

**TIJDREEKSEN IN BODEM EN WATER**



**COMMISSIE VOOR HYDROLOGISCH ONDERZOEK TNO**

**RAPPORTEN EN NOTA'S No. 20**

## **TIJDREEKSEN IN BODEM EN WATER**

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, 's-GRAVENHAGE

Tijdreeksen

Tijdreeksen in bodem en water: lezingendag, 's-Gravenhage, 26 januari 1989/[P.K. Baggelaar ... et al.; samenst. J.C. Hooghart]. -

's-Gravenhage: Commissie voor Hydrologisch Onderzoek TNO. - Ill. - (Rapporten en Nota's /Commissie voor Hydrologisch Onderzoek TNO; no. 20)

Met lit. opg., reg.

ISBN 90-6743-135-4

SISO 631.2 UDC 556.3

Trefw.: hydrologie

De verantwoording voor de inhoud van de bijdragen ligt bij de auteurs.

COPYRIGHT © Nederlandse organisatie voor  
toegepast natuurwetenschappelijk onderzoek  
TNO, 1988

**TIJDREEKSEN IN BODEM EN WATER**



**RAPPORT MET:**  
Inleidingen van de lezingendag op 26 januari 1989  
Centraal Kantoor TNO te 's-Gravenhage

**COMMISSIE VOOR HYDROLOGISCH ONDERZOEK TNO – 's-Gravenhage – 1988**  
**RAPPORTEN EN NOTA'S No. 20**



# INHOUD

	pag.
VOORWOORD	1
AUTEURS	3
1 INLEIDING F.C. van Geer	5
2 PRAKTISCHE METHODEN VOOR TRENDANALYSE IN HET WATERBEHEER J.R. Moll	9
Samenvatting	9
1 Inleiding	9
2 Trenddetectie	11
2.1 Staptrend	11
2.2 Polynomiale trend	13
3 Trendbeschrijving	16
4 Meetnetoptimalisatie	17
5 Conclusies	20
Literatuur	20
3 BOX-JENKINS ANALYSE VAN STIJGHOOGTEN P.K. Baggelaar	21
Samenvatting	21
1 Inleiding	21
2 Modelleren dynamische systemen	22
3 De stijghoogte als uitgangssignaal	23
4 De Box-Jenkins methode	24
4.1 Inleiding	24
4.2 Mogelijke benaderingen	25
4.3 De transfer functie	27
4.4 Het Arima model	29
5 Het bouwen van een Box-Jenkins model	30
5.1 Algemeen	30
5.2 Identificatie transfer functie	31
5.3 Identificatie transfer-/ruis model	33
5.4 Schatting parameters transfer-/ruis model	35
5.5 Verificatie transfer-/ruis model	36
6 Toepassing	36
7 Enkele praktische slotopmerkingen	37
Literatuur	38
4 TIJDREKSANALYSE VAN VERANDERINGEN IN CHEMISCHE SAMENSTELLING VAN DE BODEMOPLOSSING IN DE STUDIELOCATIE HACKFORT GEDURENDE 6 JAAR A. Stein, T. Feijtel, T. Pape en H. Booltink	39
Abstract	39
1 Inleiding	39
2 Studielocatie	40
3 Statistische analyses	41
3.1 De tijdreeks	41
3.2 Ruimte-tijd modellering	43
4 Resultaten en discussie	47
4.1 Aannames en keuzes	47
4.2 Resultaten	48

4.3 Discussie	53
Acknowledgement	54
Literatuur	55
Appendix	57
5 SPECTRAALANALYSE	59
A.W. Heemink	
Samenvatting	59
1 Inleiding	60
2 Stochastische processen	60
3 Tijdreeksanalyse	62
4 Spectraalanalyse	65
5 Slotopmerkingen	69
Literatuur	70
6 DE ANALYSE VAN WATERSTANDSMETINGEN IN DE OOSTERSCHELDE; TOEPASSING VAN SPECTRAAL-ANALYSE BIJ DE OOSTERSCHELDEWERKEN	71
A. Langerak en J.J. Leendertse	
Samenvatting	71
1 De Oosterscheldewerken en de modeltechniek	71
2 De fysica van het getij	75
3 Toepassing spectraal-analyse bij voorbereidend getijonderzoek	76
4 Toepassing spectraal-analyse bij randvoorwaarden modelrand	80
5 Spectraal-analyse en afregeling model	83
6 Resumé en slotopmerkingen	85
7 STATISTISCHE ANALYSE MET BETREKKING TOT MAXIMA	87
M.A.J. van Montfort en J.V. Witter	
Samenvatting	87
1 Inleiding	88
2 Veiligheid en herhalingstijd	93
3 Nederlandse neerslagmaxima	96
4 Extractiemethoden en bijbehorende modellen	99
4.1 Jaarmaxima (top-1)	99
4.2 Top-r	100
4.3 Maandmaxima	100
4.4 Overschrijding van een drempelwaarde	101
5 Schattingsmethoden van de modelparameters	102
5.1 Jaarmaxima (top-1)	105
5.2 Top-r	110
5.3 Maandmaxima	111
5.4 Peaks over threshold model	111
6 Schatten van de T-jaar-herhalingswaarde	113
6.1 Bij top-1 data	113
6.2 POT data	117
7 Conclusies en aanbevelingen	118
Dankbetuiging	119
Literatuur	119

## VOORWOORD

Sinds ruim één jaar bestaat er een nauwe samenwerking tussen de NRLO werkgroep "Ruimtelijke variabiliteit in Bodem en Water" en de CHO Studiegroep Statistiek in de Hydrologie. Onder andere organiseren beide groepen voor hun leden gezamenlijke lezingenmiddagen. Tevens werd zinvol geacht om voor een wat breder samengesteld publiek een lezingendag te organiseren over het thema "Tijdreeksen in bodem en water". Op vele plaatsen in het land wordt immers met tijdreeksen gewerkt en hierbij doen zich vele vragen voor. Tevens werd door de voorbereidingsgroep besloten om de behandelde onderwerpen in boekvorm uit te geven, hetgeen geleid heeft tot het onderhavige boekje. Hopelijk geeft deze publikatie een beter inzicht in wat met tijdreeksen kan worden gedaan en wat de beperkingen zijn.

's-Gravenhage, december 1988





## AUTEURS

P.K. Baggelaar	Keuringsinstituut voor Waterleidingartikelen KIWA N.V., Nieuwegein
H. Booltink	LUW, Vakgroep Bodemkunde en Geologie, Wageningen
T. Feijtel	LUW, Vakgroep Bodemkunde en Geologie, Wageningen
F.C. van Geer	Dienst Grondwaterverkenning TNO, Delft
A.W. Heemink	Rijkswaterstaat, Dienst Informatieverwerking, 's-Gravenhage
A. Langerak	Rijkswaterstaat, Dienst Getijdewateren, Middelburg
J.J. Leendertse	Rand Corporation, Santa Monica, California, U.S.A.
J.R. Moll	Waterloopkundig Laboratorium, Marknesse
M.A.J. van Montfort	LUW, Vakgroep Wiskunde, Wageningen
T. Pape	LUW, Vakgroep Bodemkunde en Geologie, Wageningen
A. Stein	LUW, Vakgroep Bodemkunde en Geologie, Wageningen
J.V. Witter	UvA, Vakgroep Fysisch Geografie en Bodemkunde, Amsterdam



## INLEIDING

F.C. van Geer

In dit boekje "tijdreeksen in bodem en water" zijn zes bijdragen verzameld rond het onderwerp tijdreeksen. Een logische vraag die gesteld kan worden is:

Waarom een boekje over tijdreeksen?

