

Wavix 2004

Functioneel Ontwerp

30 September 2003
In opdracht van RIKZ

In opdracht van: RIKZ
Postbus 20907
2500 EX Den Haag
Contactpersoon: Ing. E.R.A. Marsman
ITA-RIKZ
+31 70 3114 518

Uitgevoerd door: Modelit
Rotterdamse Rijweg 126
3042 AS Rotterdam
www.modelit.nl
Contactpersoon: Dr.ir. N.J. van der Zijpp
Zijpp@modelit.nl
+31 10 4623 621

Inhoud

<u>1 Inleiding</u>	<u>4</u>
<u>1.1 Voorgeschiedenis</u>	<u>4</u>
<u>1.2 Kenmerken van de te valideren data</u>	<u>4</u>
<u>1.3 Doelstelling nieuwe software</u>	<u>4</u>
<u>1.4 Systeem afbakening</u>	<u>5</u>
<u>1.5 Doelstelling van het functioneel ontwerp</u>	<u>6</u>
<u>2 Golfklimaatgegevens</u>	<u>8</u>
<u>2.1 Het karakteriseren van golven in het tijdsdomein</u>	<u>8</u>
<u>2.2 Het karakteriseren van golven in het frequentiedomein</u>	<u>9</u>
<u>3 Werkproces</u>	<u>12</u>
<u>3.1 Inleiding</u>	<u>12</u>
<u>3.2 Werkproces: Off-line validatie</u>	<u>12</u>
<u>3.3 Werkproces: Modelbouw</u>	<u>13</u>
<u>4 Beschrijving software modules voor operationele validatie</u>	<u>17</u>
<u>4.1 Inleiding</u>	<u>17</u>
<u>4.2 Visuele controle van invoergegevens</u>	<u>17</u>
<u>4.3 Completeren reeksen</u>	<u>19</u>
<u>4.4 Berekening voorspelde waarden</u>	<u>20</u>
<u>4.5 Identificeren outliers</u>	<u>21</u>
<u>4.6 Genereren verslag met outliers</u>	<u>22</u>
<u>4.7 Interactief aanpassen: valideren</u>	<u>23</u>
<u>4.8 Inhoud van het inspectiescherm voor golfparameters</u>	<u>23</u>
<u>5 Beschrijving software modules voor modelbouw</u>	<u>30</u>
<u>5.1 Toe te passen model</u>	<u>30</u>
<u>5.2 Software modules</u>	<u>34</u>
<u>6 Beschrijving software modules voor invoer en uitvoer</u>	<u>38</u>
<u>6.1 Externe relaties</u>	<u>38</u>
<u>6.2 Huidige situatie: Dia2Wavix en Wavix2DIA</u>	<u>38</u>
<u>6.3 Nieuwe situatie: invoer</u>	<u>40</u>
<u>6.4 Nieuwe situatie: uitvoer</u>	<u>40</u>
<u>6.5 Knmi2Dia</u>	<u>40</u>
<u>6.6 ASCII export optie</u>	<u>41</u>
<u>7 Technische aspecten</u>	<u>42</u>
<u>7.1 Eisen aan de gebruikers interface</u>	<u>42</u>
<u>7.2 Voorwaarden aan de implementatiewijze</u>	<u>42</u>
<u>7.3 Beheer en onderhoud</u>	<u>42</u>
<u>8 Literatuur verwijzingen</u>	<u>44</u>

1 Inleiding

1.1 Voorgeschiedenis

Sinds 1997 is door RIKZ ervaring opgedaan met het valideren van golfgegevens met behulp van een in opdracht van RIKZ ontwikkelde methode, genaamd Wavix. De vraag die bij de ontwikkeling van Wavix centraal stond was:

“Is het mogelijk om golfgegevens te valideren en aan te vullen door verbanden te leggen tussen windrichting, windsterkte, en golfhoogtes voor een aantal locaties, en aan de hand van deze verbanden ramingen (inclusief betrouwbaarheidsmarge) te maken van golfhoogtes voor de te valideren locaties en periodes”

Gezien de hoge mate van complexiteit van de verbanden tussen de genoemde grootheden, is niet gekozen voor expliciete fysische modellering, maar voor een beschrijvende aanpak op basis van neurale netwerken. Dit heeft geresulteerd in een de ontwikkeling van een verzameling modellen die zijn toegesneden op de verwerking van de gegevens van 10 locaties in de Noordzee. Deze modellen vormen de basis voor het WAVIX systeem waarmee sinds 1997 de golfgegevens voor deze locaties worden gevalideerd en dat voor de verwerking van gegevens voor deze locaties goed voldoet.

1.2 Kenmerken van de te valideren data

De golfklimaatgegevens bestaan uit golfhoogte, golfperiode en golfrichting. Deze gegevens worden waargenomen met gespecialiseerde apparatuur zoals stappenbaken, meetboeien en radar. De meetgegevens worden via een directe radio verbinding doorgegeven aan de verwerkingscentra. Met een beperkt aantal meetlocaties wordt over het algemeen een groot gebied bestreken, binnen het huidige WAVIX worden momenteel bijvoorbeeld tien meetlocaties gebruikt. Vooral in reeksen van 10 jaar en ouder komen veelvuldig hiaten voor. Vaak is ook sprake van een gelijktijdige uitval van alle detectoren.

1.3 Doelstelling nieuwe software

Nadat met het huidige Wavix systeem ervaring is opgedaan bestaat bij RIKZ nu de behoefte een soortgelijk systeem toe te passen voor de validatie van golfgegevens die afkomstig zijn van andere locaties. De huidige Wavix versie is alleen geschikt voor het valideren van de gegevens van de voornoemde tien locaties en bevat geen functies om het systeem te kalibreren voor andere locaties. In de nieuwe situatie moet het mogelijk worden om het systeem voor willekeurige configuratie van meetlocaties geschikt te maken, zonder daarbij van externe experts afhankelijk te zijn. Om dit mogelijk te maken dient een flexibele Wavix versie (werktitel: WAVIX 2004) ontwikkeld te worden die aan de volgende eisen voldoet:

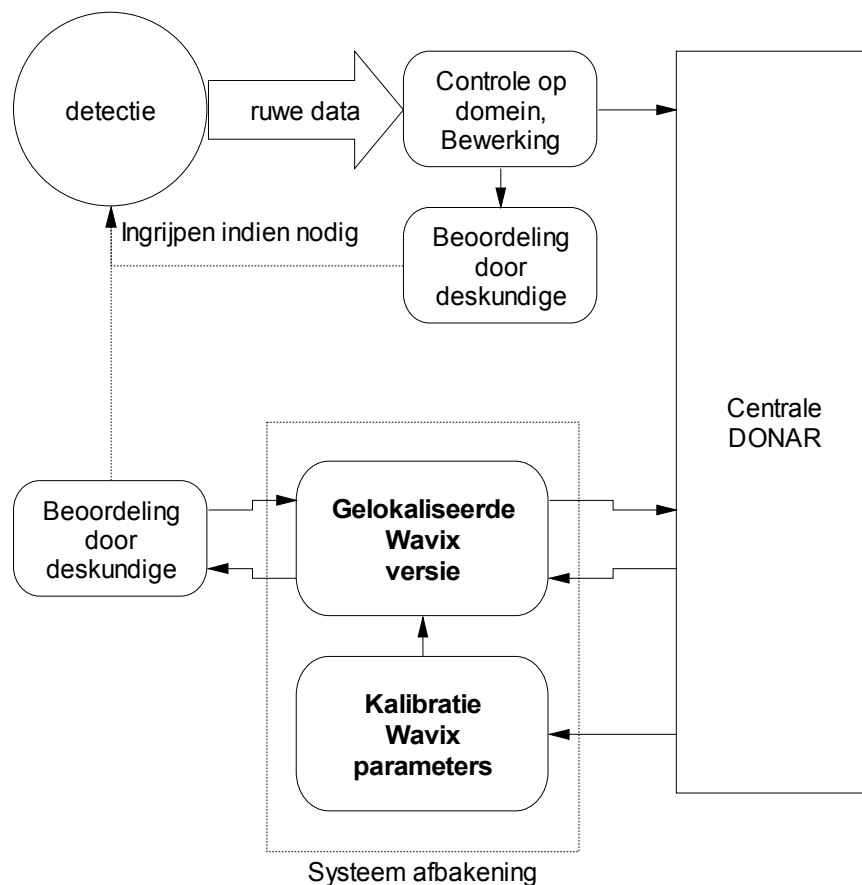
- De software moet in ruime mate generiek zijn:
 - Het model moet overweg kunnen met een geparparameteriseerde weergave van detector configuraties en relevante omgevingsfactoren;
 - De in- en uitvoerformaten van het model mogen een brede toepassing van het model niet in de weg staan;
- Het werkproces dat in de afgelopen jaren is uitgekristalliseerd moet op optimale wijze worden ondersteund;
- Naast de functionaliteit van het huidige WAVIX ook moet de software ook functionaliteit bevatten voor het kiezen van een model structuur en het inregelen van model parameters.

1.4 Systeem afbakening

Het onderstaande figuur geeft de systeem afbakening weer. De eerste controle en bewerking op de ruwe data vindt plaats in een gespecialiseerd systeem. In dit systeem worden de golfparameters gegenereerd zoals de significante golfhoogte, de laagfrequente golfhoogte en de gemiddelde golfperiode. Vervolgens worden de gegevens in DONAR opgeslagen.

De Golven Validatie begint met het ophalen met de golfgegevens uit DONAR of uit een fixed-format ASCII bestand. Op basis van deze gegevens worden overzichten met behulp van een Gelokaliseerde Wavix Versie gegenereerd en verdachte meetgegevens opgespoord. Na beoordeling door een deskundige worden gegevens geaccepteerd, verworpen of aangepast. De aangepaste dataset wordt vervolgens teruggeplaatst in DONAR, of geëxporteerd naar een fixed-format ASCII bestand.

De kalibratie van de parameters in de Gelokaliseerde Wavix Versie is een onderdeel van het te bouwen systeem. Ook de gegevens die voor de kalibratie nodig zijn worden uit DONAR betrokken.



1.5 Doelstelling van het functioneel ontwerp

Dit functioneel ontwerp moet antwoord geven op de volgende vragen:

- welke functies moet het systeem bevatten?
- hoe kunnen deze functies worden opgesplitst in deelfuncties?
 - wat is de in- en uitvoer per deelfunctie?
 - wat zijn de belangrijkste rekenkundige bewerkingen per deelfunctie?
- welke externe in- en uitvoer heeft het systeem:
 - welke stuurfiles worden gebruikt?
 - welke invoerfiles worden gebruikt?
 - welke uitvoerfiles worden gebruikt?
 - welke rapporten worden gegenereerd?
 - welke user interface-schermen zijn aanwezig?
- in welke hardware omgeving moet het systeem functioneren?

De meerwaarde van FO moet vooral liggen in het vinden van een optimale trade-off tussen de mate waarin het systeem generiek moet zijn en de specialistische kennis en de inspanning die vereist zal zijn om met het systeem te werken.

Zeker wanneer gebruik gemaakt wordt van een abstracte modelbeschrijving op basis van neurale netwerken, zoals in de huidige WAVIX programmatuur, lijkt een generieke oplossing mogelijk. Bij een generieke oplossing maakt het strikt genomen niet uit of het gaat om significante golfhogte of, bijvoorbeeld, temperatuurmetingen. Vanuit dit gezichtspunt hebben we te maken met een verzameling detectoren waarvoor het volgende geldt:

- de detectoren bevinden zich op verschillende locaties;
- deze detectoren meten tijdreeksen van golfparameters (hoogte, periode en richting);
- er bestaat een sterk verband tussen golfparameters voor één locatie;
- er bestaat een verband tussen golfparameters van verschillende geografische locaties;
- dit verband is duidelijker naarmate de locaties zich dichterbij elkaar zijn;
- er bestaat een verband tussen golfparameters van verschillende tijdstippen;
- dit verband is sterker naarmate de tussenliggende periode korter is.

Op basis van deze aannames zou het mogelijk moeten zijn om een of-the-shelf oplossing, zoals de Neural Network Toolbox voor Matlab, toe te passen, en moet hooguit voorzien worden in invoer- en uitvoermodules waarmee data uitwisseling met DONAR wordt gerealiseerd. Een dergelijke aanpak brengt echter wel met zich mee dat vele ontwerp beslissingen over het toe te passen model aan de gebruiker worden overgelaten, dat alleen gebruik gemaakt kan worden van algemene grafische weergaven, en dat het niet mogelijk is om specifieke materiekennis die betrekking heeft op golfklimaat gegevens op voorhand in te brengen.

Het andere uiterste van het spectrum is dat een volledige maatwerk oplossing wordt toegepast, waarin de invloed van de gebruiker is beperkt tot het al dan niet selecteren van voorgeprogrammeerde opties. Een dergelijke oplossing doet precies wat op het moment van implementatie vereist is maar is niet in andere situaties toe te passen, zelfs als de nieuwe situatie sterk lijkt op de oude.

Het ambitieniveau van het huidige Functionele Ontwerp ligt ergens tussen deze twee uitersten in en op de volgende wijze gekenschetst:

- Het toepassingsgebied is duidelijk beperkt tot het valideren van golfklimaatgegevens;
- Het systeem moet toepasbaar zijn in situaties die sterk verschillen voor wat betreft de waterdiepte, de nabijheid van land, het type detectoren en de onderlinge afstand

tussen deze detectoren. Hierdoor zal de rangschikking van invloedsfactoren belangrijk kunnen verschillen van situatie tot situatie;

- De kalibratie van modelparameters (bij neurale netwerken: het inleren van het systeem) vindt binnen de applicatie plaats;
- Materiekennis over golfgegevens, zoals vuistregels voor het schatten van ontbrekende golfparameters op basis van overige gegevens, zijn in het systeem ondergebracht;
- Het systeem wordt ontwikkeld op basis van de behoefte van RIKZ, maar zou in principe ook moeten kunnen worden toegepast bij andere diensten die zich met de validatie van Golfklimaatgegevens bezighouden, zowel binnen als buiten Nederland.

2 Golfklimaatgegevens

Een golf is een sinusoidale beweging die wordt gekenmerkt door een golfhoogte, golfperiode en een voortplantingsrichting. In praktijk wordt een superpositie van vele golven waargenomen en is het nodig de situatie aan de hand van een aantal afgeleide grootheden te karakteriseren. Dit kan zowel gebeuren in het tijdsdomein als in het frequentiedomein zoals in de onderstaande secties kort besproken. Deze secties zijn gebaseerd op het oorspronkelijke Wavix rapport [Van Noort (1997)], de golfklimaat site [<http://www.golfklimaat.nl/>], en een op internet gepubliceerde tekst van E. Bouws (KNMI) [<http://www.knmi.nl/~bouws/nl/editgp.htm>]. Zie verder Bouws (1984) en Kuik (1988).

