	8.89	5.97	8.23
AMZB31	TH1/3	0.00	0.15	0.07	0.42	0.10	1.07
AMZB31	Th0	0.28	3.12	-2.79	30.37	-3.90	31.07
AMZB31	Tm02	0.00	0.07	0.03	0.37	-0.05	0.89
AMZB41	H1/3	0.42	1.43	4.35	15.12	8.88	18.39
AMZB41	HTE3	0.51	0.67	1.08	2.07	2.36	3.20
AMZB41	Hm0	0.67	3.34	4.34	16.48	9.46	19.92
AMZB41	TH1/3	0.00	0.12	-0.13	0.55	-0.22	1.07
AMZB41	Th0	0.35	1.71	-5.17	20.58	-9.92	20.16
AMZB41	Tm02	-0.01	0.08	-0.07	0.54	-0.04	0.81
AMZB51	H1/3	0.60	1.62	2.09	8.76	7.31	15.44
AMZB51	HTE3	0.45	0.57	0.21	0.65	0.59	0.92
AMZB51	Hm0	0.42	2.06	2.26	10.43	7.51	17.35
AMZB51	TH1/3	-0.02	0.15	-0.06	0.33	-0.14	0.63
AMZB51	Tm02	-0.01	0.07	-0.03	0.25	-0.08	0.55

6 Aanbevelingen

In dit project is voor het eerst gebruik gemaakt van WavixIV voor het valideren van golfparameters. Dit heeft geleid tot een aantal aanpassingen en verbeteringen ten opzichte van de WavixIV versie zoals die in 2005 ontwikkeld is. Naast deze verbeteringen zijn er echter ook nog een aantal noodzakelijke toekomstige verbeteringen die nog niet tijdens dit project gerealiseerd konden worden.

De eerste opzet van WavixIV was het valideren van golfgegevens op uurlijkse basis. Voor dit project echter was het noodzakelijk om de golfparameters op tien minuten basis te valideren. WavixIV is daarom uitgebreid met de mogelijkheid om golfparameters te valideren die op een willekeurig tijdsinterval worden ingewonnen.

Aangezien frequenter inwinnen van gegevens meer systeembronnen vereist is tevens de optimalisatie routine voor het automatisch valideren (conhop) geoptimaliseerd. Konden voorheen maximaal 8.000 hiaten tegelijkertijd bijgeschat worden thans ligt deze grens op 80.000 hiaten.

Tijdens het ontwerp van WavixIV is als harde randvoorwaarde gesteld dat de windrichting en windsterkte volledig zijn. Deze windgegevens worden normaliter geleverd door het KNMI. In de praktijk echter is het KNMI niet in staat om de windgegevens op te tijd te leveren. WavixIV is daarom uitgebreid met de mogelijkheid om windgegevens door middel van lineaire interpolatie aan te vullen. Het is echter gebleken dat de windgegevens voor langere perioden kunnen ontbreken waardoor lineaire interpolatie onbruikbaar wordt. In een volgende versie van WavixIV zouden daarom de windgegevens geschat kunnen worden met behulp van neurale netwerken die verbanden leggen tussen de wind op een bepaalde locatie en de wind op de andere locaties. Aangezien de wind voor dit project alleen beschikbaar was voor de locatie Huibertgat is deze uitbreiding nog niet verder uitgewerkt.

Bij de eerste versie van WavixIV is er vanuit gegaan dat alle te valideren golfparameters gemeten worden op elke locatie. In dit project echter wordt de gemiddelde golfrichting (θ_0) niet op elke locatie gemeten. De eis dat alle hoofdsensoren gespecificeerd moeten worden is in de nieuwe versie van WavixIV dan ook komen te vervallen.

Bij simultane uitval van één of meerdere stations is te zien dat de kwaliteit van de bijgeschatte hiaten afneemt. Aangezien het validatiemodel onder deze omstandigheden afhankelijk is van de wind- en waterhoogtegegevens verdient het de aanbeveling om in de toekomst het validatiemodel te kalibreren met meer dan één locatie met windgegevens.