Het antwoord op die vraag is erg simpel. Er is in de hydrologie en de bodemkunde, ondanks het feit dat bij vrijwel elke studie gesteld wordt dat er (te) weinig gegevens zijn, een gigantische hoeveelheid gegevens verzameld. Veel van die gegevens hebben de vorm van tijdreeksen (neerslaghoeveelheden, grondwaterstanden, oppervlaktewaterstanden enz.). Kortom een boekje over de analysemogelijkheden van tijdreeksen is zinvol, omdat er in ons vakgebied een groot aantal tijdreeksen bestaat.

Een tweede logische vraag is:

Waarom dit boekje op dit moment?

Behalve de constatering dat er zeer veel gegevens als tijdreeksen beschikbaar zijn, kan ook worden gesteld dat de beschikbare informatie meestal niet optimaal wordt gebruikt. Dit vindt zijn oorzaak in het feit dat tot voor kort veel tijdreeksen moeilijk toegankelijk

waren voor routinematige analyses. De laatste jaren zijn, met de opkomst van de automatisering, veel tijdreeksen direct beschikbaar gekomen via grotere (mainframe) of kleinere (micro's) data bases. Bovendien heeft diezelfde automatisering ertoe bijgedragen dat (veelal statistische) technieken die in principe wel eerder bekend waren, zijn uit-ontwikkeld tot praktisch toepasbare gereedschappen. Daarom leek het ons (de Studiegroep Statistiek in de Hydrologie van CHO-TNO en de NRLO werkgroep "Ruimtelijke variabiliteit in bodem en water") zinvol analysetechnieken ten aanzien van tijdreeksen te presenteren, die op dit moment binnen ons vakgebied operationeel zijn.

Dit boekje is techniek-georiënteerd. Dat wil zeggen dat in elk van de bijdragen een bepaalde techniek centraal staat en wordt toegelicht met een voorbeeld. De bijdragen van Heemink en van Langerak en Leendertse zijn als zodanig als een twee-eenheid op te vatten. De technieken die in de bijdragen worden behandeld, zijn niet alleen geschikt voor de betreffende toepassing. Integendeel, de technieken zijn stuk voor stuk toepasbaar op een breed scala van problemen. Het wordt echter aan de lezer(es) overgelaten de aangedragen technieken te combineren met toepassingsmogelijkheden in het hem/haar bekende vakgebied.

Bij het samenstellen van dit boekje is uitgegaan van technieken die op dit moment in de hydrologie en de bodemkunde operationeel zijn. Een consequentie daarvan is dat de behandelde technieken geen compleet beeld geven van alle mogelijkheden van het analyseren van tijdreeksen. De volgende vijf technieken worden behandeld.

#### Trendanalyse

Een belangrijke reden om gedurende langere tijd een variabele te meten is het vaststellen van een verandering in de tijd of een controle op het uitblijven van een verandering. Het beschrijven van "blijvende" veranderingen in een tijdreeks is het doel van trend-

analyse. Moll illustreert een aantal mogelijkheden van trendanalyse aan de hand van toepassingen in het oppervlaktewater.

#### Box-Jenkins modellen

De Box-Jenkins modellen worden meestal niet gebruikt voor het beschrijven van een bepaald fenomeen in een tijdreeks, zoals de trendanalyse, maar veel meer voor de predictie van te modelleren variabelen en voor het leggen van verbanden tussen invoer- en uitvoerreeksen (oorzaak en gevolg). Een aantal mogelijkheden voor het toepassen van Box-Jenkins modellen op grondwaterstandsreeksen wordt door Bagelaar behandeld.

#### Space-time modellen

Gemeten tijdreeksen van een variabele geven per definitie informatie over die variabele op een bepaalde lokatie. Vaak is er tussen verschillende waarnemingsreeksen een ruimtelijk verband. Met space-time modellen wordt getracht gebruik te maken van dat verband waardoor ook conclusies in ruimtelijke zin getrokken kunnen worden. Stein et al. geven een voorbeeld van een space-time modellen in de bodemkunde.

#### Spectraalanalyse

De voorgaande onderwerpen hebben betrekking op het beschrijven en modelleren van tijdreeksen in het tijdsdomein, hetgeen goed aansluit op het denken in de reële tijd. Complementair hieraan is de spectraalanalyse. Dit is een benadering in het frequentiedomein, waarbij een tijdsafhankelijke variabele wordt beschreven door frequenties en amplituden. De spectraalanalyse, tradioneel veel toegepast voor het beschrijven van periodieke verschijnselen, is ook geschikt voor het vastleggen van verbanden tussen verschillende tijdreeksen. Heemink geeft een introductie in de techniek van de spectraalanalyse, waarna Langerak en Leendertse een toepassing bij getijdewater behandelen.

## Extreme waarden

Een aparte plaats in het rijtje van technieken is de beschrijving en modellering van extreme waarden. De meeste andere technieken zijn er op gericht om "zoveel mogelijk" van een tijdreeks als geheel te beschrijven. Dit heeft vaak als consequentie dat de beschrijving niet toegesneden is op uitzonderlijk hoge of lage waarden. Bij modellering van extreme waarden ligt het accent juist op het goed beschrijven van deze uitzonderlijke waarden. Daarbij spelen specifieke extreme waardenverdelingen een grote rol, zoals door Witter van Van Montfort wordt beschreven aan de hand van gegevens uit de meteorologie.

Uiteraard kunnen de bovenstaande technieken, binnen het kader van dit boekje, niet tot in alle details worden beschreven. Het mag echter, na het lezen van de bijdragen, duidelijk zijn dat er in Nederland over een breed front (zie de lijst van auteurs) kennis aanwezig is over het analyseren van tijdreeksen in de hydrologie en de bodemkunde. Hopelijk kan dit boekje ertoe bijdragen dat er in de toekomst meer gebruik gemaakt wordt van de informatie in de vele beschikbare tijdreeksen dan tot nu toe het geval is.