2.1 Het karakteriseren van golven in het tijdsdomein.

Volgens de golftheorie (bevestigd door praktijkwaarnemingen) zijn de golfhoogtes altijd op een bepaalde manier verdeeld. Een klein gedeelte van de golven is zeer hoog, een groot gedeelte is van een gemiddelde hoogte en een klein gedeelte heeft een zeer lage hoogte. Deze verdeling kan doorgaans goed worden beschreven door een zogenaamde Rayleigh verdeling. Zodoende is met behulp van enkele parameters inzicht te geven in de verdeling van golfhoogtes. De belangrijkste parameters in het tijdsdomein staan beschreven in Tabel 1. Niet alle parameters worden opgeslagen in Donar en gevalideerd. In de kolom 'validatie' in Tabel 1 is aangegeven welke parameters in Donar worden opgeslagen en dus gevalideerd moeten worden.

Tabel 1: *Definities van golfparameters in het tijdsdomein*

Categorie	Symbol	Beschrijving	Validatie
Hoogte	Hgem	Gemiddelde golfhoogte	Nee
	H1/3	Significante golfhoogte; Gemiddelde van het hoogste 3° deel van de golven.	Ja
	H1/50	Gemiddelde van het hoogste 50° deel van de golven.	Nee
	Hmax	Hoogste waargenomen golf	Nee
Periode	Tgem	Gemiddelde golf periode	Nee
	T1/3	Gemiddelde van het langste 3° deel van de golfperiodes.	Nee
	TH1/3	Gemiddelde golfperiode van het hoogste 3° deel van de golven.	Ja
Richting	n.v.t.	Er zijn geen relevante parameters voor golfrichting in het tijdsdomein	Nee

De parameters in Tabel 1 hebben betrekking op een analyseperiode van 20 minuten. Deze periodeduur is het gevolg van een trade-off tussen enerzijds het minimaliseren van steekproef effecten en anderzijds het minimaliseren van het effect van wijzigingen in de golfparameters tijdens de waarnemingsperiode. De golfparameters komen in de vorm van een tijdreeks beschikbaar, waarbij eens per uur over een 20-minuten periode wordt gerapporteerd. Waarnemingen over de overige 40 minuten worden niet gebruikt. Van deze parameters worden alleen de significante golfhoogte H1/3, en de golfperiode TH1/3, gevalideerd en opgeslagen in Donar.

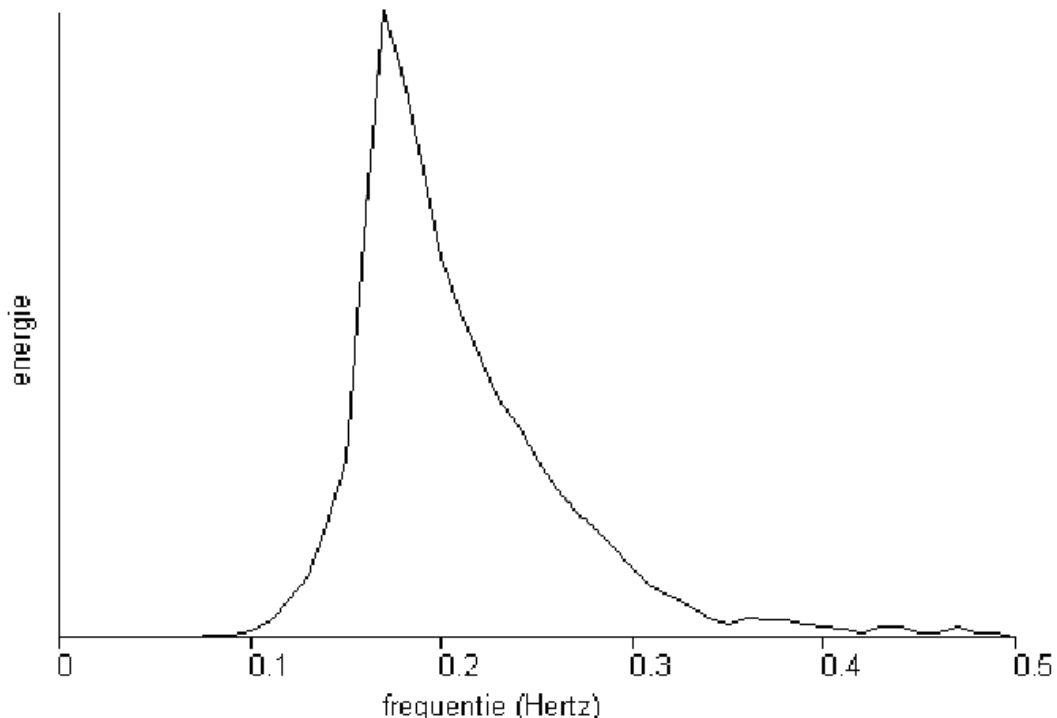
2.2 Het karakteriseren van golven in het frequentiedomein.

In het frequentiedomein is het continue signaal van de oppervlakte hoogte via de zogenaamde Fourier transformatie geconverteerd naar een frequentiespectrum. Dit spectrum geeft per frequentieklasse aan hoeveel energie door golven binnen deze klasse is vertegenwoordigd.

Frequentie is de inverse van golfperiode. De frequenties van de wind geïnduceerde golven waarop de golfenvalidatie zich richt bevinden zich tussen de 0.03 and 0.50 Hertz (golfperiode tussen de 2 en 30 seconden). De golfenergie is proportioneel aan de gekwadrateerde amplitude (=halve golfhoogte) van de golf.

Figuur 1 geeft een voorbeeld van een frequentiespectrum. Dit spectrum geeft aan hoe de golfenergie verdeeld is over verschillende frequenties en geeft voor een geoefend oog een goede indruk van de situatie. Wanneer de golfhoogtes toenemen, schuift het spectrum op richting lage frequenties en neemt de energie toe. Bij lage golfhoogtes is het spectrum gecentreerd rondom de hoge frequenties en is de energie lager. Soms kan in een spectrum een superpositie van meerdere spectra worden herkend, zoals een door de wind ter plaatse geïnduceerde golfslag en *deining* (golfslag die elders is opgewekt en zich naar de waarnemingslocatie heeft voortgeplant).

Het complete spectrum geeft een goede, maar enigszins gecompliceerde, beschrijving van het golvenveld dat gedurende een bepaalde periode aanwezig was. Net als voor de tijdsdomeinbeschrijving van golven zijn ook voor het frequentiedomein weer enige samenvattende parameters beschikbaar. De beschrijving in het frequentiespectrum speelt daarbij een belangrijke rol. Bij deze beschrijving wordt per frequentieklasse getotaliseerd hoeveel 'energie' wordt vertegenwoordigd door golven in deze klasse.



Figuur 1: Voorbeeld van een frequentiespectrum voor golfhoogte
(bron: <http://www.golfklimaat.nl/>)

Deze golfparameters in het frequentiedomein zijn bijna allemaal gebaseerd op de *momenten*

van het spectrum, met name het nulde, eerste en tweede moment. Deze momenten zijn als volgt gedefinieerd:

$$m_n = \int_{0.03}^{0.5} s(f) \cdot f^n df \quad (0)$$

met:

$s(f)$: De spectrale energiedichtheid bij frequentie f

n : De orde van het berekende moment, $n=0,1,2,3,\dots$

Vaak wordt gesproken over de energiedichtheid van een spectrum. Strikt genomen is dit echter niet correct, omdat slechts de variantiedichtheid van het hoogte signaal wordt bedoeld.

De belangrijkste parameters in het frequentiedomein zijn beschreven in Tabel 2. Niet al deze parameters worden in Donar opgeslagen (zie de kolom 'validatie'). Van deze grootheden worden alleen de significante golfhoogte $Hm0$, de laagfrequente golfhoogte $HTE3$, de golfperiode $Tm02$ en de gemiddelde hoofdgolfrichting berekend voor 20 minuut periodes en in de vorm van uurlijkse reeksen opgeslagen. Eens per drie uur wordt over een interval van 20 minuten het spectrum van 0.03 Hz tot 0.5 Hz in stappen van 0.01 Hz berekend en opgeslagen. De overige grootheden zijn desgewenst uit dit spectrum te berekenen.

Tabel 2: *Definities van golfparameters in het frequentiedomein*

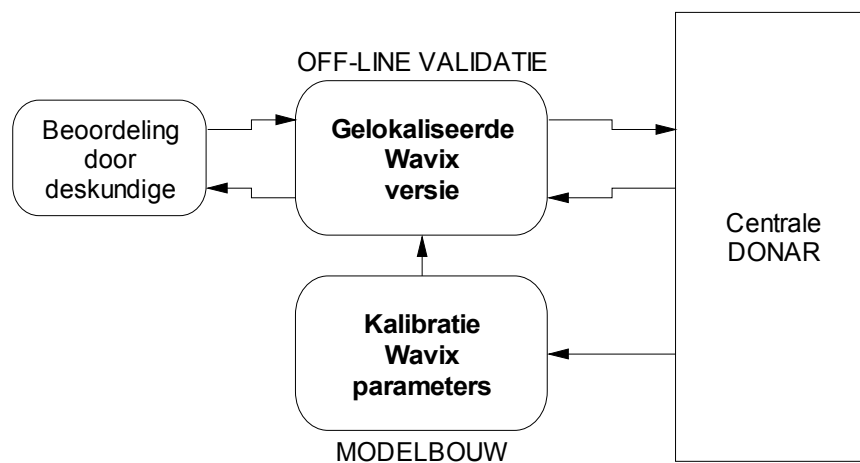
Categorie	Symbol	Beschrijving	Validatie
Hoogte	Hm0	Significante golfhoogte, $4 \cdot m_0^{1/2}$	Ja
	HTE3 (HE10)	Laagfrequente golfhoogte (voor het frequentiegebied 0.03-0.10 Hz) $HTE3 = 4 \sqrt{\int_{0.03}^{0.1} s(f) df}$ <p>Omdat het spectrum wordt opgeslagen in discrete stappen van 0.01 Hz wordt deze uitdrukking in de praktijk benaderd als de volgende sommatie:</p> $HTE3 = 4 \sqrt{\sum_{f \in F} \Delta \cdot s(f)}$ <p>met: $\Delta = 0.01$ $F = \{0.03, 0.04, \dots, 0.10\}$</p>	Ja
Periode	Tm02	Gemiddelde golfperiode Tm02: $Tm_{02} = \sqrt{m_0 / m_2}$	Ja
	Tm01	Significante golfperiode Tm01: $Tm_{01} = m_0 / m_1$	Nee
	Tm-10	Significante golfperiode Tm-10: $Tm_{-10} = m_{-1} / m_0$	Nee
	Tp	$Tp = 1 / Fp$, met Fp de frequentie waar het energiespectrum zijn piek bereikt.	Nee
Richting	Th0	Gemiddelde hoofdgolfrichting. $Th0 = \arctan(b_1/a_1)$, waarbij a_1 en b_1 de laagste fourier-richtingscoëfficiënten zijn van de spectrale richtingsverdeling $D(\theta)$	Ja
	Th1	Laagfrequente golfrichting. Idem, maar beperkt tot de lage frequenties.	Nee
	s0	Gemiddelde richtingsspreiding. $s0 = \{2(1 - M_1)\}^{1/2}$, waarbij $M_1 = (a_1^2 + b_1^2)^{1/2}$; zie hiervoor A.J.Kuik et al., 1988.	Nee

De frequentiedomein parameter Hm0 en de tijdsdomeinparameter H1/3 (beiden aangeduid als significante golfhoogte) lijken op elkaar maar zijn niet exact aan elkaar gelijk. Hm0 is meestal enkele procenten hoger dan H1/3. De golfperiode Tm02 is ongeveer gelijk aan de gemiddelde golfperiode in het tijddomein (Tgem). De golfhoogte HTE3 geeft een indicatie over de aanwezigheid van deining in het golfveld. N.B.: deze parameter wordt ook wel HE10 genoemd.

3 Werkproces

3.1 Inleiding

Binnen de golvenvalidatie kunnen twee afzonderlijke werkprocessen worden onderscheiden. Het eerste werkproces is de off-line validatie waarbij de metingen van de afgelopen periode worden gevalideerd. Dit gebeurt een aantal maal per jaar. Het tweede werkproces is het modelbouwproces waarin een schattingsmodel voor golfparameters wordt opgesteld en gekalibreerd. Dit gebeurt in principe eenmalig voor elke te valideren detector configuratie. Beide werkprocessen worden in de volgende secties kort beschreven.



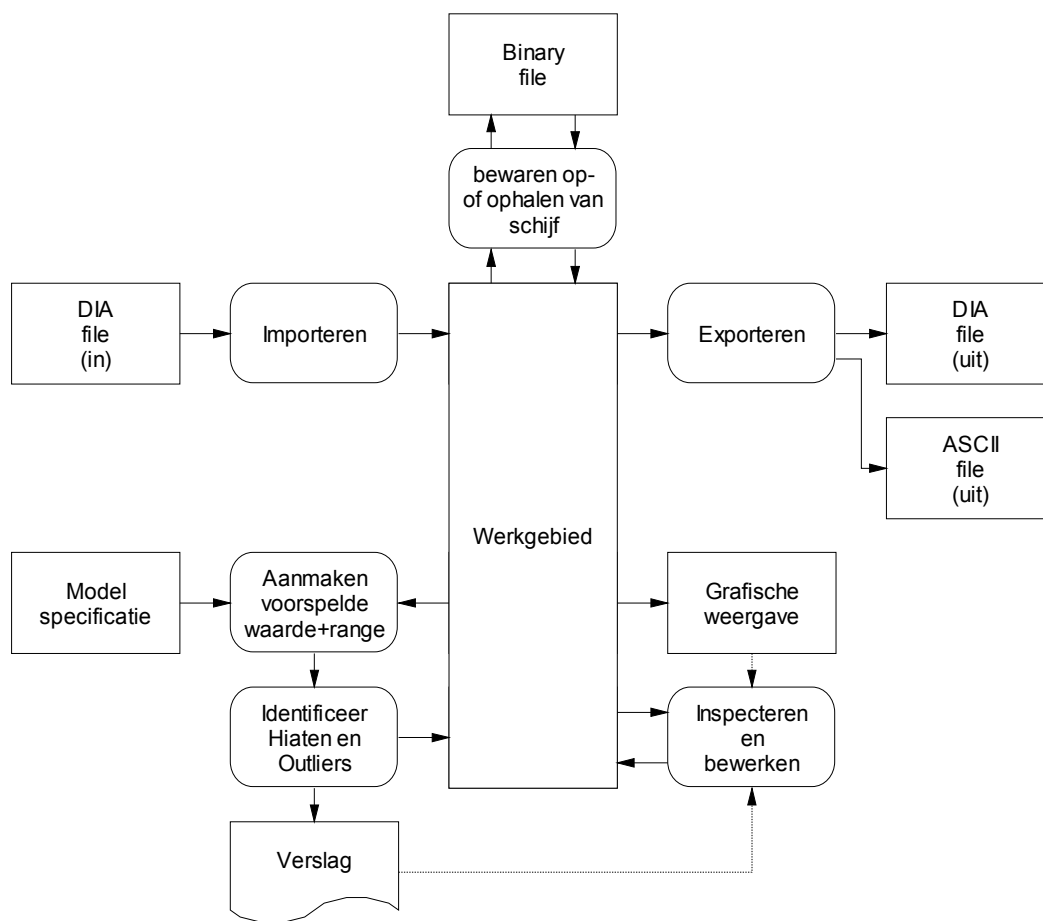
Figuur 2: *Golven applicatie: er is onderscheid tussen een software modules ten behoeve van de off-line validatie en een softwaremodule ten behoeve van de modelbouw*

