

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



info@modelit.nl
www.modelit.nl

Technische Documentatie WAVIX: Datastructuur en modules

Datum 8 Mei 2004
Wijzigingen 20100919 Titel gewijzigd
Correcties doorgevoerd

Modelit
KvK Rotterdam 24290229
Postbank 8086419

1 Inleiding

Dit document behoort tot de technische documentatie van Wavix en bevat een beschrijving van de datastructuur en een beschrijvingen van WAVIX modules

2 Representatie van gegevens in het werkgebied

2.1 Overwegingen

Als we ons beperken tot golfdata, dan kunnen we stellen dat in DONAR gegevens zijn opgeslagen in reeksen die van elkaar zijn te onderscheiden op basis van de locatie, de parametersoort en het veldapparaat. Wanneer deze reeksen geen onderlinge samenhang zouden hebben dan zouden de validatie reeks voor reeks kunnen worden afgehandeld. Voor de opslagstructuur zou dit betekenen dat alle reeksen in een lijst kunnen worden opgenomen en dat geen clustering van reeksen per locatie nodig zou zijn. Figuur 1 toont de bijbehorende opslag structuur (de waargenomen reeks en de gecorrigeerde reeks kunnen desgewenst in hetzelfde veld worden opgeslagen)

Locatie
Parameter
Sensortype
Waargenomen reeks+ stdafw
Geschatte reeks+ stdafw
Gecorrigeerde reeks+ stdafw

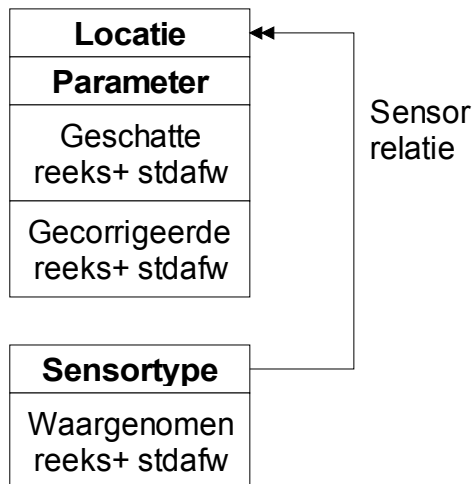
Figuur 1: *Opslag structuur, indien iedere reeks apart wordt beschouwd. De combinatie Locatie-Parameter-Sensortype bepaald eenduidig de reeks. Alle gegevens worden in één tabel opgeslagen (waarbij een reeks als een van de velden van de tabel wordt beschouwd)*

Uiteraard hangen de reeksen wel met elkaar samen. Er zijn twee soorten van samenhang tussen de reeksen te onderscheiden, waar vervolgens ieder op een eigen manier moet worden omgegaan:

- Functionele samenhang. Hiermee wordt bedoeld dat er twee verschillende grootheden worden waargenomen, maar dat deze grootheden op basis van fysische verbanden met elkaar samenhangen. Bijvoorbeeld de windsterkte en de golfslag op een locatie hangen met elkaar samen, of de golfhoogte op de ene locatie hangt samen met de golfhoogte op een locatie 20 kilometer verderop.
- Redundantie. Hiermee wordt bedoeld dat een bepaalde grootte zoals golfhoogte of golfperiode op een locatie wordt waargenomen met verschillende waarnemingsmethoden. In de huidige praktijk worden nooit meer dan twee waarnemingsmethoden per locatie toegepast. Deze worden aangeduid als de

hoofd en de nevensensor. In theorie zouden ook meer dan twee waarnemingsmethodes van toepassing kunnen zijn.

De vraag is wat dit betekent voor de opslagstructuur, met name waar het de geschatte en gecorrigeerde waarden van redundant waargenomen reeksen betreft. Per combinatie van locatie en parameter kunnen meerdere reeksen met waarnemingen beschikbaar zijn, maar uiteindelijk dient er voor een dergelijke combinatie uiteindelijk maar één reeks met geschatte en gecorrigeerde data over te blijven.