J.R. Moll

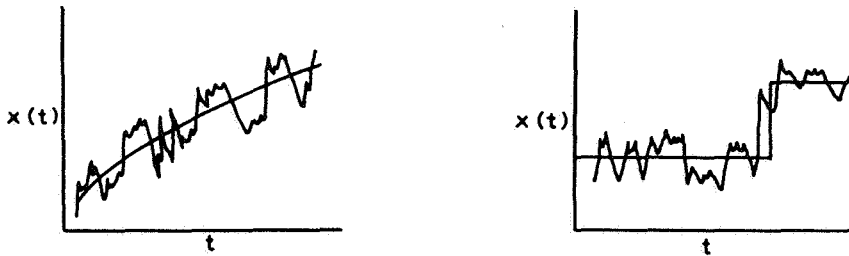
## SAMENVATTING

Het ontdekken en volgen van trendmatige ontwikkelingen is een belangrijk doel van veel meetinspanningen in het waterbeheer. Ingegaan wordt op verschillende vormen van trend (staptrend, lineaire trend), en op de relatie met de relevante tijdschaal. Voor het detecteren van trends in een meetreeks van gegeven lengte zijn een aantal klassieke toetsen beschikbaar: t-toets, Wilcoxon, likelihood ratio test, tekentoets, Kendall's toets. Deze toetsen worden kort beschreven en m.b.v. eenvoudige voorbeelden geïllustreerd. In verband met de verdere verwerking van de meetgegevens is het verwijderen van trends soms van belang. Een aantal technieken hiervoor (gemiddelde helling, polynomen) wordt gepresenteerd. Het optimaliseren van een meetset voor trenddetectie wordt toegelicht aan de hand van voorbeelden uit o.a. het waterkwaliteitsbeheer.

## 1 INLEIDING

Eén van de belangrijkste doelstellingen van veel meetinspanningen in het waterbeheer is conditiebewaking; voorbeelden zijn: het letten op verhoging of verlaging van de grondwaterstand, het controleren van de kwaliteit van drinkwater, het opsporen van verzakkingen in dijklichamen, kunstwerken of rivierbodems, en het nauwgezet volgen van de stijging van de zeespiegel. De meetgegevens die ten behoeve van deze conditiebewaking vergaard worden zijn beschikbaar in de vorm van tijdreeksen. Stapsgewijs of geleidelijke toe- of afnames in het niveau van een gemeten variabele noemt men trends, zie Figuur 1.



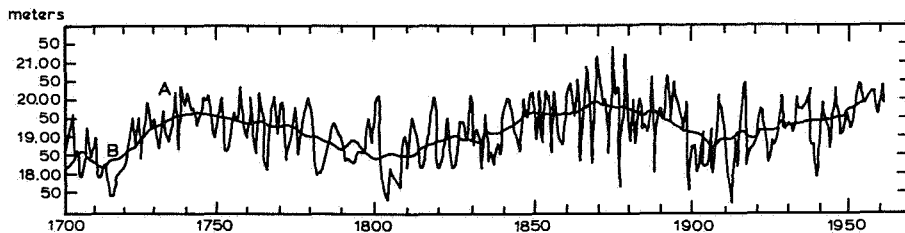


Figuur 1 Voorbeeld van een staptrend en een polynomiale trend

Trends zijn van primaire interesse bij de conditiebewaking. Ook bij andere aspecten van waterbeheer kunnen trends een belangrijke rol spelen. Een voorbeeld is het ontwerp van reservoirs: de benodigde capaciteit kan worden afgeleid uit historische reeksen van toe- en afvoer, en een prognose over de ontwikkeling van gesignaleerde trends.

Het toepassen van technieken voor trendanalyse kan vaak in belangrijke mate ondersteund worden door intuïtie. Voor een langs mathematische weg aangetoonde trend zoekt men voorts graag een fysische verklaring. Een voorbeeld is een staptrend, die in een tijdreeks kan zijn ontstaan bij het wijzigen van de meetmethode. Een te groot gewicht toekennen aan intuïtie kan echter ook merkwaardige resultaten opleveren. Een bekend voorbeeld is het "wegtoetsen" van afwijkingen van het intuïtief juiste beeld door een conservatieve nulhypothese te kiezen.

Bij het analyseren van trends is de beschouwde tijdshorizon van groot belang. De opkomende waterstand tussen eb en vloed is op een tijdshorizon van 12 uur als trend op te vatten; op een tijdshorizon van enkele dagen wordt de waterstandstijging als onderdeel van een periodiek getijdencyclus gezien. Iets dergelijks doet zich voor bij waterstanden langs de Nijl, zie Figuur 2. Op verschillende momenten kan men hier staprends of lineaire trends identificeren. Over de lengte van de gehele reeks gezien is niet vast te stellen dat de Nijl aan het opdrogen is.



Figuur 2 Waterstanden langs de Nijl, 1700-1960

In dit artikel worden verschillende praktische methoden voor het uitvoeren van trendanalyse kort besproken. Op mathematische achtergronden wordt niet in detail ingegaan en doorverwezen naar de literatuur (Kendall en Stuart, 1977). De indeling van dit artikel is als volgt: hoofdstuk 2 bevat technieken voor trenddetectie, en hoofdstuk 3 voor trendbeschrijving; hoofdstuk 4 gaat in op meetnetoptimalisatie in relatie tot trenddetectie, en hoofdstuk 5 bevat een aantal conclusies.

## 2 TRENDDETECTIE

### 2.1 Staptrend

In beginsel kan men hier twee situaties onderscheiden, afhankelijk van het feit of de plaats van de mogelijke staptrend bekend is of niet. In het eerste geval is de vraag: gegeven de tijdreeks  $x_1, \dots, x_m, \dots, x_n$ , is er een staptrend (of sprong) aantoonbaar tussen  $x_m$  en  $x_{m+1}$ ?

De reeks is met het volgende model te beschrijven:

$$x_i = \mu + \epsilon_i \quad i = 1, \dots, m \quad (1)$$

$$= \mu + \Delta + \epsilon_i \quad i = m+1, \dots, n$$

Hierin zijn  $\epsilon_i$  onderling onafhankelijke normaal verdeelde variabelen met verwachting  $a$  en met variantie  $\sigma^2$ ;  $\mu$  en  $\Delta$  representeren de onbekende verwachting van de reeks. De gestelde vraag komt neer op het toetsingsprobleem met

$$H_0: \Delta = 0$$

$$H_1: \Delta \neq 0$$

Onder  $H_0$  heeft de volgende toetsingsgrootheid  $\underline{t}$  een student's t-verdeling met  $(n-2)$  vrijheidsgraden:

$$\underline{t} = \left( \frac{1}{m} \sum_{i=1}^m x_i - \frac{1}{n-m} \sum_{i=m+1}^n x_i \right) / \left( \sqrt{\frac{1}{m} + \frac{1}{(n-m)}} \cdot \underline{s} \right) \quad (2)$$

Hierin is  $\underline{s}^2$  de steekproefvariantie van de reeks. Met een tabel van de t-verdeling is bij een gekozen betrouwbaarheid  $\alpha$  de toets uit te voeren.