3.2 Werkproces: Off-line validatie

De operationele validatie kan worden samengevat in het schema van Figuur 3. De volgende stappen worden doorlopen:

- Gegevens worden ingelezen uit een extern DIA bestand;
- Gegevens worden (eventueel) onderworpen aan een visuele controle. Het systeem is in staat om verschillende, nog nader in te vullen, grafische en alfanumerieke weergaven van de golfgegevens te genereren;
- Voor alle grootheden en waarnemingslocaties worden zogenaamde ‘voorspelde waarden’ berekend. Tevens worden bij deze puntschattingen ook standaarddeviaties berekend, aan de hand waarvan betrouwbaarheidsintervallen kunnen worden berekend. Voor het berekenen van de voorspelde waarden wordt gebruik gemaakt van functionele verbanden die eenmalig in de modelbouw fase bepaald zijn (zie volgende sectie);
- Na het instellen van het outlier criterium kan voor iedere waargenomen waarde worden vastgesteld of deze in vergelijking met de voorspelde waarde al dan niet teveel afwijkt. Het outlier criterium wordt doorgaans ingesteld als het maximaal aantal maal de standaardafwijkingen dat een gemeten waarde van de voorspelde waarde mag afwijken;
- De outliers en de geconstateerde ontbrekende waarden (de hiaten) worden weggeschreven in een verslag, en grafisch weergegeven;

- Aan de hand van het verslag of de grafische weergave kan het gedeelte van de reeks dat de hiaten of outliers bevat worden opgezocht. Vervolgens kunnen voor outliers en hiaten eventueel geschatte waarden worden ingevuld. De bijbehorende statuscode komt dan op ‘geïnterpoleerd’ te staan, waardoor deze datapunten later weer van werkelijke meetdata te onderscheiden zijn. Naast het bijgissen, bestaat ook de mogelijkheid om datapunten definitief als hiaat aan te merken, of om outliers alsnog goed te keuren;
- De laatste stap is het exporteren van de gevalideerde data na een extern dataformaat zoals DIA of ASCII tabel, in dat laatste geval gaat mogelijk een gedeelte van de Meta informatie verloren, omdat deze niet in de vorm van een ASCII tabel is te representeren;
- De inhoud van een werkgebied kan tussentijds op harde schijf worden opgeslagen of kan van de harde schijf worden ingelezen.



Figuur 3: *Off-line validatie*

3.3 Werkproces: Modelbouw

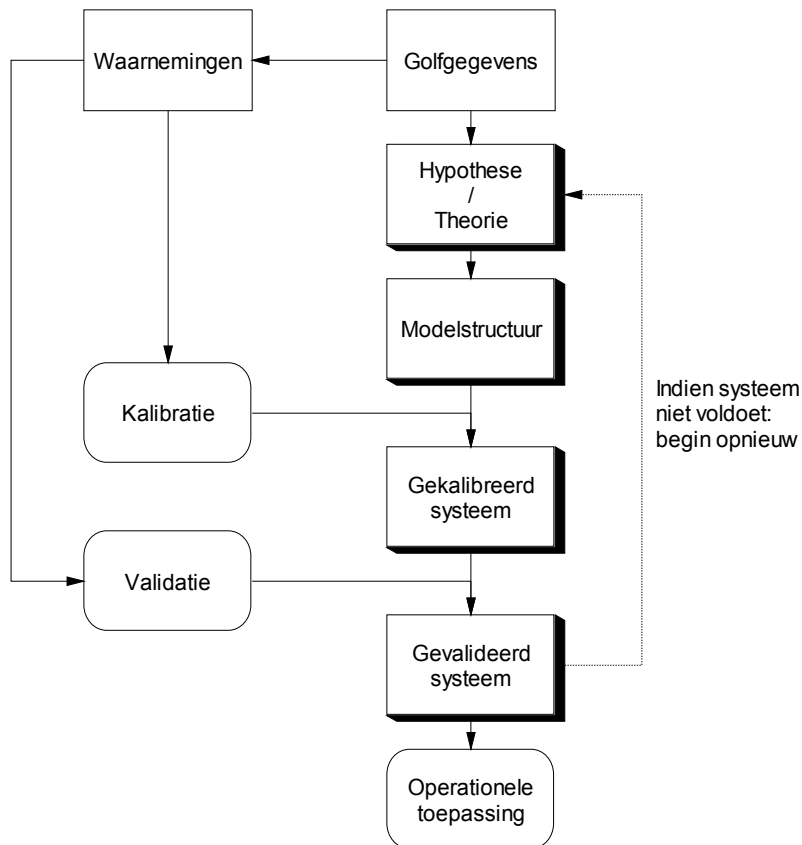
3.3.1 Algemeen

Het hart van de operationele validatie wordt gevormd door de stap waarin de voorspelde waarden en bijbehorende ranges worden aangemaakt. Hiervoor is een model nodig. Het huidige Wavix systeem beschikt hiertoe over een speciaal ontwikkeld model op basis van Neurale netwerken. Dit model is mogelijk te generaliseren tot meer algemene situaties.

Toch zal de gebruiker grote invloed op de vorm van het model kunnen uitoefenen, bijvoorbeeld door aan te geven welk type neuraal netwerk wordt gebruikt, welke afhankelijkheden wel, en welke afhankelijkheden niet mogen worden verondersteld, en dergelijke. Dit proces noemen we de modelbouw. Het aantal vrijheidsgraden waarover de gebruiker moet beslissen is nog nader te bepalen in dit Functioneel Ontwerp.

Het bouwen van een model is doorgaans een iteratief proces, dat in Figuur 4 geschematiseerd is weergegeven¹. Dit begint met het maken van hypothesen over de onderlinge samenhang van de gegevens, zoals de aanname dat de waterdiepte een bepaalde invloed heeft op de golfhoogte, of dat de golfhoogte van nu op een bepaalde manier samenhangt met de golfhoogte van enkele uren terug. Op basis van dit soort aannames kan een modelstructuur worden gevormd, waarin nog een aantal parameters te kalibreren zijn. De volgende stap is dat de modelparameters daadwerkelijk worden gekalibreerd, dat wil zeggen: deze parameters worden zodanig gekozen dat het model optimaal past bij een verzameling waarnemingen die voor de kalibratie wordt gebruikt. Deze set is bij voorkeur van te voren ontdaan van fouten. De laatste stap in het modelbouw proces is de modelvalidatie, waarin op basis van data die nog niet in de kalibratiestap zijn gebruikt een controle op het model wordt toegepast. Vaak is de conclusie naar aanleiding van de validatiestap dat modelverbeteringen nog mogelijk zijn, door rekening te houden met extra factoren die in eerste instantie verwaarloosd waren. Op dat moment kan een volgende iteratie binnen het modelbouwproces worden ingezet, net zo lang totdat nieuwe verbeteringen aan het model niet meer lonend worden geacht.

¹ In tegenstelling tot Figuur 3, verwijst Figuur 4 niet naar bestanden of softwarecomponenten, maar naar proces-stappen.



Figuur 4: *Stappen in het modelbouw proces*

3.3.2 Kalibratie neuraal netwerk

Het uitgangspunt is dat ook in de nieuwe WAVIX versie de kern van het model wordt gevormd door een neuraal netwerk (of leiver gezegd: door een verzameling neurale netwerken). Het specificeren van een neuraal netwerk gebaseerd model bestaat uit de volgende stappen:

- Het selecteren en opschonen van een dataset. Een bruikbare dataset bevat allereerst een gezonde variatie in de stimuli (in dit geval golfhoogte, richting, windsnelheid etc. bij naburige stations). Foutieve waarnemingen dienen zoveel mogelijk te worden weggefilterd. Ten slotte moeten situaties die afwijken ten gevolge van niet waargenomen omstandigheden buiten de dataset gehouden worden. Een (buitensporig) voorbeeld daarvan zou een periode met zware ijsgang zijn;
- Het onderverdelen van de opgeschoonde dataset in een kalibratie- en een controleset. Nadat het model gekalibreerd is, vindt aan de hand van een aparte dataset een controle plaats. Deze controle is nodig om de voorspellende waarde van het model te toetsen. Vooral als het neuraal netwerk te groot en de trainingsset te klein wordt gekozen kan het model overgeparameteriseerd raken. In dit geval zal het model na een voldoende lange inleerperiode een goede score laten zien op de trainingsset, maar wordt deze score niet bereikt op grond van fysische verbanden die door het model worden onderkend, maar eerder op grond van het herkennen en reproduceren van specifieke situaties. Een controle met behulp van data die niet in het kalibratieproces werd gebruikt kan deze tekortkoming aan het licht brengen;

- Het kiezen van een basisvorm voor het neurale netwerk. Net zoals er vele schattings- en optimalisatie algoritmes bestaan zijn er ook een aantal verschijningsvormen van neurale netwerken. De meest bekende vorm is het Back Propagation Network dat ook in de huidige WAVIX versie wordt toegepast. Zelfs als echter op voorhand voor dit type netwerk wordt gekozen moeten nog een aantal configuratie beslissingen worden genomen, zoals het aantal layers en het aantal knopen per layer;
- Het definiëren afhankelijkheden in de ruimte en de tijd. De gebruiker kan de kans op ongewenste uitkomsten verminderen en de voorspellende waarde van het model verhogen door op voorhand vast te leggen welke afhankelijkheden mogelijk wel en welke afhankelijkheden in ieder geval niet bestaan. Deze afhankelijkheden betreffen de detectorlocaties, de gegevenssoort en de tijdstippen.
- Het inleren van het neurale netwerk;
- Het controleren van het gekalibreerde model. Hiervoor wordt gebruik gemaakt van de controlezet. Aan de hand van grafieken en overzichten kan worden geïnspecteerd in hoeverre het gekalibreerde model past bij de controlezet;
- Het vullen van de modelcomponent waarmee de nauwkeurigheid van de geproduceerde puntschattingen wordt voorspeld.

4 Beschrijving software modules voor operationele validatie

4.1 Inleiding

De operationele validatie werd al samengevat in het schema van Figuur 3. Dit hoofdstuk beschrijft de software modules die de verschillende bewerkingen ondersteunen.

4.2 Visuele controle van invoergegevens

4.2.1 Overzichtsscherm

Wanneer de applicatie opstart verschijnt als eerste het overzichtsscherm. Zodra data geladen zijn bevat dit scherm een samenvatting van de detectoren die in het werkgebied aanwezig zijn. Per detector worden een aantal kentallen getoond zoals het percentage hiaten per datacategorie. Ook kan vanuit het overzichtsscherm worden aangegeven voor welke detectoren een detailscherm moet worden getoond en welke variabelen in deze detailschermen moeten worden weergegeven.

4.2.2 Weergave van tijdreeksen

De volgende tijdreeksen van grootheden zijn in de golven applicatie te inspecteren:

Tabel 3: *Gegevenssoorten en bijbehorende bewerkingen*

Variabele	Code	Bewerkingen		
		Inspecteren	Afkeuren	Bijgissen
Significante golfhoogte	Hm0	Ja	Ja	Ja
Significante golfhoogte	H1/3	Ja	Ja	Ja
Laagfrequente golfhoogte	HE10	Ja	Ja	Ja
Gemiddelde golfperiode	Tm02	Ja	Ja	Ja
Significante golfperiode	TH1/3	Ja	Ja	Ja
Gemiddelde hoofdgolfrichting	Th0	Ja	Ja	Ja
Laagfrequente golfrichting	Th1	Ja	Ja	Nee
Gemiddelde richtingsspreiding	s0	Ja	Ja	Nee
Windsnelheid	-	Ja	Nee	Nee
Windrichting	-	Ja	Nee	Nee
Waterstand	-	Ja	Nee	Nee
Golfspectrum	-	Eventueel	Nee	Nee

Deze grootheden zijn beschikbaar per locatie, bovendien zijn voor de golfgegevens per locatie gegevens beschikbaar uit zowel een hoofdsensor als een nevensensor. Deze reeksen dienen te kunnen worden gepresenteerd als grafieken in het zogenaamde inspectiescherm. Het inspectiescherm voor de diverse reeksen zal tevens gebruikt worden als interactief bewerkingsscherm (zie sectie 4.7).

Het weergeven van het golfspectrum is optioneel en dient in de (aanbesteding van) implementatie apart behandeld te worden.

Om het mogelijk te maken om de gegevens van alle detectoren in één scherm weer te geven dient voor een compacte presentatie gekozen te worden, waarbij meerdere lijnen in één grafiek worden weergegeven. Hierbij zijn de volgende manieren van presentatie te overwegen:

- Geef alle reeksen die bij een detector horen in één grafiek weer (dit is de situatie in het huidige WAVIX). Omdat niet alle reeksen dezelfde eenheid hebben (als eenheden komen voor golfhoogte, golfperiode en golfrichting, waterstand, windsterkte en windrichting) is het niet mogelijk de y-as eenduidig te definiëren;
- Splits de weergave per detector in meerdere grafieken op basis van de weer te geven eenheid. In verband met de beschikbare ruimte op het beeldscherm is deze optie niet toepasbaar bij het weergeven van alle stations tegelijk;
- Geef reeksen van dezelfde soort in één grafiek weer. Dus één grafiek HMO reeksen, één grafiek met de H1/3 reeksen enzovoort. In dit geval kan worden volstaan met acht grafieken, ongeacht het aantal weer te geven detectoren.

De eerste weergave optie voldoet momenteel goed in de praktijk. Mogelijkheid 2 en 3 dienen daarom hooguit als optioneel te worden beschouwd.