Figuur 2: *Opslag structuur, indien reeksen in hun onderlinge samenhang worden beschouwd. De combinatie Locatie-Parameter bepaalt eenduidig de grootheid. Per grootheid kunnen meerdere waarnemingsreeksen beschikbaar zijn, ieder gekenmerkt door hun eigen sensortype. De gegevens worden verspreid over twee tabellen opgeslagen.*

Het nadeel van de opslagstructuur zoals in Figuur 2 weergegeven, is dat de relatie tussen de gecorrigeerde reeks en de oorspronkelijke meetreeks niet meer eenduidig is. Wanneer de reeks vervolgens wordt teruggeplaatst in DONAR is niet meer aan te geven wat de herkomst van de gegevens is en zouden strikt genomen alle datapunten de status ‘geïnterpoleerd’ moeten krijgen. Een dergelijke aanpak zou een trendbreuk beteken in vergelijking met de validatie van Waterstanden (Melissa) of Morfologie gegevens (Maria), waarbij steeds de relatie met de oorspronkelijke reeks behouden blijft.

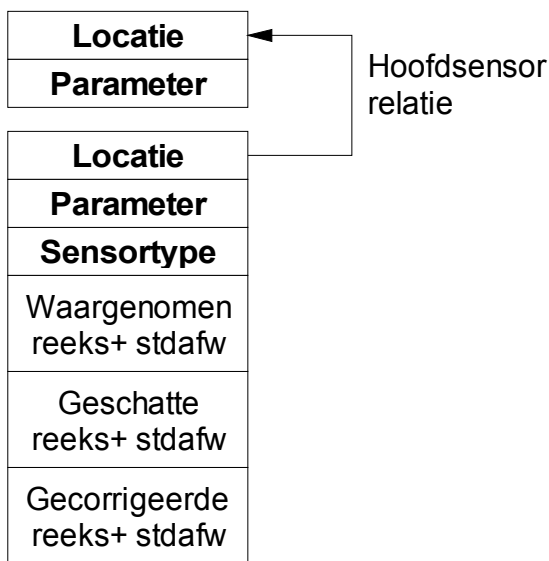
2.2 Gemaakte keuze

Als tussen oplossing wordt gekozen voor de opslagstructuur waarbij er per combinatie van locatie en parametersoort één slechts reeks wordt gevalideerd (de hoofdsensor), ook al zijn er meerdere reeksen aanwezig (nevensensors). De gegevens van nevensensors kunnen wel worden gebruikt om ontbrekende waarden van de hoofdsensors bij te schatten. Welke sensoren als hoofdsensor fungeren kan de gebruiker middels een handmatige keuze of een stuurfile aangeven. Verder geldt het volgende:

- Bij het visualiseren van de data worden alleen de als hoofdsensor aangeduide reeksen getoond;
- Nevensensors worden eventueel getoond, maar dan alleen als een ‘passieve reeks’ die verder niet te manipuleren is;

- Bij het markeren van bepaalde perioden als hiaat of outlier wordt alleen rekening gehouden met de hoofdsensors;
- Wanneer het validatieproces is afgerond, worden alleen de hoofdsensors weggeschreven naar DONAR;
- De inhoud van de nevensensors wordt niet aangepast binnen het validatieproces;
- De gegevens van nevensensors worden kunnen wel worden gebruikt om ontbrekende waarden van de hoofdsensors bij te schatten.

De gekozen situatie is weergegeven in Figuur 3, en is een tussen vorm van de situaties in Figuur 1 en Figuur 2. Enerzijds kunnen een onbeperkt aantal reeksen in het werkgebied worden ingelezen, ook als de bijbehorende combinatie locatie-parameter reeds aanwezig is. Anderzijds kan per combinatie van locatie en parameter hooguit één hoofdsensor worden aangegeven (dit is in het figuur aangegeven doordat de pijl een enkele punt heeft).



Figuur 3: *gekozen opslag structuur. Een werkgebied kan een onbeperkt aantal reeksen bevatten. Er kan echter per combinatie van locatie en parameter maar 1 hoofdsensor worden aangewezen. Alleen de hoofdsensor wordt gevisualiseerd en gevalideerd.*

2.3 Voorlopige inhoud van het werkgebied

Naarmate de functies van WAVIX worden uitgebreid zal het werkgebied worden uitgebreid met nieuwe velden.

Een reeks die uit DONAR wordt geïmporteerd heeft de volgende attributen:

- Een verzameling attributen die de tijd-as bepaald, zoals de begin en de eind periode en de tijdstap;
- De gemeten waarden;
- De bijbehorende status. Deze kan zijn:
 - 0: normaal waargenomen punt
 - 99: hiaat
 - 25: geïnterpoleerde of geschatte waarde

- Een aantal meta-informatie velden die aangeven om wat voor gegevens het gaat. De belangrijkste hiervan zijn:
 - W3H.sLoccod: de meetlocatie
 - W3H.sParcod: de waargenomen parameter
 - W3H.sVatcod: de gebruikte sensor

Tijdens de verwerking in WAVIX worden nog een aantal corresponderende attributen gegenereerd:







- Voorspelde waarden; Dit zijn geschatte golfhoogtes of golfperiodes. Voor dit doel worden binnen WAVIX een aantal schattingsmethoden geïmplementeerd.
- Standaardafwijking van de voorspelde waarden.