Indien geen normaliteit mag worden verondersteld is de verdelingsvrije rangorde-toets van Wilcoxon te gebruiken. De procedure hiervoor is:

- i) elke meetwaarde krijgt een rangnummer  $r(x_i)$  in opklimmende volgorde  
 ii) bereken  $\underline{W} = \sum_{i=1}^m r(x_i)$  (3)

- iii) De kansverdeling voor  $\underline{W}$  onder  $H_0$  wordt dan benaderd door een standaard normale verdeling  $\phi$ :

$$P(\underline{W} > W; H_0) \approx 1 - \phi \left[ (W - \frac{1}{2}m(n+1)) / \sqrt{\frac{1}{12} m(n-m)(n+1)} \right] \quad (4)$$

Indien niet van te voren aangegeven kan worden op welk punt in de reeks een staptrend verondersteld wordt, is met de volgende procedure een maximum-likelihood schatting voor dit punt te geven in het geval de reeks normaal verdeeld verondersteld mag worden.

- i) Bereken  $S_k = \sum_{i=1}^k (x_i - \bar{x})$   $k = 1, \dots, n$  (5)

- ii) Bereken  $D_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  (6)

- iii) Bereken  $V = \max_{1 \leq k \leq n-1} \{ |S_k| / [D_x \{k(n-k)\}^{\frac{1}{2}}] \}$  (7)

- iv) De ML-schatting voor het bewuste punt  $m$  is die  $k$  waarvoor  $V$  maximaal is.

Met de zo berekende  $m$  is nu via de al beschreven  $t$ -toets de trend-detectie uit te voeren op de grootte van de staptrend. Hiertoe is echter ook rechtstreeks gebruik te maken van de berekende  $V$ , zie Buishand (1984). In dat artikel wordt hiervoor tevens nog een alternatieve steekproefgrootte,  $U$ , gepresenteerd.

Een toepassing in het operationeel waterkwaliteitsbeheer vormt de volgende situatie: een stochastisch fluktuerende waterkwaliteitsparameter kan onvoorzien een blijvende sprong omhoog maken. De beheerder is niet alleen geïnteresseerd in overschrijding van een absoluut nivo (calamiteit) maar ook in staprends.

De detector voor een staptrend ter grootte  $\lambda$  kan als volgt beschreven worden (Basseville, 1988):

i) Neem  $m_n = 0$

ii) Bereken  $U_n = \sum_{k=1}^n (x_k - \mu - \frac{\lambda}{2})$   $n=1, \dots$  (8)

iii) Bereken  $m_n = \min (m_{n-1}, U_n)$  (9)

iv) Alarm als  $U_n - m_n > \lambda$  (10)

## 2.2 Polynomiale trend

Voor het ontdekken van geleidelijke schommelingen in het gemiddelde nivo van een tijdreeks zijn een aantal methoden beschikbaar. Aan de orde komen de tekentoets Kendall's tau, Spearman's rangcorrelatie en een regressiemethode.

Een eenvoudige toets op trend is de tekentoets. Het aantal positieve verschillen  $\underline{p}$  tussen opeenvolgende waarden in een reeks wordt geteld. Voor de nulhypothese (geen trend) geldt dat bij reekslengte  $n$

$$E \{ \underline{p} \} = \frac{1}{2}(n-1) \quad (11)$$

$$\text{var} \{ \underline{p} \} = \frac{1}{12} (n+1) \quad (12)$$

Voor grote  $n$  is  $\underline{p}$  bij benadering normaal verdeeld. Een bezwaar van deze toets is het geringe onderscheidend vermogen ervan.

De tau-test van Kendall verloopt volgens de volgende procedure:

Tel in de reeks  $x_1, \dots, x_n$  het aantal keren  $p$  dat in alle combinaties  $(x_i, x_j; j > i)$   $x_j$  groter is dan  $x_i$ . Het aantal paren  $(x_i, x_j; j > i)$  bedraagt  $\frac{1}{2}n(n-1)$ . In geval van positieve trend zal  $p$  dus in de buurt van  $\frac{1}{2}n(n-1)$  liggen, in geval van negatieve trend in de buurt van 0.

De steekproefgrootte  $\tau$ , gedefinieerd door

$$\underline{\tau} = 4p/(n-1)n - 1 \quad (13)$$

heeft als verwachting 0 en als variantie  $2(2n+5)/9n(n-1)$ .

Voor grote  $n$  nadert  $\tau$  snel tot een normale verdeling.

De test is hiermee:

$$(\text{geen trend}) H_0: \phi(|\tau|/\sqrt{\text{var } \tau}) < \alpha/2$$

$$(\text{trend}) H_1: \phi(|\tau|/\sqrt{\text{var } \tau}) > \alpha/2$$

Het is mogelijk om de gevonden positieve verschillen bij de uitwerking van Kendall's toets te wegen met de grootte van het verschil. Dus:

$$(\text{Kendall}) p = \sum_{i < j} h_{ij} \quad (14)$$

$$\begin{aligned} \text{met } h_{ij} &= 1 \text{ als } x_i > x_j \\ h_{ij} &= 0 \text{ als } x_i < x_j \end{aligned}$$

$$(\text{variant}) Q = \sum_{i < j} h_{ij}(i-j) \quad (15)$$

Het blijkt dat deze toetsingsgrootte te herleiden valt tot Spearman's rangcorrelatie - coefficient  $r$ :

$$\underline{r} = 1 - 12Q/n(n^2-1) = 1 - \frac{6}{n(n^2-1)} \sum_{t=1}^n (x_t - t)^2 \quad (16)$$

De toetsingsgrootheden  $\tau$  en  $r$  zijn zeer sterk gecorreleerd ( $\rho > 0.98$ ).

Het toetsen op trend is ook mogelijk d.m.v. lineaire regressie. Hierbij wordt verondersteld dat de reeks al dan niet een lineaire trend bezit en de volgende opbouw heeft:

$$x_i = \beta + \alpha \cdot i + \xi_i \quad (17)$$

met  $\{\xi_i\}$  een stationaire normaal verdeelde witte ruis reeks.

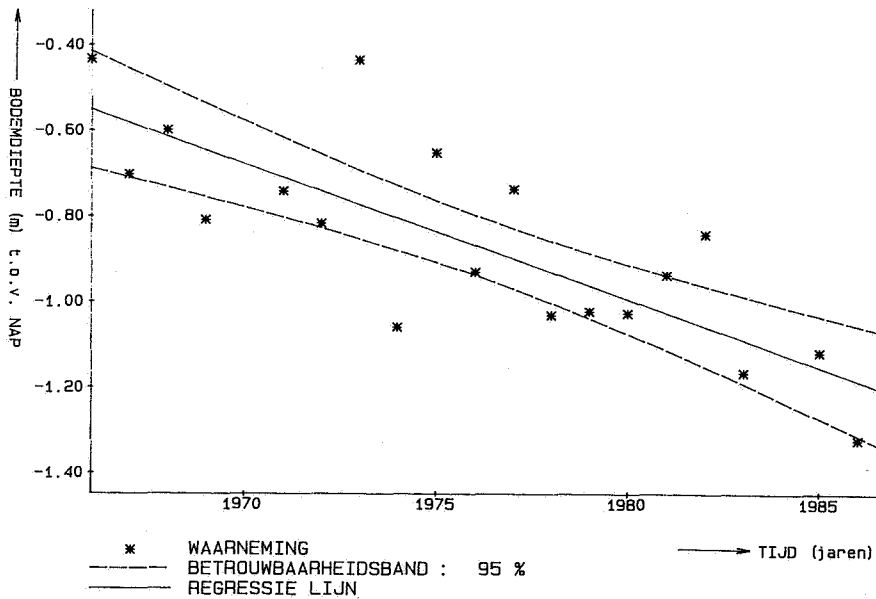
De procedure bestaat uit het schatten van  $\alpha$  en de variantie van  $\alpha$ , aan te duiden met  $\hat{\alpha}$  en  $s_\alpha^2$

De steekproefgrootheid  $t = \hat{\alpha} / s_\alpha$  is verdeeld als een Student's t-variabele, en via de t-tabel is de toets uit te voeren.