De weergave van de grafieken dient door de gebruiker op de volgende wijze beïnvloed te kunnen worden:

- De weer te geven detectoren dienen te kunnen worden geselecteerd uit de lijst van detectoren die in het werkgebied aanwezig zijn;
- De weer te geven grootheden moeten vrij kunnen worden gekozen;
- Indien meerdere manieren van presentatie worden ondersteund: selecteer de presentatie wijze;
- Selecteer hoe de diverse grafieken gerangschikt worden. Wanneer meerdere grafieken gelijktijdig worden weergegeven dient te worden aangegeven hoeveel kolommen worden gebruikt. 8 panels kunnen bijvoorbeeld alle boven elkaar worden weergegeven (1 kolom), of in twee kolommen van 4 panels, of in 3 kolommen, etc;
- Zoom in op het weer te geven tijdsinterval. De gebruiker kan met de muis en andere navigatiehulpmiddelen (bijvoorbeeld een kalender) een bepaald gedeelte van de tijd-as selecteren. Het dient mogelijk te zijn om de tijd-assen van de verschillende grafieken automatisch te laten synchroniseren.

4.2.3 Ruimtelijke weergave

Op dit moment wordt geen gebruik gemaakt van ruimtelijke visualisaties van golven velden en eventueel andere gegevens zoals waterstand en wind. Toch kunnen dergelijke visualisaties een nuttige rol spelen in de controle van de gegevens en het specificeren van modellen. Het gaat dan om een beknopte grafische weergave zonder interactieve mogelijkheden. Omdat het belang van een dergelijke visualisatie ondergeschikt is aan de overige functies gaat Functioneel Ontwerp er niet verder op in.

Er wordt daarom aanbevolen om het al dan niet opnemen van een beknopte ruimtelijke weergave in het verdere implementatietraject als optioneel te beschouwen.

4.3 Completeren reeksen

Indien in de nieuwe versie de huidige berekeningsmethode ongewijzigd wordt overgenomen geldt als eis voor de invoer van deze berekening dat geen hiaten mogen optreden. Hiaten die aanwezig zijn in de invoerdata dienen dan te worden vervangen door de best mogelijke benadering. Data die op deze manier zijn bijgeschat dienen te worden voorzien van een code ‘extrapolatie’.

4.3.1 Automatische completering

Deze completering kan grotendeels automatisch plaatsvinden door gebruik te maken van een *substitutie tabel*. Deze tabel bevat voor ieder combinatie van detector en gegevenstype een lijst van substitutiemogelijkheden, in volgorde van voorkeur. De combinatie detector-gegevenstype wordt in deze tabel geïdentificeerd aan de hand van de velden locatie code, sensortype (hoofd of neven) en gegevens type. Het attribuut sensortype is alleen relevant voor golfgegevens. Voor windsnelheid en waterstand (die voor WAVIX alleen als invoer gelden) wordt dit onderscheid niet gemaakt.

De in WAVIX gebruikte combinatie locatiecode-sensortype-gegevenssoort correspondeert eenduidig met een combinatie locatiecode (DONAR W3H veld: sLoccod), veldapparaattypecode (DONAR W3H veld: sVatcod) en gegevenssoort (DONAR W3H veld: sParcod) in DONAR. DONAR kent het begrip hoofd- of nevensensor dus niet. De invoer- en uitvoer modules (zie hoofdstuk 6) zorgen voor een correcte conversie.

Opmerking: momenteel wordt deze conversie uitgevoerd door de software module Wavix2dia (zie RIKZ, 2002)

4.3.2 Completeren met de Wavix98 methode

In plaats van het substitueren van ontbrekende waarden kunnen ze ook worden bijgeschat. In Wavix98 is hiervoor een methode ontworpen. In deze methode worden voor een beperkt aantal windrichting en windsnelheidsklassen en voor elke te valideren parameter de gemiddelde verhoudingen bepaald, en worden ook bijbehorende spreidingen afgeleid. Bij het completeren wordt een geschatte waarde voor de ontbrekende parameter bepaald als het gewogen gemiddelde van alle aparte voorspellingen die op basis van de berekende verhoudingsgetallen kunnen worden bepaald. De weegfactoren zijn de inverse van de eerder bepaalde spreidingen:

$$\hat{y}_{i|V} = \frac{1}{D} \sum_{v \in V} \frac{\hat{y}_{i|v}}{\sigma_{i|v}^2} \quad D = \sum_{v \in V} \frac{1}{\sigma_{i|v}^2}$$

met:

- V : De verzameling van waargenomen locaties
- $\hat{y}_{i|V}$: Voorspelde waarde voor locatie i op basis van alle waargenomen locaties (0)
- $\hat{y}_{i|v}$: Voorspelde waarde voor locatie i op basis van locatie v
- $\sigma_{i|v}^2$: Spreiding van y_i rondom $\hat{y}_{i|v}$

4.3.3 Substitutietool

De substitutietabel kan via een apart tool, het zogenaamde substitutietool, worden aangemaakt en kan voor hergebruik worden bewaard op schijf. Dit tool zorgt ervoor dat de substitutietabel op correcte wijze wordt aangemaakt zodat het aantal controles dat achteraf moet worden uitgevoerd beperkt blijft. Het substitutietool dient ook voor de inspectie van een eerder aangemaakte substitutietabel. Het substitutietool is voorzien van een Grafische User Interface die als hoofdelementen drie lijsten heeft (zie bijvoorbeeld Figuur 5). In de linkerlijst worden alle detector-sensortype-gegevenstype combinaties weergegeven die in het werkgebied (of de te bewerken substitutietabel) vertegenwoordigd zijn. Door uit deze lijst een regel te selecteren wordt het te bewerken item geselecteerd. In de middelste tabel worden vervolgens alle kandidaten gepresenteerd die aan het te bewerken item kunnen worden gekoppeld. In de meeste rechtse tabel wordt de actuele inhoud van de substitutietabel weergegeven voor het te bewerken item. De Grafische User Interface bevat verder knoppen voor het invoegen, verplaatsen of verwijderen van een regel in de substitutietabel. De substitutietabel moet kunnen worden uitgevoerd als een alfanumerieke verslag.

<p>Lijst: <i>Te bewerken detector-sensortype-gegevenstype combinaties.</i> Alle combinaties die in het werkgebied aanwezig zijn worden weergegeven. Er is steeds hooguit uit één item geselecteerd. Dit is het te bewerken item.</p>	<p>Lijst: <i>Kandidaten voor de substitutie tabel.</i> Alleen kandidaten die voor wat betreft gegevens type in aanmerking komen worden afgebeeld.</p>	<p>Tabel: <i>Substitutietabel.</i> In deze tabel wordt de huidige substitutietabel weergegeven voor de te bewerken detector-sensortype-gegevenstype combinatie.</p>
--	---	---

Figuur 5: *Mogelijke schermindeling van de Grafische User Interface voor het bewerken/inspecteren van de substitutietabel*

4.3.4 Handmatige completering

Naast het automatisch completeren moet het mogelijk zijn om voor of na deze completering handmatig wijzigingen aan te brengen in de invoerreeksen. De procedure die hiervoor wordt voorgesteld bestaat uit de volgende stappen:

- Selecteer een aantal perioden uit een bepaalde tijdreeks (zie sectie 4.7 voor meer details);
- Vul de tijdreeks ‘voorspelde waarde’ met de nieuwe tijdreeks (zie sectie 4.4.3);
- Kies de optie ‘Geselecteerde waarden vervangen door voorspelde waarden’.

4.4 Berekening voorspelde waarden

4.4.1 Inleiding

Voor alle tijdreeksen die binnen WAVIX gevalideerd worden kunnen tijdreeksen van *voorspelde waarden* worden berekend. De voorspelde waarden worden gebruikt om een vergelijking met de gemeten waarden uit te voeren (de identificatie van outliers) en desgewenst om gemeten waarden door voorspelde waarden te vervangen.

4.4.2 Bereken voorspelde waarden met model

Op basis van een eerder opgesteld model (zie hoofdstuk 5) en de detector data uit het werkgebied worden de voorspelde waarden en de bijbehorende betrouwbaarheidsintervallen berekend.

In de huidige WAVIX versie is het met het oog op rekestijd mogelijk een dergelijke berekening uit te voeren voor een gedeelte van de periode waarvoor data aanwezig zijn. In de toekomstige versie zou deze mogelijkheid mogen vervallen, mits de berekening voor de volledige dataset in beperkte tijd (bijvoorbeeld 1 minuut) kan worden voltooid.

De berekening wordt gestart met behulp van een knop of menu. Voordat de berekening start wordt gecontroleerd of aan alle nodige voorwaarden is voldaan. Deze voorwaarden hangen af van de toe te passen berekeningsmethode.

Tijdens de berekening wordt een progress-bar getoond. Nadat de berekening voltooid is wordt een verslag weggeschreven.

4.4.3 Vul de voorspelde waarden met een kopie van een andere tijdreeks

In plaats van waarden die met een model zijn berekend, kunnen ook handmatig waarden worden aangewezen waarmee de verzameling van voorspelde waarden wordt gevuld. In dit geval is er sprake van substitutie.

4.5 Identificeren outliers

Tegelijk met het berekenen van voorspelde waarden wordt ook een betrouwbaarheidsmarge berekend. Deze marge is gedefinieerd in de vorm van een standaarddeviatie (de wortel uit de gemiddelde kwadratische afwijking tussen voorspelde en waargenomen waarde).

Er is sprake van een outlier wanneer de werkelijke kwadratische afwijking van voorspelde en waargenomen waarde significant groter is dan de bovengenoemde variantie, of equivalent: indien de absolute afwijking groter is dan de wortel van de variantie (de standaardafwijking). Niet elke outlier wordt veroorzaakt door een fout, indien een correcte waarneming toch een outlier veroorzaakt, spreken we van een *vals alarm*.

De kans op een vals alarm hangt samen met het aantal maal dat de standaardafwijking moet worden overschreden voordat een outlier wordt getriggerd. Als de afwijkingen normaal verdeeld zijn, dan is de kans dat een absolute afwijking van 3.1σ of groter optreedt kleiner dan 1 op 500. Met andere woorden: als we het outliercriterium instellen op een maximaal acceptabele afwijking van 3.1σ (op basis van een tweezijdig begrensde interval) dan genereren we minder dan 1 vals alarm op de 500 datapunten. De onderstaande tabel bevat de vals-alarm kansen voor andere instellingen.

Tabel 4: *Drempelwaarden met bijbehorende vals alarm kansen*

Ingestelde drempelwaarde voor de acceptabele afwijking (maal σ)	kans op een onterechte kwalificatie als outlier (%)
D=1.3	20
D=1.6	10
D=2.0	5
D=2.3	2
D=2.6	1
D=3.1	0.2
D=3.3	0.1
D=3.9	0.01
D=4.4	0.001

Er moet een goede trade-off gevonden worden tussen het optreden van valse alarmen en het detecteren van fouten. Wanneer een outlier pas wordt getriggerd bij een overschrijding van 10 maal de standaardafwijking, dan kunnen we er vrijwel zeker van zijn dat elke outlier een fout betreft, maar zullen er tevens vele foutieve waarnemingen zijn die geen outlier triggeren. Wanneer een (te) lage waarde voor D kiezen zullen er echter vele valse alarmen worden gegenereerd die allemaal visueel geïnspecteerd moeten worden. Hierdoor neemt het valideren van de data onnodig veel tijd.

Een veelgebruikte waarde in de huidige WAVIX praktijk is D=4.

4.6 Genereren verslag met outliers

In de huidige WAVIX versie fungeert een verslag met outlierperiodes als hulpmiddel bij het inspecteren van de outliers. Voor terugwaartse compatibiliteit dient dit verslag ook door de nieuwe versie geproduceerd te kunnen worden. In dit verslag moeten minimaal de volgende gegevens terug zijn te vinden:

- de locatie, datum en tijd waarop de outlier geconstateerd is;
- een regel met de gemeten waarden voor de golfparameters voor dit tijdstip;
- een regel met de voorspelde waarden voor de golfparameters voor dit tijdstip;
- een regel met het verschil tussen gemeten en voorspelde waarden voor dit tijdstip, uitgedrukt als het aantal maal de standaardafwijking.

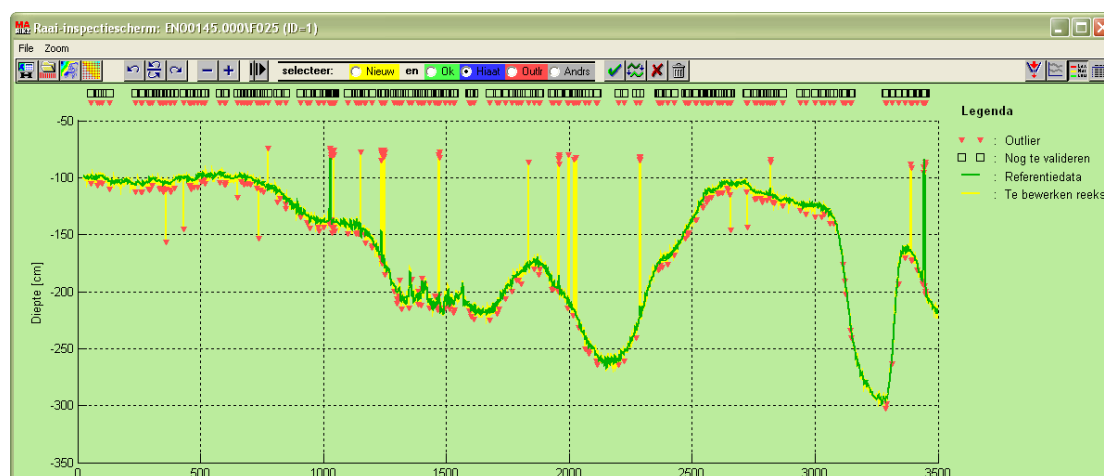
Naast een verslag dient de nieuwe versie ook een aantal hulpmiddelen te bevatten om geconstateerde outliers snel in hun context aan de gebruiker zichtbaar te maken. Ten eerste bevat het overzichtsscherm (zie sectie 4.2.1) een lijst met de in het werkgebied vertegenwoordigde locaties waarin per gegevenssoort het aantal outliers is weergegeven. Uit deze lijst kan in een oogopslag worden afgelezen voor welke locaties de grootste problemen optreden. Ten tweede zijn in de grafische weergave van de afzonderlijke locaties met symbolen in de bovenmarge van de grafiek de tijdstippen gemarkeerd waarbij voor ten minste één van de reeksen een outlier optreedt. Bovendien bevatten deze grafische weergaveschermen navigatieknoppen waarmee naar vorige en volgende outliers kan worden gesprongen (zie sectie 4.7).