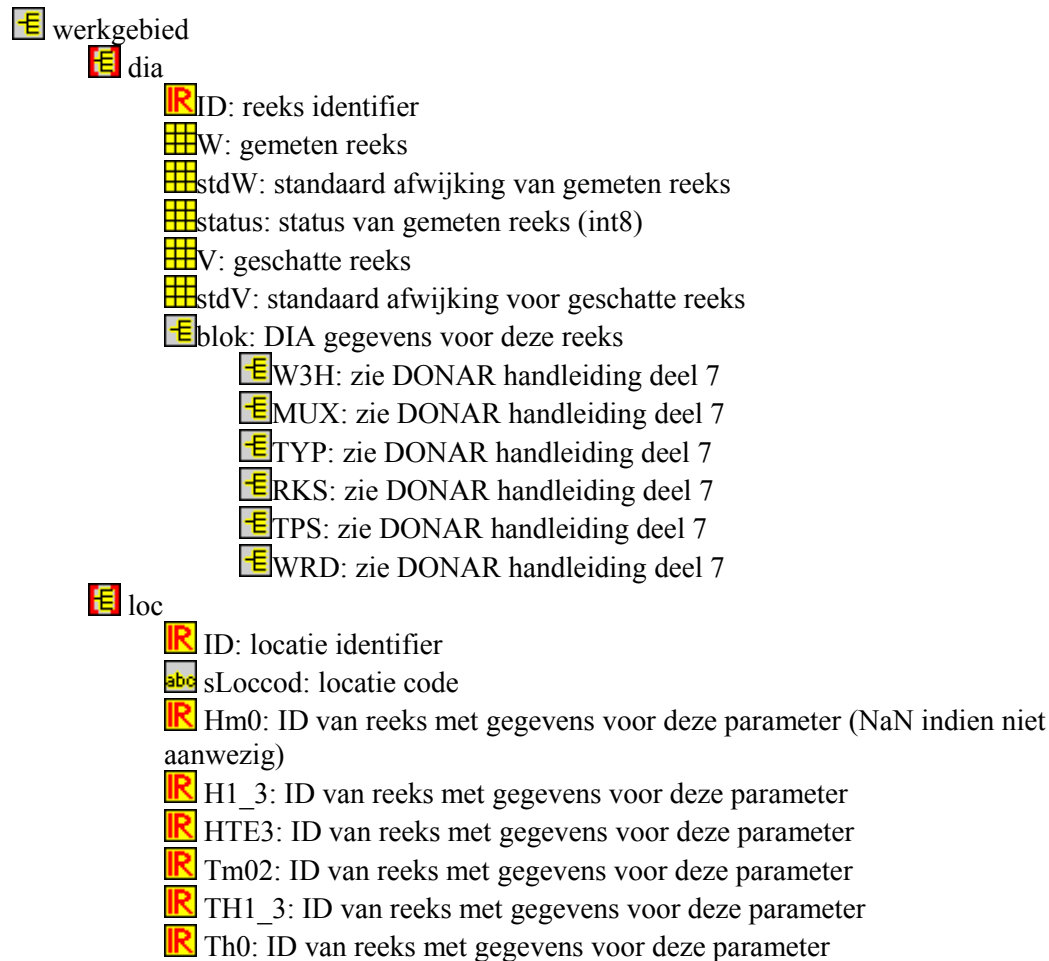
In de huidige implementatie van WAVIX (WAVIX 98) worden deze geschatte waarden en bijbehorende standaardafwijkingen niet opgeslagen, maar worden zij over de waargenomen datapunten heen geschreven. In de nieuwe WAVIX implementatie zal de mogelijkheid zijn om de geschatte waarden te visualiseren voordat ze worden toegepast.

Verder worden de reeksen die als hoofdsensor fungeren gekoppeld aan een locatie.

Dit alles leidt tot een datastructuur die binnen de programmatuur zal worden aangeduid als de werkgebied-structure. Deze is weergegeven in Figuur 4 op basis van een aantal conventies die zijn uitgelegd in Tabel 1. Zoals gezegd betreft het hier een moment opname. De datastructuur zal worden uitgebreid wanneer nieuwe functies worden toegevoegd.