Een voorbeeld van toepassing van trenddetectie is gegeven in Figuur 3. In de figuur zijn uitgezet de gemiddelde bodemligging in de rivier de Waal tussen kilometerpaal 917 en 918 over de jaren 1966-1987.

Toepassing van Kendall's tau-test resulteerde in een negatieve trend. Deze is gevisualiseerd met een regressielijn. De betrouwbaarheidsintervallen hebben betrekking op de positie van de regressielijn.

RAAI : 917.000 - 917.875



Figuur 3 Gemiddelde bodemligging in de Waal tussen km 917 en 918.

### 3 TRENDBESCHRIJVING

Voor het beschrijven en hierna eventueel verwijderen van een gedetecteerde trend in een reeks staan verschillende mogelijkheden ter beschikking.

Een eenvoudige procedure is de "average slope" methode. Een reeks

$$\{x_i | i=1, \dots, n\} \text{ wordt beschreven door } x_i = \bar{x} + \bar{\alpha} (t - \frac{1}{2}n) + u_i$$

Hierin is  $\bar{x}$  de gemiddelde waarde, en  $\bar{\alpha}$  de gemiddelde helling van de reeks. De reeks  $\{u_i\}$  heeft gemiddelde nul en geen trend, en vormt dus de op trend gecorrigeerde reeks.

De gemiddelde helling  $\bar{\alpha}$  wordt geschat door

$$\bar{\alpha} = \frac{1}{v(n-v)} \left[ \sum_{i=n-v}^n x_i - \sum_{i=1}^v x_i \right] \quad (18)$$

met  $v$  = het grootste gehele getal kleiner of gelijk aan  $n/3$ .

Een gebruikelijke methode voor trendbeschrijving is het fitten van een polynoom van lage orde aan de meetreeks m.b.v. een kleinste kwadraten methode. De trend wordt beschreven door:

$$\hat{x}(t) = \sum_{k=0}^K b_k i^k \quad i=1, \dots, n \quad (19)$$

De coëfficiënten  $b_k$  worden zo gekozen dat  $J$  geminimaliseerd wordt:

$$J = \sum_{i=1}^n [x_i - \hat{x}(t)]^2 = \sum_{i=1}^n [x_i - \sum_{k=0}^K b_k i^k]^2 \quad (20)$$

$$\text{In het optimum geldt: } \frac{\partial J}{\partial b_k} = 0 \quad k=0, \dots, K \quad (21)$$

Uit deze  $(K+1)$  lineaire vergelijkingen zijn de coëfficiënten  $b_k$  op te lossen.

Als  $K = 0$  ontstaat één vergelijking, de gemiddelde lijn:

$$b_0 = \frac{1}{n} \sum_{i=1}^n x_i \quad (22)$$

Als  $K = 1$  ontstaan er twee vergelijkingen, met als oplossing voor de lineaire trend:

$$b_0 = [2(2n+1) \sum_{i=1}^n x_i - 6 \sum_{i=1}^n ix_i] / [n(n-1)] \quad (23)$$

$$b_1 = [12 \sum_{i=1}^n x_i - 6(n+1) \sum_{i=1}^n ix_i] / [n(n-1)(n+1)] \quad (24)$$

Voor trendbeschrijving wordt ook vaak gebruik gemaakt van lineaire regressierekening. Een probleem hierbij kan zijn dat de op trend gecorrigeerde reeks autocorrelatie bevat. Een belangrijke aanname bij regressierekening is dat de residuen onderling onafhankelijk zijn, m.a.w. dat de residureeks geen autocorrelatie bevat. Past men toch lineaire regressierekening toe, waarbij de coëfficiënten geschat worden met een kleinste kwadratenmethode, dan zijn deze schattingen wel zuiver, maar zijn de varianties ervan erg groot. Het toepassen van standaardformules voor de varianties leidt tot onderschattingen.

#### 4 MEETNETOPTIMALISATIE

Gelet op de kosten van meetinspanningen is meetoptimalisatie een onderwerp voor blijvende interesse. Optimaliseren van meetinspanningen gericht op trenddetectie komt neer op het kiezen voor het bemonsteringsinterval. Een zeer klein bemonsteringsinterval, ofwel een zeer hoge bemonsteringsfrequentie, is vaak niet aantrekkelijk gelet op de hiervoor noodzakelijke kosten of gelet op de overweging dat het bemonsterde proces binnen dit interval nauwelijks verandert. Een groot bemonsteringsinterval, ofwel een lage bemonsteringsfrequentie, kan onaantrekkelijk zijn indien zich binnen een interval niet waargenomen, maar wel van belang zijnde fluktuaties kunnen voordoen. Een optimum dient gezocht te worden.

In Schilperoort (1986) is uitgewerkt hoe de kans een trend met grootte  $Tr$  te ontdekken afhangt van de intrinsieke variatie van het proces en de autocorrelatieschaal. De intrinsieke variatie wordt meestal gerepresenteerd door de procesvariantie  $\sigma_x^2$ . De autocorrelatieschaal is (intuïtief) die (tijds-)

afstand tussen meetwaarden waarop de autocorrelatie net verdwenen is. Schilperoort geeft de volgende uitdrukking van de detectiekans  $P$  voor een trend ter grootte  $Tr$  bij een meetduur  $T$  en de bemonsteringsinterval  $\Delta$ :

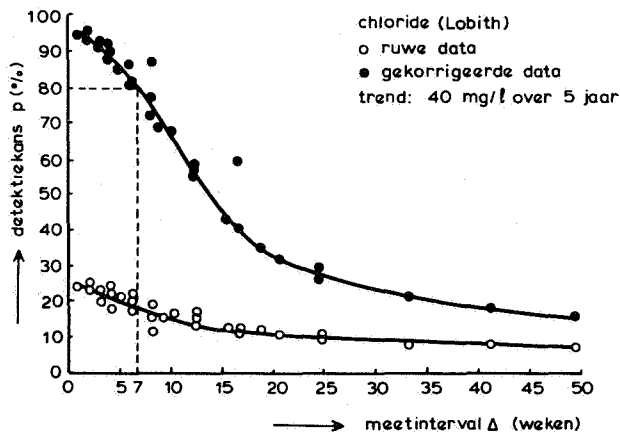


$$P(\Delta, Tr, T) = \phi \left\{ \frac{1}{c} \frac{Tr}{\sigma_x} \sqrt{N^*(\Delta, T)} - \xi\left(\frac{\alpha}{2}\right) \right\} \quad (25)$$

Hierin is  $\phi$  de standaard normale verdeling en  $\xi\left(\frac{\alpha}{2}\right)$  het percentiepunt bij een  $\alpha/2$ - overschrijdingskans. De constante  $c$  is afhankelijk van de vorm van de trend, en  $N^*$  staat voor het effectief aantal waarnemingen in meetperiode  $T$ .