4.7 Interactief aanpassen: valideren

4.7.1 Inleiding

Nadat de outliers zijn geïdentificeerd, is het aan te gebruiker om te beslissen welke outliers (en mogelijk ook andere data) moeten worden verworpen of bijgeschat en welke data goedgekeurd worden. Dit gebeurt in een bewerkingsscherm waarbij de diverse tijdreeksen voor één locatie in een grafiek zijn weergegeven. Het bewerkingsscherm lijkt op de in sectie 4.2.2 genoemde inspectie schermen. Maar omdat het niet nodig is om meerdere detectoren tegelijk te bewerken mag het bewerkingsscherm grotere afmetingen hebben, zodat ruimte is voor meer details zoals een toolbar, legenda, annotatie bij de grafiek en marge waarin met behulp van iconen de validatiestatus kan worden aangegeven. Ook is het mogelijk om de verschillende grootheden zoals golfhoogte, golfrichting, windsterkte en eventueel waterstand in aparte maar voor wat betreft de tijd-as gesynchroniseerde assen stelsels recht boven elkaar af te beelden, op een soortgelijke manier als nu gebeurt voor de loodlijn en de diepte in het raai-inspectiescherm van MARIA (Morfologie Applicatie voor Ruimtelijke Interpolatie en Analyse).

Figuur 6 toont het bewerkingsscherm voor raagegevens zoals dat momenteel in MARIA te bekijken is. Hoewel raagegevens geen tijdreeks maar een ruimtelijke reeks vormen is MARIA toch als concreet voorbeeld te gebruiken hoe de bewerking zou kunnen plaatsvinden. MARIA beschikt ook over een alfanumerieke wijze van presenteren. En dergelijke presentatie is voor de nieuwe WAVIX versie nog optioneel. De volgende sectie beschrijft aan hoe de bedieningswijze zou kunnen verlopen.



Figuur 6: Voorbeeld: bewerkingsscherm zoals dit in MARIA (Morfologie Applicatie voor Ruimtelijke Interpolatie en Analyse) gebruikt wordt

4.8 Inhoud van het inspectiescherm voor golfparameters

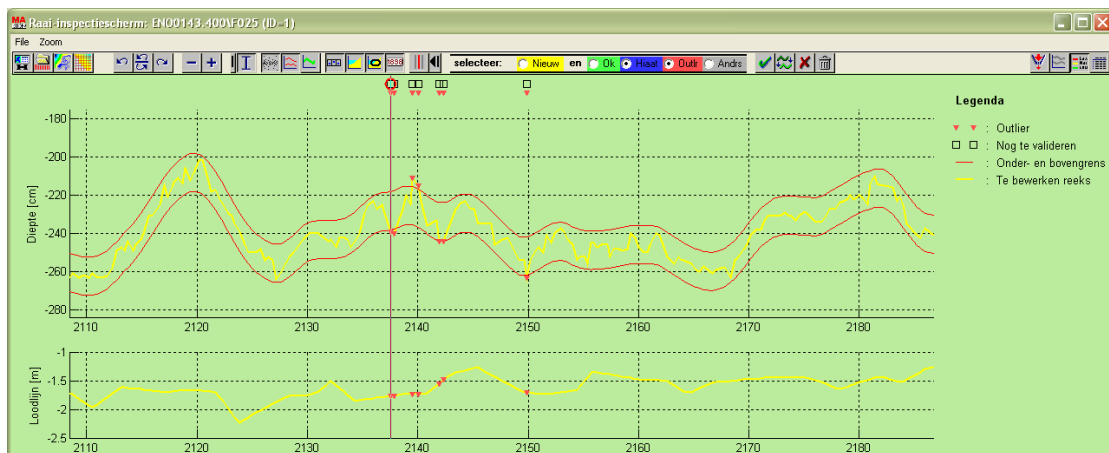
Bij golfparameters is sprake van meerdere parameters per locatie die in hun samenhang beoordeeld moeten worden. Het is daarom nodig om al deze parameters gelijktijdig in grafieken uit te zetten. Sommige van deze parameters zijn in het tijdsdomein gedefinieerd, terwijl andere in het frequentiedomein zijn gedefinieerd. Omdat parameters voor golfhoogte, golfperiode en golfrichting ook verschillende eenheden hebben, is het ten behoeve van de overzichtelijkheid gewenst om de diverse parameters boven elkaar weer te geven in drie

grafieken waarvan de x-as synchroon loopt. Figuur 7 geeft een voorbeeld van dit principe dat afkomstig is uit MARIA. Hier zijn de diepte en de loodlijn boven elkaar geprojecteerd, maar hetzelfde kan worden toegepast op golfhoogte, golfperiode en golfrichting.

Het inspectiescherm voor golfparameters bevat de volgende elementen:

- Een grafiek met golfhoogtes;
- Een grafiek met golfperiodes;
- Een grafiek met golfrichtingen;
- Een grafiek met wind en waterstandsgegevens (niet te valideren);
- Een statusbar (een balk met statuscodes per periode);
- Curves met gemeten waarden (alle parameter soorten);
- Curves met voorspelde waarden (alle parameter waarvoor voorspelling aanwezig is);
- Curves met onder- en bovengrens (alle parameter waarvoor voorspelling aanwezig is);
- Symbolen die situaties periodes markeren, zoals hiaten, outliers en bijgeschatte data.
- Een legenda
- Een toolbar voor het aan- en uitzetten van verschillende beeldelementen

Meestal zullen niet alle elementen tegelijk geactiveerd zijn, omdat de weergave dan onoverzichtelijk kan worden. Via een keuzemenu, en eventueel shortcut toetsen in de toolbar, kan de gebruiker aangeven welke beeldelementen zichtbaar dienen te zijn.



Figuur 7: Tweede voorbeeld: tweede parameter (loodlijn) wordt gevisualiseerd in 'gesynchroniseerde' grafiek

4.8.1 Mechanisme van valideren

Te valideren gegevenssoort

Er zijn een aantal soorten parameters die gevalideerd kunnen worden (zie Tabel 3 op blz 17). In principe worden de reeksen voor alle parameters gelijktijdig gevalideerd. Dit betekent bijvoorbeeld dat als de H_{m0} parameter wordt afgekeurd, ook de parameters voor de andere parameters worden afgekeurd. Voor het gemak van de gebruiker dient het daarom mogelijk te zijn om bepaalde acties, zoals het op hiaat zetten voor een bepaald interval of het vervangen door voorspelde waarde in één keer op alle parameters toe te passen. Ook de classificatie van bepaalde periodes als outlierperiode of hiaatperiode moet mogelijk zijn op basis van alle parametersoorten tegelijk. Dit wil zeggen: een periode wordt als hiaat aangemerkt als tenminste één parameter ontbreekt voor deze periode.

Datapunt categorieën

Het valideren begint met het berekenen van een referentiereeks, ook wel voorspelde waarde genoemd. Door de tijdreeks met de referentiereeks te vergelijken kunnen alle datapunten worden onderverdeeld in de volgende, elkaar onderling uitsluitende, categorieën:

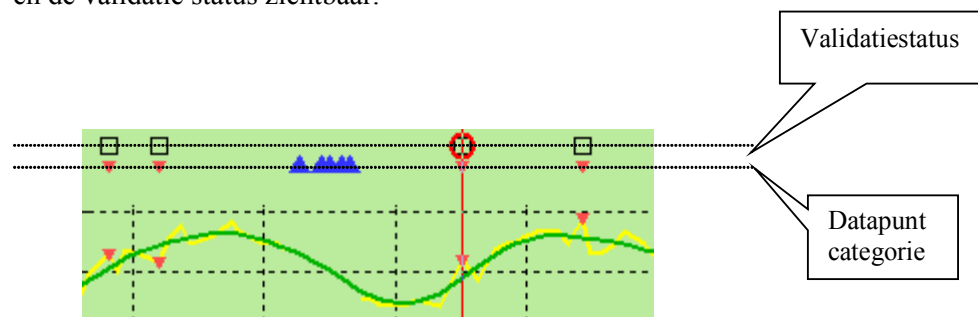
- Akkoord bevonden datapunten, kort aangeduid als *ok* waarden
- Ontbrekende datapunten, zogenaamde *hiaten*. Een periode wordt als hiaat aangemerkt indien tenminste voor één parametersoort de gegevens ontbreken;
- Datapunten waarvan de afwijking t.o.v. de referentiereeks de opgegeven drempelwaarde overschrijdt, zogenaamde *outliers*. Een periode wordt als outlier aangemerkt indien tenminste voor één parametersoort een outlier optreedt. Bij het gelijktijdig optreden van een outlier en een hiaat prevaleert het predikaat hiaat;
- Datapunten waarvoor in de referentiereeks geen corresponderende waarden beschikbaar zijn. Deze worden ingedeeld in de categorie *anders*. Deze situatie treedt op vlak nadat de te valideren in het werkgebied zijn ingelezen. Er zijn dan nog geen voorspelde waarden aangemaakt;

Datapunten die nog niet akkoord bevonden zijn, moeten door de gebruiker worden beoordeeld voordat ze als gevalideerd mogen worden beschouwd.

Los van de bovenstaande indeling wordt een reeks daarom ook ingedeeld in *ongevalideerde* en *gevalideerde* datapunten:

- Nieuw (ongevalideerd): datapunten die voor een beoordeling door de gebruiker in aanmerking komen, maar nog niet beoordeeld (gevalideerd) zijn;
- Niet nieuw (gevalideerd): datapunten die door de gebruiker beoordeeld zijn of waarvoor geen gebruikersbeoordeling nodig is.

De datapunt categorie en de validatie status zijn met behulp van een icon zichtbaar gemaakt in een balk boven de grafische weergave van de reeks (zie Figuur 8). Ook in het alfanumerieke overzicht zijn (indien de kolommen door de gebruiker geactiveerd zijn) de datapunt categorie en de validatie status zichtbaar.



Figuur 8: In een balk boven de grafiek is de validatiestatus (bovenste regel) en de datapunt categorie (onderste regel) via een symbool aangegeven. In de legenda worden alle gebruikte symbolen verklaard

Selectie van datapunten in grafiek met behulp van selectiecriteria

De datapunt categorie en de validatie status dienen als een hulpmiddel om datapunten te selecteren en een systematische validatie mogelijk te maken. Het selecteren van de datapunten verloopt in twee stappen:

Stap 1: het instellen van een selectie criterium.

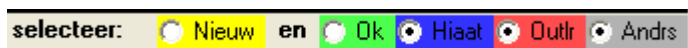
In de toolbar van het raai-inspectiescherm wordt aangevinkt welke datapunten te selecteren zijn. Tabel 5 geeft een overzicht van de selectie mogelijkheden

Stap 2: het selecteren van datapunten

Vervolgens kunnen met de muis datapunten worden geselecteerd. Tabel 6 geeft hier voorbeelden van. De eenvoudigste manier van selecteren is het simpelweg klikken in de figuur. Het eerste punt rechts van het aangeklikte punt dat aan de ingestelde selectiecriteria voldoet wordt dan geselecteerd. *Indien dit punt buiten het zichtbare vlak valt wordt het zichtbare gedeelte van de grafiek automatisch naar links gescrolld*. Op deze manier kan door het herhaaldelijk klikken ter rechterzijde van het laatst zichtbare geselecteerde punt door de hele grafiek worden heengestapt.

Tip: Bij het valideren van grote hoeveelheden gegevens is het met het oog op het aantal uit te voeren handelingen aan te raden eerst alle datapunten te bekijken door steeds ter rechterzijde van de grafiek te klikken en daarna alle punten in één keer goed of af te keuren.

Er kunnen ook meerdere punten tegelijk worden geselecteerd. Dit kan door de shift knop ingedrukt te houden terwijl u met de muis door middel van slepen het gewenste interval aanwijst. Hetzelfde effect wordt bereikt indien eerst met een enkele muisklik het linker punt van het interval wordt bepaald en daarna de rode cirkel die dit beginpunt markeert naar links of rechts wordt geslept.



Figuur 9: Instellen selectie criteria voor het selecteren van datapunten

Tabel 5: Instel mogelijkheden voor de selectiecriteria en hun uitwerking

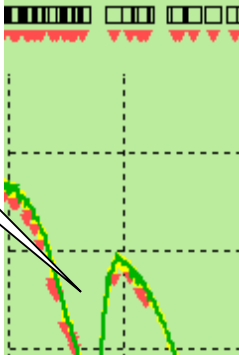
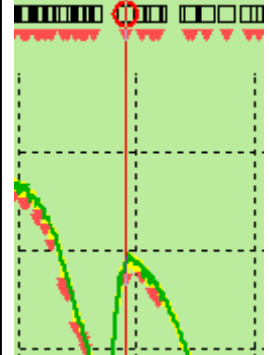
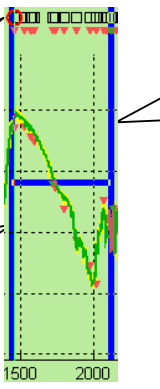

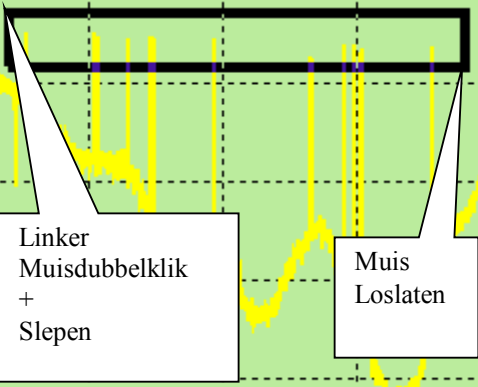
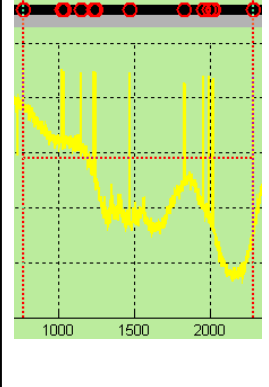
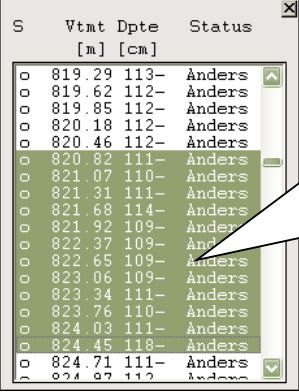
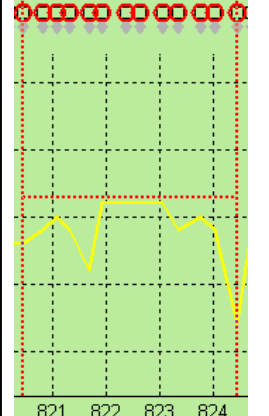
<input checked="" type="radio"/> Nieuw	Selecteer alleen <i>ongevalideerde</i> datapunten
<input type="radio"/> Nieuw	Selecteer zowel <i>ongevalideerde</i> als <i>gevalideerde</i> datapunten
<input checked="" type="radio"/> Ok	Maak datapunten van de categorie <i>ok</i> selecteerbaar
<input type="radio"/> Ok	Voorkom dat datapunten van de categorie <i>ok</i> worden geselecteerd
<input checked="" type="radio"/> Hiaat	Maak datapunten van de categorie <i>hiaat</i> selecteerbaar
<input type="radio"/> Hiaat	Voorkom dat datapunten van de categorie <i>hiaat</i> worden geselecteerd
<input checked="" type="radio"/> Outlr	Maak datapunten van de categorie <i>outlier</i> selecteerbaar
<input type="radio"/> Outlr	Voorkom dat datapunten van de categorie <i>outlier</i> worden geselecteerd
<input checked="" type="radio"/> Andrs	Maak datapunten van de categorie <i>anders</i> selecteerbaar
<input type="radio"/> Andrs	Voorkom dat datapunten van de categorie <i>anders</i> worden geselecteerd

Directe selectie van datapunten

Datapunten kunnen ook op directe wijze worden geselecteerd. Dit kan door te dubbelklikken en vervolgens slepen van de muis. Alle datapunten die binnen het nu gevormde vierkant vallen worden nu geselecteerd, ongeacht hun validatie status en datapunt categorie. Deze selectiemethode is bij uitstek geschikt voor het selecteren van datapunten die duidelijk outliers zijn, maar om wat voor reden dan ook niet door de automatische outlierdetectie zijn aangemerkt.