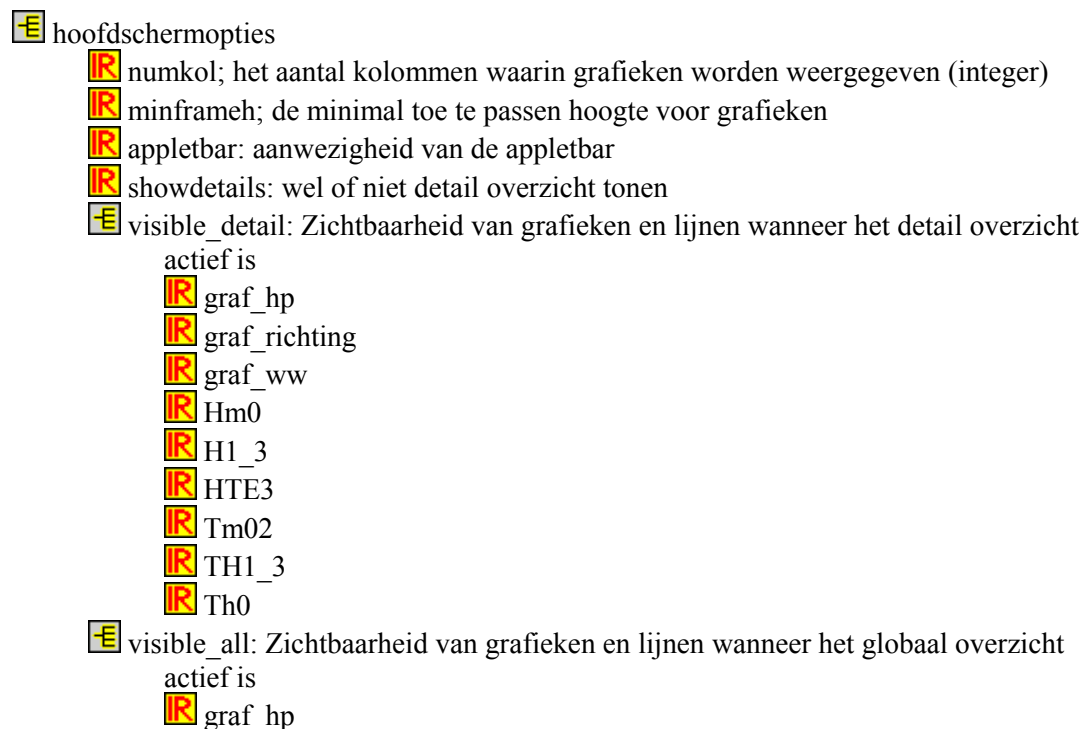
Tabel 1: *Weergave conventies voor Matlab datastructuur*
































Symbol	Betekenis
	Een enkele structure
	Een structure array met mogelijk meerdere elementen
	Een scalar. Opmerking: er wordt in de weergave nog geen onderscheid gemaakt tussen doubles en integers
	Een numeriek array met mogelijk meerdere elementen. Opmerking: er wordt in de weergave nog geen onderscheid gemaakt tussen doubles en integers
	Een karakter array met mogelijk meerdere elementen
 Zie ...	Een placeholder voor een meer gedetailleerde uitleg, die elders gegeven wordt



Figuur 4: *Werkgebied structure*

3 Inhoud optie structure voor het hoofdscherm



-  graf_richting
-  graf_ww
-  Hm0
-  H1_3
-  HTE3
-  Tm02
-  TH1_3
-  Th0
-  DefineHiaat: definieer een periode als hiaat wanneer een of meerdere van de volgende reeksen een hiaat is
 -  Hm0
 -  H1_3
 -  HTE3
 -  Tm02
 -  TH1_3
 -  Th0
-  DefineOutlier: definieer een periode als outlier wanneer een of meerdere van de volgende reeksen een outlier is
 -  Hm0
 -  H1_3
 -  HTE3
 -  Tm02
 -  TH1_3
 -  Th0
-  Factor: definieer de factor die moet worden toegepast bij het weergeven van de diverse lijnen
 -  Hm0
 -  H1_3
 -  HTE3
 -  Tm02
 -  TH1_3
 -  Th0
-  selectkolom: 0
-  kaartexport: 0

4 Module: definieer afhankelijkheden

Doel: het vastleggen van de afhankelijkheden tussen reeksen in een afhankelijkheids file.