Deze methode is toegepast op de chloridemetingen bij Lobith. Geëist werd dat met een kans  $P$  van 80% trends gedetecteerd moesten kunnen worden met een grootte  $Tr$  van 20% van de gemiddelde waarde over een periode  $T$  van 5 jaar. Bij uitwerking bleek dat de autocorrelatie van de originele chloride data zo hoog was dat de gestelde eis niet haalbaar was.

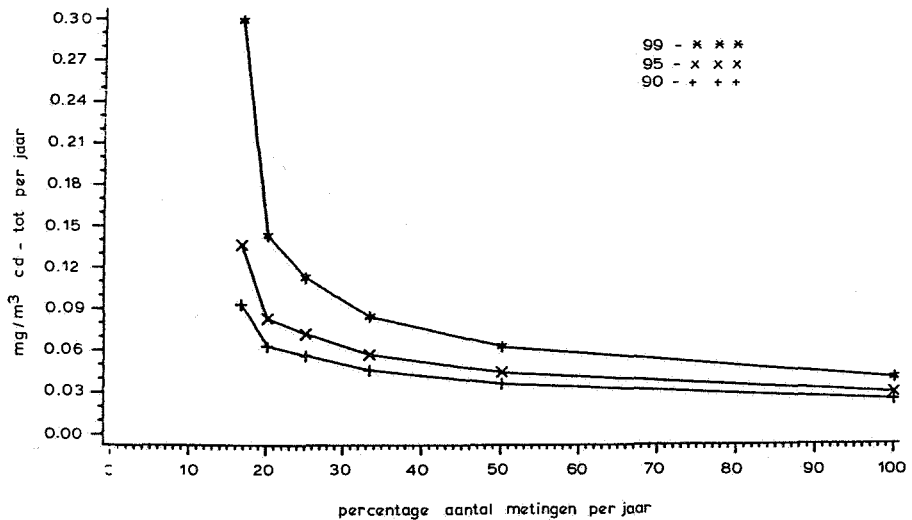
Een belangrijk deel van de autocorrelatie bleek echter samen te hangen met de afvoervariaties. Door op dit effect te corrigeren ontstond een reeks met een aanzienlijk kleinere correlatieschaal. Hiervoor bleek een bemonsteringsinterval van 7 weken voldoende, zie Figuur 4.



Figuur 4 Trenddetecteerbaarheid voor chloride bij Lobith

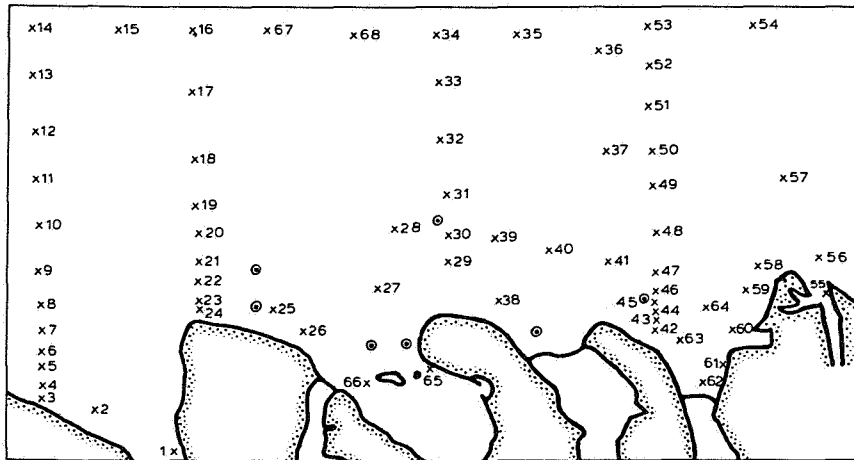
Een in de praktijk vaak gehanteerde methode voor meetnetoptimalisatie bestaat uit het "uitdunnen" van het gegevensbestand. Dit komt er op neer dat onderzocht wordt bij welke verdunningsfactor de statistische uitspraken die op het oorspronkelijke gegevensmateriaal gebaseerd zijn, te handhaven zijn na verdunning hiervan. Een voorbeeld hiervan is gegeven in Figuur 5. De detectiekans voor een trend van  $y$  mg/m<sup>3</sup> cadmium per jaar

is uitgezet tegen de verdunningsfactor voor een drietal betrouwbaarheden. De oorspronkelijke verzameling gegevens bestaat uit 2-wekelijkse waarnemingen van de concentratie cadmium bij meetpost Noordwijk.



Figuur 5 Trenddetecteerbaarheid voor cadmium bij Noordwijk

Een voorbeeld van een onderzoek waarbij de kans op aanwezigheid van een staptrend groot geacht werd is een waterkwaliteitsonderzoek in de zuidelijke Noordzee. In Figuur 6 zijn de lokaties gegeven waar vierwekelijks met een helikopter een groot aantal parameters bemonsterd zijn. Tevens aangegeven zijn de lokaties waar met vaste meetopstellingen hoogfrequent o.a. chloride is gemeten. De vraag was of de chloridewaarnemingen uit het ruimtelijk dichte helikopternet rechtstreeks gecombineerd konden worden met die van het temporeel dichte net van meetpalen. Hiertoe is regressie-analyse toegepast op die metingen uit de verschillende netten, die qua plaats en tijdstip vergelijkbaar zijn. De twee meetmethoden bleken vergelijkbare resultaten op te leveren: geen staptrend was aantoonbaar.



Figuur 6 Meetlokaties waterkwaliteitsonderzoek Zuidelijke Noordzee

## 5 CONCLUSIES

Er is een groot aantal praktische, direct toepasbare technieken beschikbaar voor het detecteren en beschrijven van trends. In sommige gevallen is een beheerder primair in trends geïnteresseerd. De relatie tussen meetnetoptimalisatie en trenddetectie is besproken. In andere gevallen vormen trends niet de specifieke interesse van de beheerder. Toch kan het ook dan van groot belang zijn een meetreeks eerst op trend te corrigeren alvorens verdere bewerkingen, zoals autocorrelaties e.d., uit te gaan voeren: een aanwezige trend kan het resultaat van deze verdere bewerkingen drastisch beïnvloeden.