Indien de gegevens ook in een alfanumerieke lijst worden weergegeven, kunnen datapunten ook met behulp van deze lijst worden geselecteerd.

Tabel 6: Voorbeelden van het selecteren van datapunten met behulp van de muis

Instelling	Actie	resultaat
<p><input checked="" type="radio"/> Nieuw</p> <p><input checked="" type="radio"/> Outlr</p>	<p>Linker Muis Klik</p> 	
<p><input checked="" type="radio"/> Nieuw</p> <p><input checked="" type="radio"/> Outlr</p>	<p>Linker Muisklik op cirkel + Slepen</p> <p>of</p> <p>Shift + Linker Muisklik + Slepen</p> 	
<p>niet van toepassing</p> <p>(directe selectie op basis van voetmaat en diepte)</p>	<p>Linker Muisdubbelklik + Slepen</p> <p>Muis Loslaten</p> 	
<p>niet van toepassing</p> <p>(directe selectie op basis van positie in lijst)</p>	<p>Selecteer Regels met muis (LM klik, Shift+LM klik, of CTRL + LM klik)</p> 	




Acties op geselecteerde datapunten

Nadat één of meerdere datapunten zijn geselecteerd zijn een aantal acties mogelijk, welke bekrachtigd kunnen worden door drie buttons in de toolbar van het inspectiescherm van golfparameters (zie Figuur 10).



























Figuur 10: Knoppen voor het bewerken van datapunten. Van links naar rechts: Goedkeuren, Interpoleren naar referentiewaarde, en Markeren als hiaat

Tabel 7: Acties op geselecteerde datapunten en hun uitwerking

Button	Actie	Invloed op Validatiestatus	Invloed op datapunt categorie		Invloed op meetwaarde	Invloed op meetwaarde status
			Oude categorie	Nieuwe categorie		
	Goedkeuren	Validatiestatus wordt 'gevalideerd'	Ok	Ok	Geen	Geen
			Hiaat	Hiaat		
			Outlier	Ok		
			Anders	Anders		
	Interpoleren naar referentiewaarde	Validatiestatus wordt 'gevalideerd'	Categorie wordt 'ok'		Meetwaarde wordt gewijzigd naar referentiewaarde	Status wordt 'interpolatie' ²
	Markeren als hiaat	Validatiestatus wordt 'gevalideerd'	Categorie wordt 'hiaat'		Meetwaarde wordt gewist	Status wordt 'hiaat'

In principe werken deze knoppen voor alle parameter soorten tegelijk, maar het is mogelijk om deze acties ook apart per parametersoort toe te passen. Voor dit doel kan bijvoorbeeld een apart panel in het inspectie scherm worden geopend, waarin de knoppen uit Figuur 10 voor elke parameter soort apart herhaald worden. Bij normaal gebruik van de applicatie blijft dit panel gesloten.

² Geïnterpoleerde punten zijn in de grafiek te herkennen doordat ze met een cirkel geaccentueerd zijn.

Alle parameters			
Significante golfhoogte			
Significante golfhoogte			
Laagfrequente golfhoogte			
Gemiddelde golfperiode			
Significante golfperiode			
Gemiddelde hoofdgolfrichting			
Laagfrequente golfrichting			
Gemiddelde richtings spreiding			

Figuur 11: *Lay-out van een panel waarmee de status van individuele parametersoorten kan worden gewijzigd voor de geselecteerde periode(s)*

Inspectie van niet hiaten en niet outliers

Bij het valideren moet naast de mogelijkheid van het inspecteren van outliers en hiaten ook de mogelijkheid bestaan om de gehele tijdreeks te doorlopen door steeds een panelbreedte op te schuiven (met een kleine overlap). Hiertoe dienen aparte knoppen aanwezig te zijn.

5 Beschrijving software modules voor modelbouw

5.1 Toe te passen model

5.1.1 Eisen aan het model

Centraal in het modelbouwproces staat de toe te passen rekenmethodiek of rekenmethodieken. Het modelbouwproces houdt in dat deze methodiek wordt geconfigureerd voor een bepaalde combinatie van detectoren. Dit betekent dat binnen de ruimte die de methodiek hiervoor biedt de afhankelijkheden worden aangegeven en dat de nog onbekende parameters worden gekalibreerd. Het resultaat van deze processtap is een gekalibreerd model dat moet passen in de werkmethode van het operationeel valideren zoals beschreven in 3.2.

Als minimale eisen aan de toe te passen rekenmethodiek gelden daarom:

- De methodiek moet puntschattingen kunnen leveren voor de grootheden waarvoor in kolom 'afkeuren' van Tabel 3 'ja' is ingevuld. Daarbij mogen alle in Tabel 3 genoemde grootheden als invoer worden gebruikt;
- Dergelijke puntschattingen moeten onder alle omstandigheden kunnen worden geproduceerd, ook als één of meer van de invoervariabelen niet beschikbaar is;
- Deze puntschattingen moeten vergezeld zijn van een standaardafwijking;
- De methodiek moet configureerbaar zijn, dat wil zeggen dat het model overdraagbaar moet zijn in de ruimte. Voorzover het model parameters bevat die plaatsgebonden zijn, moeten deze op eenvoudige wijze te kalibreren zijn voor de nieuwe situatie.

Een speciale omstandigheid die benadrukt moet worden is dat tijdens de toepassing van het model de invoer van het model bestaat uit *niet gecontroleerde* gegevens. Dit wil zeggen dat deze data hiaten kunnen bevatten, of erger: data die tengevolge van technische of menselijke fouten volledig fout zijn maar niet als zodanig zijn aangemerkt. Ten slotte bevatten de invoerdata ook afwijkingen ten gevolge van meetonnauwkeurigheid. Van het toe te passen model wordt echter verwacht dat het onder alle omstandigheden met een puntschatting vergezeld van een realistische standaardafwijking komt.

Overige criteria die meespelen bij het kiezen van een toe te passen methodiek zijn:

- De benodigde inspanningen voor het ontwikkelen van de methode;
- De benodigde inspanningen voor de implementatie;
- Het gemak waarmee de methode voor nieuwe situaties kan worden geconfigureerd, inclusief de noodzaak om na een dergelijke configuratie controles uit te voeren;
- De precisie van de geleverde puntschattingen;
- Het realisme van de door de standaardafwijking geïmpliceerde betrouwbaarheidsmarges;
- De rekentijd die nodig is bij operationele toepassing van de methode;
- De afmetingen van de dataset die kan worden verwerkt;
- De transparantie van de rekenmethode voor gebruiker.

Desgewenst kan gebruik gemaakt worden van het model dat momenteel in WAVIX gebruikt wordt. Dit model voldoet aan de meeste van de bovenstaande voorwaarden. Een praktisch probleem zou kunnen bestaan uit het feit dat er een gedetailleerde specificatie van het model op grond waarvan het model kan worden nagebouwd vooralsnog niet beschikbaar is. De volgende sectie beschrijft het huidige model in wat grotere lijnen. Verdere informatie is te vinden in Van Noort (1997).

5.1.2 Huidige model

De gedachte achter het huidige model is dat altijd schattingen moeten worden geproduceerd, ongeacht de hoeveelheid hiaten in de invoer. Indien een goede schatting vanwege het optreden van hiaten niet gemaakt kan worden dan komt dit tot uiting in de geschatte betrouwbaarheidsmarge. Deze marge varieert dus in de tijd. Het schattingsproces is beschreven in Van Noort (1997). Dit document vormt echter geen eenduidige specificatie van de methode, omdat het document informatie zoals schema's die de berekeningswijze weergeven en wiskundige definities van doelfuncties en stopcriteria niet bevat. De volgende paragrafen en Figuur 12 moeten daarom beschouwd worden als een poging om het schattingsproces samen te vatten, voor het opstellen van volledige technische specificaties is nog nader onderzoek nodig.

Het schattingsproces is feitelijk een iteratief proces waarbij met een serie aparte backpropagation neurale netwerken voorspelde waarden worden gegenereerd voor alle combinaties van detector en datatype. De afhankelijkheden die door deze neurale netwerken worden gemodelleerd zijn vooraf gespecificeerd (afhankelijk van de te schatten detector-datatype combinatie wordt dus een bepaalde selectie van de beschikbare invoer gebruikt).

De bijbehorende nauwkeurigheid wordt met een numerieke methode (die geen deel uitmaakt van een neuraal netwerk) bepaald. In een eveneens numerieke afweging wordt bepaald of de verzameling van geschatte waarden 'consistent' is. Het criterium dat hiervoor gebruikt wordt komt erop neer dat de voorspelde waarden (de afhankelijke variabelen voor iteratie n) een optimale gelijkens met de waarden waarop deze voorspellingen zijn gebaseerd (de onafhankelijke variabelen voor iteratie n). Zo niet dan wordt de invoer voor de volgende iteratie gelijk gesteld aan de uitvoer van de huidige iteratie en worden nieuwe voorspelde waarden berekend voor alle detector-datatype combinaties.

De doelfunctie die door dit proces wordt geminimaliseerd is gegeven door de onderstaande vergelijking (zie [Van Noort (1997)], vgl (31)):

$$E(S) = \frac{1}{N} \sum_{i=1}^N \frac{(S_i(t+1) - S_i(t))^2}{(\sigma_i^2(t+1) - \sigma_i^2(t))^2} \quad (0)$$

met:

$S_i(t)$: De geschatte waarde voor parameter i in iteratie t

$\sigma_i^2(t)$: De bijbehorende geschatte variantie

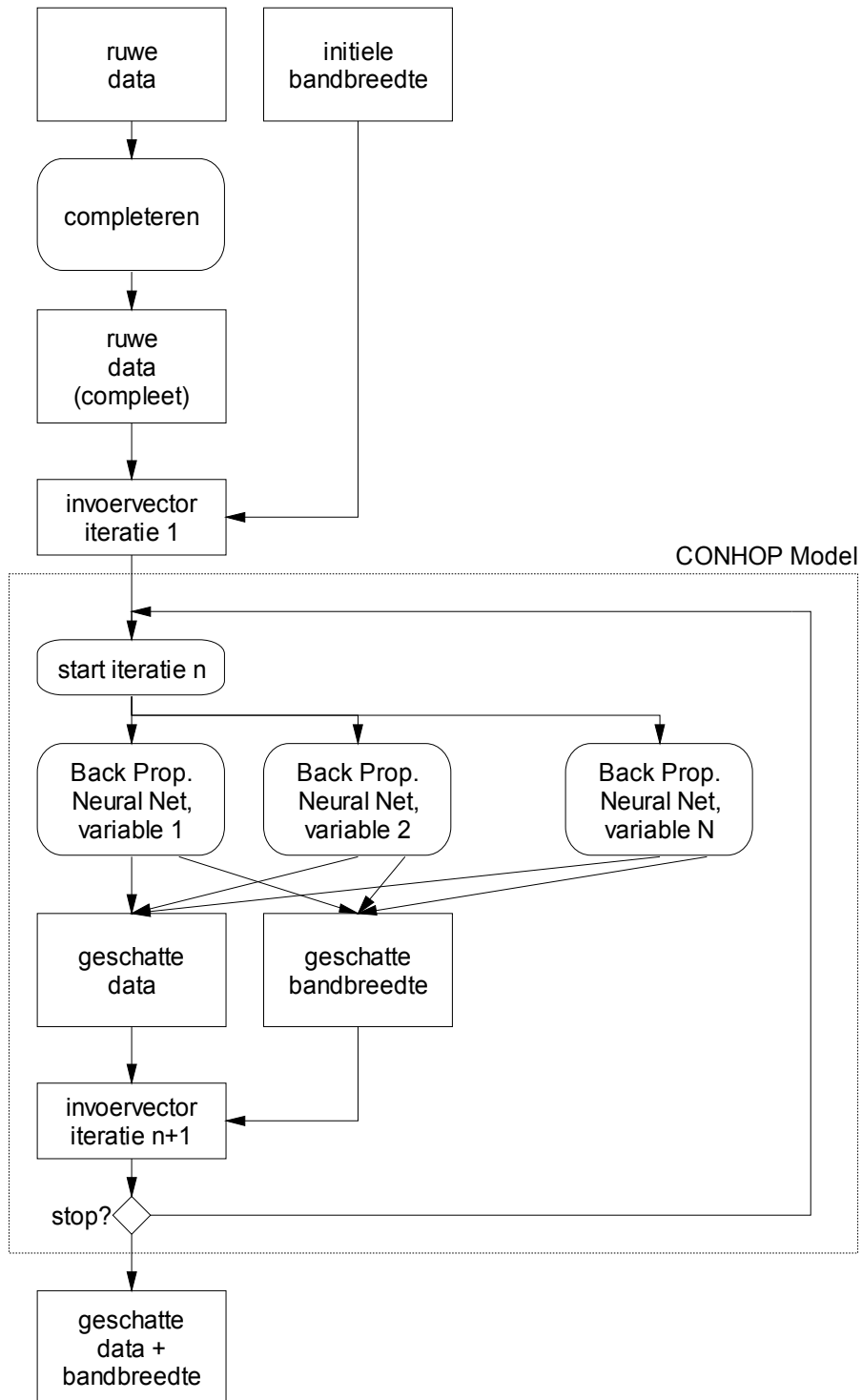
Wat overigens opvalt in deze vergelijking is dat deze alleen test op de consistentie tussen twee opeenvolgende iteraties, en niet op consistentie met de oorspronkelijke gemeten waarden. Ten behoeve van de robuustheid van de methode lijkt het wenselijk om een term toe te voegen die het sterk afwijken van de oorspronkelijke meetwaarde bestraft, zoals bijvoorbeeld in de onderstaande vergelijking.

$$E(S) = \frac{1}{N} \sum_{i=1}^N \frac{(S_i(t+1) - S_i(t))^2}{(\sigma_i^2(t+1) - \sigma_i^2(t))^2} + \frac{(S_i(t+1) - S_i(0))^2}{(\sigma_i^2(t+1) - \sigma_i^2(0))^2} \quad (0)$$

met:

$S_i(0)$: De waargenomen waarde voor parameter i

$\sigma_i^2(0)$: De bijbehorende geschatte variantie



Figuur 12: *Rekenschema CONHOP model*

5.1.3 Alternatieve modelvormen

Hoewel in het tot stand komen van het huidige model veel tijd is geïnvesteerd mogen varianten van het huidige model of geheel nieuwe modelvormen niet op voorhand worden uitgesloten. De twee belangrijkste alternatieven die in aanmerking komen zijn:

- *Multivariate lineaire regressie*. Als onafhankelijke variabelen kunnen naast de grootheden uit Tabel 3 ook hogere machten van deze grootheden worden toegepast. Dit laatste is alleen nodig indien er aanleiding bestaat om te veronderstellen dat er een niet lineair verband bestaat tussen, bijvoorbeeld, windsterkte en golfhoogte;
- *Neurale netwerken*. In dit geval wordt feitelijk de huidige oplossing toegepast, maar wordt afgezien van het iteratieve optimalisatieproces. Mogelijk zijn nu uitbreidingen nodig om te voorkomen dat bij het optreden van hiaten geen voorspelling kan worden geleverd.