Eindproduct: een definitie van afhankelijkheden, vastgelegd in een afhankelijkheidsfile

Voorbeeld van een afhankelijkheidsfile:

```
1, AUKPFRM, Hm0, WAVEC, 0
0, AUKPFRM, Hm0, WAVERDR, 0
1, AUKPFRM, HTE3, WAVEC, 0
0, EIELSGT, HTE3, WAVEC, 1
0, IJMDMNTSSPS, HTE3, WAVEC, -1
```

Uitleg over inhoud:

- kolom 1: bevat een 1 wanneer een afhankelijke variabele wordt aangeduid in deze regel
 bevat een 0 wanneer een onafhankelijke variabele wordt aangeduid
- kolom 2: bevat de locatie code W3H.sLoccod
- kolom 3: bevat de parameter code W3H.sParcod
- kolom 4: bevat de veldapparaat type code (W3H.sVatcod)
- kolom 5: bevat de tijdsverschuiving in uren

Implementatie: de afhankelijkheids file kan in een ASCII editor of EXCEL worden aangemaakt.

Op termijn dient ook een functie gerealiseerd te worden die de afhankelijkheids file automatisch gegenereerd.

5 Module: definieer condities

De hypothese is dat bij verschillende weer- en golfcondities condities verschillende modellen nodig zijn. De verschillende categorieën zullen in een zg conditie file worden gedefinieerd. Voorlopig wordt echter nog van de toepassing van een enkel model uitgegaan.

6 Module: train de neurale netwerken

Doel: het trainen van meerde neurale netwerken op basis van de gegevens in een werkgebied, een afhankelijkheidsfile en eventueel een conditie file.

Invoer:

- een of meerdere afhankelijkheidsfiles
- een of meerdere netwerkfiles
- een NN stuurfile
- een conditie file (toekomstige optie)
- de gegevens uit een netwerk file

Uitvoer:

een netwerkfile

Stappen:

- De afhankelijkheids file wordt ingelezen.
- De definities van het neural netwerk wordt ingelezen uit een stuurfile van de vorm

```
Keyword1 Waarde1  
Keyword2 Waarde2  
Keyword3 Waarde3
```

- Er wordt een lege uitvoerfile aangemaakt op basis van de invoer.
- Het is ook mogelijk om netwerk specificaties toe te voegen uit eerdere aangemaakte uitvoer files.
- Alle netwerken worden weergegeven inclusief hun status. De status kan zijn:
 - wacht op Kalibratie
 - bezig met kalibreren
 - gekalibreerd

- Het is mogelijk om de status van een gekalibreerd netwerk te wijzigen in ‘wacht op calibratie’
- Start het trainen. Alle netwerken worden nu een voor een getraind. Steeds wanneer een netwerk getraind is, wordt het resultaat opgeslagen. Het is mogelijk om het trainingsproces tussentijds af te breken. alleen de trainingsgegevens van het laatste netwerk gaan dan verloren.
- Het is mogelijk om bepaalde netwerken uit een set te verwijderen.

7 Module: visualiseer netwerk resultaten

Over te nemen uit Stormnet en vervolgens aan te passen

8 Module: pas NN toe

Doel: het aanmaken van voorspelde waarden en bijbehorende standaard afwijking op basis van een getraind netwerk

Invoer:

Een netwerk file

De gegevens uit een werkgebied

Stappen:

- haal de gegevens uit een dia blok en roep een rekenfunctie aan.
- bepaal in deze rekenfunctie de voorspelde waarde en de bijbehorende nauwkeurigheid

9 Module schat betrouwbaarheidsmarge bij ruwe metingen

Doel: het koppelen van een betrouwbaarheid in de vorm van een standaard afwijking aan de ruwe data

Invoer: een diablok (uit het veld sParcod wordt de parametersoort afgeleid)

Stappen: pas de vuistregels uit het WAVIX rapport toe

10 Module: schat parameters op basis van verhoudingsgetallen

Doel: het maken van schattingen met betrouwbaarheidsmarge op basis van een aantal vuistregels die in het WAVIX rapport zijn beschreven.

Invoer: alle reeksen uit het werkgebied.

Uitvoer: een geschatte waarde en een betrouwbaarheidsmarge voor alle reeksen in het werkgebied

Stappen: de manier van schatten is beschreven in het WAVIX rapport.