## LITERATUUR

- BASSEVILLE, M., 1988. Detecting changes in signals and systems - a survey, *Automatica* vol. 24, no. 3, p. 309-326.
- BUISSHAND, T.A., 1984. Test for detecting a shift in the mean of hydrological time series, *Journal of Hydrology* 73, p. 51-69.
- KENDALL, M.G. and STUART, A., 1977. *The advanced theory of statistics*, 4th. ed. London, Griffin.
- SCHILPEROORT, T., 1986. Optimalisatie van bemonsteringsfrequenties bij waterkwaliteitsonderzoek, CHO-TNO Rapporten en Nota's 14, p. 79-97.

# BOX-JENKINS ANALYSE VAN STIJGHOOGTEN

P.K. Baggelaar

## SAMENVATTING

Een tijdreeks van de stijghoogte op een bepaalde locatie bevat waardevolle informatie over het geohydrologische systeem dat de stijghoogte heeft gegenereerd. De Box-Jenkins methode is een data-georiënteerde benadering waarmee zulke informatie uit de tijdreeks gefilterd kan worden. Er wordt daartoe een statistisch model afgeleid dat de stijghoogte relateert aan invloedsvariabelen (zoals het neerslagoverschot en grondwaterwinningen). Aan de hand van dit model kunnen er vervolgens uitspraken over het systeem worden gedaan. De methode kan bijvoorbeeld informatie opleveren over de manier waarop het neerslagoverschot de stijghoogte beïnvloedt, of over de verlaging van de stijghoogte als gevolg van een winning. Met het ruimtelijke beeld verkregen uit meerdere analyses kunnen zelfs bodemconstanten worden afgeleid, of kan een grondwatermodel worden geijkt.

## 1 INLEIDING

De Box-Jenkins benadering heeft nog enigszins de reputatie slechts black-box modellen te leveren, die beperkt bruikbaar zijn en geen inzicht verschaffen. Maar het blijkt vaak wel degelijk mogelijk de black-box te openen en er een beter

inzicht in het beschouwde systeem uit te halen. Dit zal hier worden toegelicht aan de hand van een systeemtheoretische beschouwing van de stijfhoogte, een voor veel geohydrologische onderzoeken centrale variabele.

Een systeem kan overigens enigszins intuïtief worden omschreven als een deel van de werkelijkheid, dat via ingangsen uitgangssignalen wisselwerkingen vertoont met z'n omgeving. Als de tijd een essentiële rol in het systeemgedrag speelt, spreekt men van een dynamisch systeem.

## 2           MODELLERING DYNAMISCHE SYSTEMEN

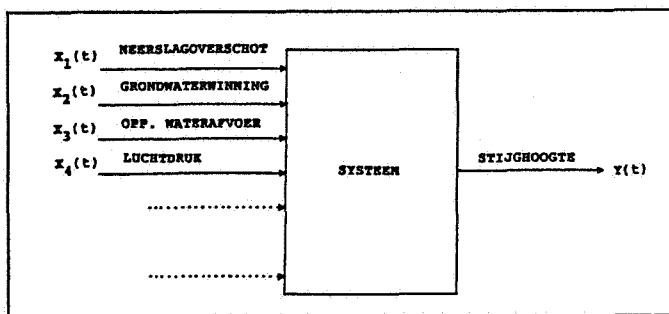
Bij onderzoeken in meerdere vakgebieden spelen dynamische systemen een centrale rol. Uitspraken worden daarbij doorgaans onderbouwd met behulp van een model van het systeem. In het geval er voldoende theoretische fysische kennis van de dynamica van het systeem is, kunnen de uitspraken worden onderbouwd met behulp van een mechanistisch model. Dit geeft een grote flexibiliteit, omdat daarmee in principe over allerlei theoretische situaties uitspraken gedaan kunnen worden. Deze benadering kan in de hydrologie vaak met succes worden gevolgd (bijvoorbeeld met een grondwatermodel dat gebaseerd is op de wet van Darcy en de wet van behoud van massa).

Uitspraken kunnen natuurlijk ook worden onderbouwd met een empirisch (= statistisch) model. Men noemt dit wel de data-georiënteerde benadering. Als er relevante tijdreeksen beschikbaar zijn biedt de Box-Jenkins methode hiertoe een mogelijkheid. In het geval van een gecompliceerd dynamisch systeem is dit vaak zelfs de enige manier om uitspraken te onderbouwen. De Box-Jenkins methode heeft dan ook vooral in de economie, met z'n gecompliceerde systemen, opgang gemaakt.

Om de Box-Jenkins benadering toe te lichten zullen we de stijghoogte op een bepaalde locatie opvatten als het uitgangssignaal van een dynamisch systeem, dat wordt aangedreven door meerdere ingangssignalen, voornamelijk van meteorologische, geohydrologische en waterhuishoudkundige aard (zie figuur 1). Zo'n systeem kan algemeen worden weergegeven als:

$$Y(t) = g(G, X, t) \quad [1]$$

met  $Y$  de stijghoogte,  $g(\cdot)$  één of andere functie,  $G$  een verzameling modelparameters,  $X$  de verzameling ingangssignalen en  $t$  de continue tijd.



Figuur 1 De stijghoogte als uitgangssignaal van een continu dynamisch systeem met meerdere ingangssignalen

Uit praktische overwegingen verdient het bij de empirische modellering van de stijghoogte aanbeveling deze te beschouwen als het uitgangssignaal van een discreet lineair dynamisch stochastisch systeem:

- . discreet: omdat relevante gegevens over de stijghoogte en ingangssignalen meestal slechts beschikbaar zijn in de vorm van discrete tijdreeksen;
- . lineair: omdat dit de minste theoretische en praktische problemen oplevert, bovendien is gebleken dat veel systemen redelijk beschreven kunnen worden met lineaire modellen;

. stochastisch: omdat zo'n systeem nooit exact (deterministisch) kan worden beschreven, enerzijds vanwege de benadering die is ingevoerd door de discretisatie en linearisatie, anderzijds vanwege het feit dat nooit alleingangssignalen bij de beschrijving kunnen worden meegenomen.

Zo'n systeem kan algemeen worden weergegeven als:

$$Y_t = f(F, X', t) + N_t \quad [2]$$

met  $Y$  de stijghoogte,  $f(\cdot)$  een lineaire functie,  $F$  een verzameling parameters,  $X'$  een deelverzameling van de ingangssignalen,  $N$  de ruis en  $t$  de discrete tijdsindex.

Omdat de algemene vorm van de Box-Jenkins modellen gelijk is aan [2] is de Box-Jenkins methode bij uitstek geschikt om zo'n systeem langs empirische weg te beschrijven. En omdat de methode er op is gericht het stochastische element terug te brengen tot een verschijnsel dat minimaal is en dat bekende waarschijnlijkheidswetten volgt, kunnen er uitspraken omtrent het systeem worden gedaan met minimale en kwantificeerbare onzekerheden.