In beide gevallen kan een spreidingsparameter worden berekend door de residuen te analyseren. Deze spreidingsparameter is echter niet afhankelijk van de invoerdata, zoals bij Conhop model wel het geval is. Ook weken de bovengenoemde modellen alleen in het geval dat alle benodigde invoerdata aanwezig zijn. Om deze modellen in situaties toepasbaar te maken waarin invoerdata ontbreken moeten er of meerdere versies van deze modellen worden gekalibreerd (analoog aan de situatie bij het voorspellen van waterstanden in MELISSA) of moeten de gegevens gecompliceerd worden voordat het model wordt toegepast.

Bij de Neurale Netwerk toepassing is het voorstelbaar dat de invoer van het neurale netwerk bestaat uit zowel de gemeten waarden als de corresponderende codes die aangeven welke van deze waarden hiaten zijn. Bij het inleren zou dan een modelvorm moeten volgen die robuust is voor het uitvallen van detectoren. Dit is echter een methode waarvan het succes nog niet experimenteel is aangetoond.

Bij de methode gebaseerd op multivariate lineaire regressie kunnen de geschatte waarden worden uitgedrukt als een lineaire combinatie van de invoer variabelen. Wanneer de invoer variabelen nu zelf geschatte waarden zijn, waarvan de spreiding bekend is, dan kan een uitdrukking worden gevonden voor de spreiding van de geschatte waarde op basis van de spreiding van de invoer variabelen. Deze uitdrukking kan worden toegepast indien een of meerdere invoer variabelen op basis van vuistregels is vastgesteld.

Beide genoemde bovengenoemde methodes scoren dankzij hun eenvoud hoog op de aspecten implementatiekosten en transparantie. Omdat er regelmatig hiaten opduiken in de dataset is het bij toepassing van deze methode noodzakelijk dat er meerdere versies van deze modellen worden gekalibreerd, zodat bij het ontbreken van de invoer voor het ene model overgeschakeld kan worden naar een andere versie.

5.1.4 Toe te passen model: conclusie

Er tekenen zich drie mogelijkheden af voor wat betreft de keuze van een model:

- Methode 1: het toepassen van de huidige modellen. Het betreft hier een combinatie van neurale netwerken en een specifiek ontwikkelde numerieke methode;
- Methode 2: het toepassen van neurale netwerken. Het betreft hier de huidige methode, maar dan exclusief het iteratieve gedeelte;
- Methode 3: het toepassen van multivariate lineaire regressie.

Al deze drie methodes voldoen aan de minimum eisen zoals genoemd in sectie 5.1.1. In hoeverre ze aan de overige beoordelingscriteria voldoen kan op dit moment alleen subjectief worden ingeschat:

Tabel 8: *Subjectieve inschatting van de score van de voorgestelde methodes op de criteria uit sectie 5.1.1*

	Methode 1: Huidige methode	Methode 2: Neuraal Net (Huidig excl. numerieke methode)	Methode 3: MLR
Ontwikkelingskosten	laag	beperkt	beperkt
Implementatiekosten	hoog	beperkt	laag
Inspanningen voor herconfiguratie	hoog	beperkt	laag
Precisie puntschattingen	goed	acceptabel	acceptabel
Realisme betrouwbaarheidsmarges	goed	acceptabel	acceptabel
Rekentijd	hoog	laag	laag
Maximale Afmetingen dataset (uurlijkse metingen)	onbekend	> 100 jr	> 100 jr
Transparantie	slecht	redelijk	goed

Op dit moment lijkt aan een definitieve keuze voor de huidige rekenmethode nog een risico te kleven omdat er geen technische specificatie van de methode beschikbaar is. Wanneer voor deze methode wordt gekozen zal enig experimenteren nodig zijn om de precieze vorm van het gewenste model te bepalen. Omdat het idee achter de bestaande methode, zeker bij de betrokkenen, bekend is en eerdere experimenten al hebben aangetoond dat de methode werkt worden hier echter geen onoverkomelijke problemen verwacht.

In hoeverre de huidige methode zich leent voor het herconfigureren ervan voor nieuwe detectorcombinaties door non-expert gebruikers is nog niet uit de praktijk bekend, maar het zou vreemd zijn als de toepasbaarheid van de methode gebonden zou zijn aan bepaalde locaties.

Een meer principieel punt is de theoretische onderbouwing van de huidige methode. Vooral het principe dat voorspellingen worden gebaseerd op voorspellingen zou na een aantal iteraties gemakkelijk tot een accumulatie van fouten of zelfs instabiliteit kunnen leiden. Om de methode meer robuust te maken lijkt het raadzaam om tenminste een strafterm in de doelfunctie op te nemen die grote afwijkingen ten opzichte van de oorspronkelijke meetwaarde tegengaat, zoals gesuggereerd in vergelijking (0).

De andere twee oplossingen brengen geen risico met zich mee. Wanneer methode twee wordt geïmplementeerd, dan kan methode drie tegen betrekkelijk geringe meerkosten ook worden geïmplementeerd. Dit komt omdat beide methodes wat betreft hun in- en uitvoer sterk op elkaar lijken, en doordat het algoritme achter methode drie betrekkelijk eenvoudig is.

5.2 Software modules

In deze sectie wordt een overzicht gegeven van de softwaremodules die nodig zijn om het modelbouwproces te ondersteunen.

5.2.1 Module voor het handmatig opschonen van de dataset

Hiervoor hoeven geen aparte software modules vervaardigd te worden. Met de functies die beschikbaar zijn ten behoeve van de off-line validatie kan deze taak worden uitgevoerd.

5.2.2 Module voor het aanmaken van een kalibratie- en controleset

Ook hier hoeven geen aparte software modules te worden geschreven. De functies voor het databeheer dienen voldoende algemeen te zijn om het mogelijk te maken dat een gedeelte van de dataset wordt weggeschreven. Ten behoeve van de kalibratie en de controle kunnen vervolgens apart werkgebieden worden aangemaakt.

5.2.3 Module voor de definitie van de aan te brengen afhankelijkheden

Ongeacht welk model gebruikt wordt, de gebruiker zal in alle gevallen moeten aangeven welke afhankelijkheden aanwezig zijn. De modelkalibratie zal zich vervolgens tot deze afhankelijkheden beperken. Deze afhankelijkheden hebben de volgende vorm:

De reeks voor de combinatie:

detector	datatype
----------	----------

hangt af van de reeksen voor de volgende combinatie:

detector1	datatype1
detector2	datatype2
detector3	datatype3
detector4	datatype4
detector5	datatype5
detector6	datatype6

Het komt regelmatig voor dat zich hiaten in bevinden in (groepen van) variabelen. In dit geval kan het model geen voorspellingen leveren. De huidige methode (sectie 5.1.4, methode 1) ondervangt dit door in dit geval de ontbrekende invoerwaarden te vervangen door schattingen van invoerwaarden.

Wanneer methodes 2 of 3 worden toegepast zijn aanvullende maatregelen nodig, zoals het werken met een aantal aparte definities van afhankelijkheid, analoog aan de wijze waarop dit gebeurt bij de validatie van waterstanden in MELISSA (zie RIKZ, 1999). Indien de ene set van afhankelijkheden niet geëvalueerd kan worden, wordt overgeschakeld op de tweede set. De aanpak met meerdere definities van afhankelijkheid kan worden verfijnd door op een aantal niveaus een keuze toe te staan. In dit geval kan de totale verzameling afhankelijkheidsrelaties automatisch worden bepaald. Beschouw bijvoorbeeld de volgende definitie:

detector	datatype
----------	----------

hangt af van:

een van de volgende:	detector1a	datatype1a
	detector1b	datatype1b
	detector1c	datatype1c
en	detector2a	datatype2a
een van de volgende:	detector2b	datatype2b
en	detector3	datatype3
en	detector4a	datatype4a
een van de volgende:	detector4b	datatype4b
en	detector5	datatype5
en	detector6	datatype6

Op deze wijze zijn in feite $3 \times 2 \times 2 = 12$ sets van afhankelijkheden gedefinieerd, zonder dat de daarvoor een bewerkelijk invoerproces nodig is. Ieder van deze afhankelijkheidsdefinities impliceert een apart model dat gekalibreerd dient te worden.

Dit is overigens een methode die sterk lijkt op een methode die in het pre-Wavix tijdperk is onderzocht met matige resultaten. Bij golfgegevens treden relatief veel hiaten op (dit geldt misschien voor de laatste 10 jaar in mindere mate), er zijn weinig meetlocaties, dus ook veel minder meetlocaties in de nabijheid. Bovendien komt het redelijk vaak voor dat door een storing bij het inwinnen/verwerken alle gegevens over bijvoorbeeld 24 uur ontbreken.

5.2.4 Implementatie van het voorspellende model

Zie ook sectie 4.4.2. Het voorspellende model wordt aangeropen als onderdeel van het kalibratieproces.

5.2.5 Kalibratiemodule

De inhoud van het kalibratieproces hangt sterk af van de gekozen model vorm, maar zal in alle gevallen uit de volgende stappen bestaan:

- De gebruiker zet de kalibratiedata klaar in werkgebied. Deze data zijn zo goed mogelijk gecontroleerd. Verdachte data zijn uit de dataset verwijderd. De dataset moet voldoende groot zijn en bijvoorbeeld een periode van enkele jaren beslaan;
- De gebruiker definieert de afhankelijkheden die tijdens de kalibratie gebruikt worden;
- De automatische kalibratie wordt opgestart. Het kalibratie proces neemt geruime tijd in beslag. Tijdens dit proces wordt een progress bar getoond. Na afloop van het proces wordt een verslag gegenereerd;
- De gebruiker inspecteert het resultaat van de kalibratie. Hiertoe kunnen verschillende soorten overzichten worden gegenereerd.

Voor het uitvoeren van de derde stap, de automatische kalibratie moet specifieke software worden geschreven afhankelijk van de toe te passen methode (zie sectie 5.1.4). De onderstaande tabel geeft aan wat minimaal nodig is:

Tabel 9: *Benodigde software voor de kalibratie van het toe te passen model*

methode	benodigde software
methode 1	<ul style="list-style-type: none"> • Kalibratie software voor de neurale netwerken. Alle netwerken dienen apart getraind te worden • Het schatten van de overige parameters in het CONHOP model
methode 2	<ul style="list-style-type: none"> • Kalibratie software voor de neurale netwerken. Alle netwerken dienen apart getraind te worden
methode 3	<ul style="list-style-type: none"> • Algoritme voor het schatten van de coëfficiënten in een kleinste kwadraten probleem

5.2.6 Overzichten

Vaak worden tijdens het bouwen van een model beslissingen genomen naar aanleiding van grafieken die niet of juist wel een bepaald verband laten zien. Het is moeilijk van te voren te bepalen welke grafieken precies nodig zijn. De mogelijkheid om elke willekeurige reeks voor een locatie uit te zetten tegen elke ander willekeurige reeks is echter vrij eenvoudig te realiseren en dient minimaal aanwezig te zijn.

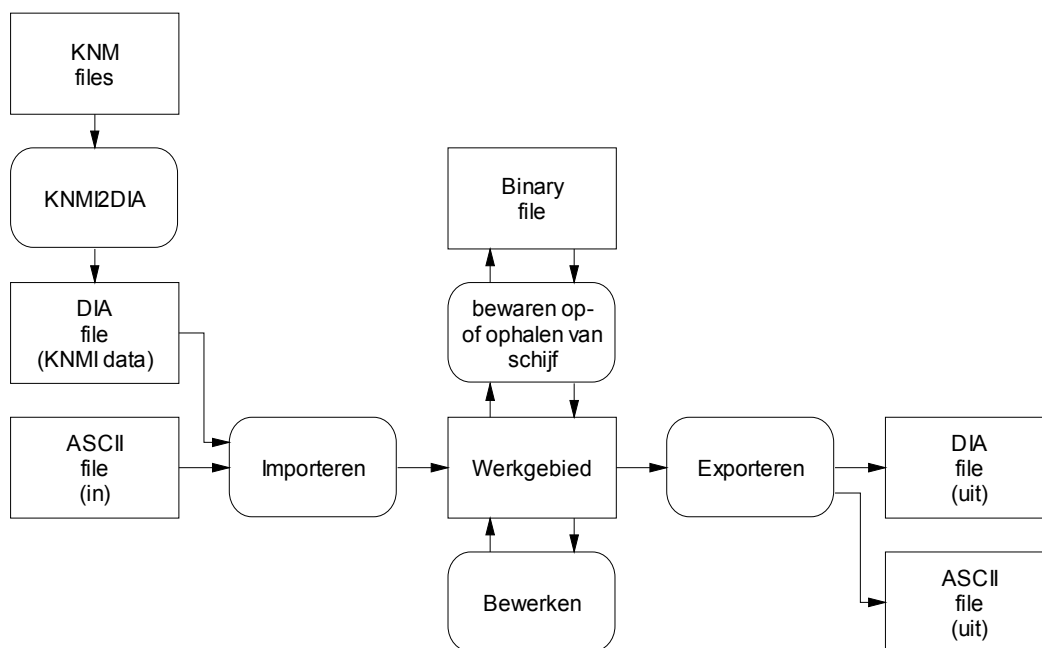
Ten behoeve van de meer uitgebreide analyses dient tevens een export mogelijkheid naar bijvoorbeeld ASCII (en daarmee ook Excel) te bestaan.

Omdat veel numerieke analyses binnen RIKZ in Matlab worden uitgevoerd is een exportmogelijkheid naar binair Matlab formaat ook wenselijk.

6 Beschrijving software modules voor invoer en uitvoer

6.1 Externe relaties

De invoer voor de golvenvalidatie bestaat uit golf-, waterstands- en windmetingen. De uitvoer betreft alleen golfgegevens. Golf- en waterstandsgegevens worden uit de DONAR database betrokken in de vorm DIA (Donar Interface ASCII) files. De windmetingen worden rechtstreeks van het KNMI betrokken in een specifiek ASCII formaat. Via een aparte utility (KNMI2dia) worden deze gegevens naar dia formaat geconverteerd.



Figuur 13: Externe relaties

6.2 Huidige situatie: Dia2Wavix en Wavix2DIA

In de huidige versie bestaat een aparte conversie module, DIA2WAVIX, die de invoer op compleetheid controleert en zo nodig completeert en ten behoeve van WAVIX in een specifiek ASCII formaat wegschrijft. Iets soortgelijks gebeurt met de uitvoer van WAVIX: deze wordt geconverteerd naar het DIA formaat en verrijkt met de relevante Meta-informatie opdat deze data weer in de DONAR database kan worden weggeschreven.

De in Tabel 3 genoemde variabelen zijn binnen DONAR is gekenmerkt door:

- Een locatie (in DONAR te identificeren aan de hand van het W3H veld sLoccod)
- Een detectortype (in DONAR te identificeren aan de hand van het W3H veld sVatcod)
- Een gegevenssoort (in DONAR te identificeren aan de hand van het W3H veld sParcod)

Binnen WAVIX wordt voor golfgegevens tevens onderscheid gemaakt tussen hoofdsensoren en nevensensoren. DONAR kent een dergelijk onderscheid niet. Via een conversie tabel worden de diverse reeksen uit DONAR gekoppeld aan de hoofd- en nevensensoren binnen WAVIX. Tabel 10 geeft aan hoe de koppeling in de huidige WAVIX versie is gedefinieerd. Tabel 11 geeft daarbij aan hoe de golf-, wind- en waterstandsgegevens kunnen worden uitgesplitst naar soort. Omdat WAVIX geen mogelijkheden bevat om wind- en waterstandsgegevens bij te schatten, zijn deze velden momenteel verplichte invoer.

Tabel 10: Karakterisering van de WAVIX invoergegevens

Locatie	Golfgegevens sVatcod1: hoofdsensoren sVatcod2: nevensensoren			Windgegevens (hoofdsensoren en nevensensoren aan elkaar gelijk)		Waterstanden (hoofdsensoren en nevensensoren aan elkaar gelijk)	
	sLoccod	sVatcod1	sVatcod2	sLoccod	sVatcod	sLoccod	sVatcod
AUK	AUKFPFM	WAVEC	NB	WESTTSLG	FASTRCDR	DENHDR	DNM
K13	K13APFM	WAVEC	WAVERDR	KOOS	FASTRCDR	DENHDR	DNM
SON	SCHIERMNOND	WAVEC	NB	WESTTSLG	FASTRCDR	HUIBGT	DNM
ELD	EIELSGT	WAVEC	NB	KOOS	FASTRCDR	DENHDR	DNM
YM6	IJMDMNTSPS	WAVEC	WAVERDR	IJMDSMFR	FASTRCDR	IJMDBTHVN	DNM
MPN	NOORDWMPT	NB	MARN300	NOORDWMPT	FASTRCDR	NOORDWMPT	MARN300
EUR	EURPFM	WAVEC	MARN300	EURPFM	FASTRCDR	EURPFM	DNM
LEG	LICHELGRE	MARN300	NB	LICHELGRE	FASTRCDR	LICHELGRE	ETMTA
SWB	SCHOUWBK	WAVERDR	NB	LICHELGRE	FASTRCDR	BROUWHVSGT02	NB
SCW	SCHEURWT	WAVERDR	VLAKTVDRN0 2 WAVERDR	LICHELGRE	FASTRCDR	BROUWHVSGT02	NB

Tabel 11: Parameteromschrijving van in WAVIX tabel op te nemen kolommen

Categorie	Parameter	sParcod	Status	(WAVIX)Eenheid	(DONAR)Eenheid
Golfgegevens	1	Hm0	facultatief	[cm]	[cm]
	2	H1/3	facultatief	[cm]	[cm]
	3	HTE3(*)	facultatief	[cm]	[cm]
	4	Tm02	facultatief	[.1 s]	[s]
	5	TH1/3	facultatief	[.1 s]	[s]
	6	Th0	facultatief	gr	gr
Windgegevens	7	WINDRTG	verplicht	gr	gr
	8	WINDSHD	verplicht	[.1 m/s]	[m/s]
Waterstanden	9	WATHTE	verplicht	[cm]	[cm] tov NAP/MSL

(*): Voor omstreeks 1989 werd HTE3 niet direct in DONAR opgeslagen. In plaats daarvan werd TE3 opgeslagen, met $HTE3=HE10= 4*\sqrt{TE3}$

In de huidige situatie wordt het proces van inlezen en controle op compleetheid vergemakkelijkt door het feit dat de detectorconfiguratie vast is. Hierdoor kunnen bepaalde beslisregels als onderdeel van de software worden toegepast.

6.3 Nieuwe situatie: invoer

De functies die nu vervuld wordt door Dia2Wavix zal in de nieuwe situatie geïntegreerd moeten worden in de software. In de nieuwe situatie is bij het opstellen van een nieuw model de detectorconfiguratie niet langer vast. Uit het model toe te passen wel af te leiden welke combinaties van detector en datatype nodig zijn. Hierdoor is het mogelijk om de invoer ook in de nieuwe situatie op compleetheid te controleren.

Het verwerkingsproces bestaat uit de volgende stappen:

- Het controleren van de meegegeven invoer;
- Het inlezen van de KNMI-ASCII files. Deze worden eerst naar DIA files geconverteerd. De manier waarop dit gebeurt is beschreven in de sectie;
- Het inlezen van de invoerdia's. De invoer kan uit meerdere dia's betrokken worden, maar moet wel binnen één operatie worden ingelezen. Het is dus niet mogelijk om achteraf gegevens aan een werkgebied toe te voegen zonder ook de oorspronkelijke data weer opnieuw in te lezen;
- Eventueel: het inlezen van ASCII data met golfgegevens uit andere applicaties;
- Het opslaan van de metagegevens in het werkgebied. omdat bij het wegschrijven van de DIA bepaalde meetgegevens nodig zijn, zoals het waarnemingssoortnummer, worden deze samen met reeks in het werkgebied opgeslagen;
- Het samenvoegen van de invoerdia's tot één geheel;
- Het afdrukken van een overzicht van de invoerdia's;
- Het synchroniseren van de verschillende reeksen naar één tijd as en het invullen van hiaten voor de ontbrekende periodes;

6.4 Nieuwe situatie: uitvoer

De functie die nu vervuld wordt door Wavix2Dia dient geïntegreerd te worden in de nieuwe software. Naast de uitvoerreeksen bevat de uitvoer dia per reeks minimaal de volgende meta informatie:

- W3H.IWnsnum
- W3H.sDomein
- W3H.sParcod
- W3H.sEhdcod
- W3H.sRkstyp
- RKS.sTydehd
- RKS.iTydstp

Deze meta informatie is bij het inlezen van de data al in het werkgebied opgeslagen.

6.5 Knmi2Dia

De windgegevens worden door het KNMI aangeleverd. Via de bestaande utility KNMI2DIA kunnen deze files naar DIA formaat worden geconverteerd. Deze utility voert tevens enige controles uit en kan desgewenst in de nieuwe applicatie worden geïntegreerd. In het andere geval dienen de gegevens direct uit de KNMI-ASCII files te worden ingelezen.

Anno 2003 zijn de KNMI-ASCII files te downloaden van de locatie:
<http://www.knmi.nl/samenw/hydra>

Opmerking:

De KNMI-windgegevens zijn meestal pas na langere tijd (orde halfjaar) beschikbaar (via internet). Wanneer recente golfgegevens moeten worden gecontroleerd is dit niet werkbaar. In dat geval kunnen de windgegevens worden gebruikt die dagelijks door het HMR worden geleverd. Deze moeten dan natuurlijk ook bij de training worden gebruikt. De KNMI-gegevens van de samenw/hydra-site zijn gevalideerd en dus beter. Om deze te gebruiken voor de eerste validatie, moet wellicht een (betaalde) overeenkomst met het KNMI worden gesloten om deze gegevens bijvoorbeeld binnen een maand beschikbaar te krijgen. Het is overigens de vraag of de bovengenoemde site up to date zal worden gehouden nu het samenwerkingsproject HYDRA is afgelopen.

Indien de gedownloadede files als invoer worden meegegeven aan Dia2Wavix, worden deze files naar een gelijknamige DIA geconverteerd. Hierbij gelden de volgende opmerkingen:

- KNMI hanteert GMT tijd. Ten behoeve van Wavix is hier een uur bij opgeteld;
- In KNMI files kunnen negatieve windsnelheden voorkomen. Deze dienen als hiaat te worden aangemerkt;
- In KNMI files kunnen negatieve windrichtingen voorkomen. Deze dienen als hiaat te worden aangemerkt;
- In KNMI files kunnen windrichtinggegevens groter dan 360 graden voorkomen. Het betreft hier de richting '990'. Dit getal wordt door het KNMI ingevuld indien de wind veranderlijk is. Ten behoeve van WAVIX dienen deze records als hiaat te worden weggeschreven;
- De windrichting '0' staat volgens de in de KNMI files specificatie voor 'de richting die bij windstil weer optreedt'. Ten behoeve van WAVIX worden dienen deze records als hiaat te worden weggeschreven;
- De onderstaande tabel geeft aan welke waarden met de bijbehorende kwaliteitscodes door de software naar de wavixtabel worden weggeschreven:

code	omschrijving	aangetroffen in voorbeeld files	weggeschreven naar WAVIX tabel (*)
-1	no data	ja	nee
0	valid data	ja	ja
2	data taken from WIKLI-archives	nee	ja
3	wind direction in degrees computed from points of the compass	nee	ja
4	onbekend	ja	ja
6	added data	ja	ja
7	missing data	ja	nee
100	suspected data	nee	ja

(*): Indien 'ja' worden de bijbehorende periodes als OK (kwaliteitscode 0) weggeschreven. Indien 'nee' worden de bijbehorende periodes als hiaat aangemerkt.

6.6 ASCII export optie

Om gegevens uit het werkgebied ook in andere applicaties te kunnen verwerken dient een eenvoudige export optie voor ASCII files aanwezig te zijn. Met deze optie worden de numerieke reeksen voor de geselecteerde detector-datatype combinaties geëxporteerd, zonder metadata.

7 Technische aspecten

In dit hoofdstuk zijn een aantal randvoorwaarden van uiteenlopende aard opgenomen, die betrekking hebben op de implementatie wijze.

7.1 Eisen aan de gebruikers interface

Deze gebruikersinterface moet aan de volgende eisen voldoen:

- Voor alle bewerkingen die worden uitgevoerd, moet gelden dat ze ook weer ongedaan kunnen worden gemaakt door de gebruiker door ‘undo’ en ‘redo’ functies. Dit geldt met name voor de validatie beslissingen, waarvan er vele op een rij worden genomen;
- De applicatie dient toegankelijk te zijn voor nieuwe gebruikers door functies op logische wijze te ordenen in menu’s;
- De applicatie dient gebruiksvriendelijk te zijn voor regelmatige gebruikers door veel gebruikte functies toegankelijk te maken vanuit shortcut-keys en een toolbar;
- De applicatie dient van een on-line help faciliteit voorzien te zijn met de volgende faciliteiten:
 - Zoeken op basis van een hoofdstuk indeling;
 - Zoeken op basis van een keywords;
 - Op vragen van hulp bij specifieke beeldscherm elementen (beter bekend als context sensitive help);
 - Het weergeven van toelichtende informatie door middel van zogenaamde tooltips.

7.2 Voorwaarden aan de implementatiewijze

De implementatie dient aan het volgende te voldoen:

- Indien de implementatie met behulp van een ontwikkeltool, zoals Matlab, wordt gebouwd dient compileerbaar te zijn, zodat de applicatie zonder een Matlablicentie te gebruiken is;
- Er moeten voorzieningen gebouwd worden die het mogelijk maken om ook met zeer grote datasets te werken.

7.3 Beheer en onderhoud

- De performance van de applicatie, en dan vooral de interactieve gedeeltes, moet van een ‘voldoende’ niveau te zijn, af te meten op een machine met een CPU snelheid van 2,5Mhz of beter en tenminste 512 Mb aan werkgeheugen;
- RIKZ zal verantwoordelijk voor het onderhoud van de applicatie zijn;
- De Morfologie applicatie zal op een inzichtelijke manier gedocumenteerd dienen te zijn, waarbij de documentatie bestaat uit:
 - Dit functioneel ontwerp;
 - Een apart document dat de werking van de modules in stroomschema’s en tekst uitlegt (aan te duiden als Technische Specificatie);
 - Documentatie op broncodeniveau die per module het volgende bevat:
 - een beschrijving van de invoer, aanroepwijze en de uitvoer;
 - een gedetailleerde beschrijving van de werking van de module;
 - toelichtend commentaar bij de softwarecode.

- Alle broncode dient meegeleverd te worden, opdat RIKZ in staat is de reguliere upgrades uit te voeren van compilers en andere bij RIKZ in gebruik zijnde tools (zoals Matlab).

8 Literatuur verwijzingen

Bouws, E., (1984) Beoordeling van golfgegevens van het Meetnet Noordzee bewerkt door het nieuwe CIC-systeem in Hoek van Holland. KNMI TR-62, 1985

Bouws, E., (2001) EDITGP-files: Compacte jaarreeksen met golfdata van de Noordzee, <http://www.knmi.nl/~bouws/nl/editgp.htm> 30 jan 2001

Kuik, A.J., G.Ph. van Vledder and L.H.Holthuijsen (1988) A method for the routine analysis of pitch-and-roll buoy wave data. J.Physical Oceanogr. 18, 1020-1034, 1988

RIKZ (1999) Handleiding Melissa, 14 dec. 1999, MobiData in opdracht van RIKZ

RIKZ (2002) DIA2WAVIX en WAVIX2DIA, Versie 2, Documentatie, MobiData in opdracht van RIKZ

RIKZ (2003) Golfklimaatsite, <http://www.golfklimaat.nl>

Van Noort, G.J.H.L. (1997) Wavix, Een Neuraal Systeem voor Controle en Correctie Werkzaamheden op Golfmetingen, All Fours Neural Tech i.o.v. RIKZ, Contactpersoon RIKZ P.F. Heinen