## 4 DE BOX-JENKINS METHODE

### 4.1 Inleiding

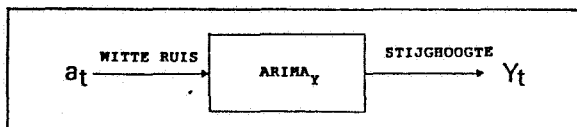
De klassieke statistische methodologie is zelden geschikt voor een statistische modellering van tijdreeksen, omdat deze gebaseerd is op onafhankelijke waarnemingen. De waarnemingen in een tijdreeks daarentegen zijn vaak niet onafhankelijk van elkaar. Bij gebruikelijke waarnemingsfrequenties als 12 of 24 per jaar bijvoorbeeld, wordt een hoge (lage) stijghoogte vaker gevolgd door een andere hoge (lage) stijghoogte, dan door een lage (hoge). Aanvangende met het werk

van Yule (1927) hebben een groot aantal prominente statistici zich met dit probleem beziggehouden, hetgeen culmineerde in het werk van Box en Jenkins (Box and Jenkins, 1970). Hun belangrijkste bijdragen bestonden uit de formulering van een familie van statistische modellen om tijdreeksen te beschrijven en uit aanwijzingen tot een verantwoorde modelbouw.

#### 4.2 Mogelijke benaderingen

Voor de modellering van een tijdreeks ontwikkelden ze twee benaderingen, namelijk de univariate modellering aan de hand van een Arima model en de transfer modellering aan de hand van een transfer-/ruis model.

Bij de univariate modellering wordt uitsluitend uitgegaan van de te modelleren tijdreeks, die wordt beschouwd als het uitgangssignaal van een discreet lineair dynamisch stochastisch systeem dat wordt aangedreven door een stochastisch ingangssignaal. Dit signaal is onbekend en ontstaat eigenlijk pas in de modelleerfase als het modelresidu. De modellering is er dan ook op gericht dit terug te brengen tot een signaal dat bekende waarschijnlijkheidswetten volgt ("witte ruis", zie figuur 2) en een minimale variantie heeft.



Figuur 2 Het principe van een Arima model

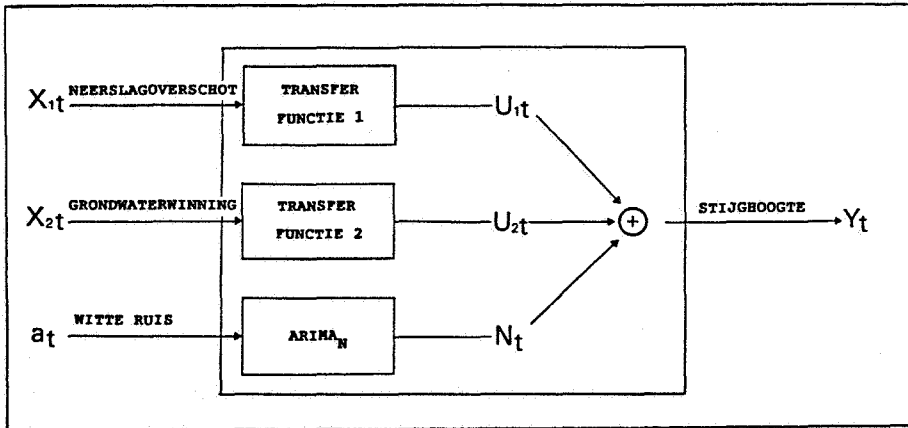
Witte ruis kan worden opgevat als een volledig toevallig signaal, bestaande uit een opeenvolging van onafhankelijke trekkingen uit een normale kansverdeling, met gemiddelde nul en variantie  $\sigma_a^2$ .

Het zal duidelijk zijn dat deze benadering nauwelijks inzicht levert in het beschouwde systeem. Het wordt veel toegepast om voorspellingen van een signaal te genereren.



Deze zijn dan uitsluitend gebaseerd op een analyse van het gedrag van dit signaal in het verleden.

Bij de transfer modellering wordt de te modelleren tijdreeks beschouwd als het uitgangssignaal van een discreet lineair dynamisch stochastisch systeem dat wordt aangedreven door meerdere ingangssignalen, waaronder ook witte ruis (zie figuur 3).



Figuur 3 Het principe van een transfer-/ruis model

Elk ingangssignaal ( $X_i$ ) levert hierbij volgens een afzonderlijke lineaire transfer functie een component van de stijghoogte ( $U_i$ ). De resterende ruis ( $N$ ) vertegenwoordigt de component van de stijghoogte die wordt geleverd door alle ingangssignalen die niet bij de modellering zijn betrokken. Deze ruis kan op zijn beurt worden weergegeven als het uitgangssignaal van een deelsysteem, dat volgens een Arima model (in feite ook een soort transfer functie) wordt aangedreven door witte ruis. De verschillende componenten geven door middel van lineaire superpositie de uiteindelijke waargenomen stijghoogte, volgens:

$$Y_t = U_{1t} + \dots + U_{mt} + N_t \quad [3]$$

Het principe van lineaire superpositie van afzonderlijke invloeden wordt overigens bij hydrologische modelleringen vaak toegepast.

De kracht van het transfer-/ruis model is dat een groot aantal dynamische situaties beschreven kunnen worden met slechts een gering aantal modelparameters.

Het moge duidelijk zijn dat de transfer modellering het meeste inzicht in het beschouwde systeem oplevert, daarom wordt daaraan in het nu volgende de meeste aandacht besteed.

#### 4.3 De transfer functie

De formulering van de dynamische relatie tussen een ingangsen uitgangssignaal van een systeem in een Box-Jenkins model is afgeleid van de algemene weergave van een continu lineair dynamisch systeem met een lineaire differentiaal vergelijking, volgens:

$$(1+A_1\frac{d}{dt}+\dots+A_R\frac{d^R}{(dt)^R})Y(t) = g(1+B_1\frac{d}{dt}+\dots+B_S\frac{d^S}{(dt)^S})X(t) \quad [4]$$

met Y respectievelijk X de afwijking ten opzichte van het gemiddelde van het uitgangssignaal respectievelijk ingangssignaal, t de continue tijd,  $A_1 \dots A_R$  en  $B_1 \dots B_S$  modelparameters en g het evenwichtseffect van X op Y ("steady-state gain"). Deze laatste is de waarde die Y uiteindelijk aanneemt als X continu op de waarde +1 wordt gehouden (bijvoorbeeld de stationaire verlaging van de stijghoogte als gevolg van een bepaalde eenheid grondwaterwinning).

De oplossing van deze vergelijking kan worden uitgeschreven als een convolutie integraal:

$$Y(t) = \int_0^{\infty} v(u) \cdot X(t-u) du \quad [5]$$

met u de tijdsvertraging en v(u) de continue impuls-respons functie. Het uitgangssignaal wordt dus beschreven als een lineaire functie van het ingangssignaal. Deze integraal

convergeert voor een stabiel systeem (een systeem dat bij een eindige verandering in het ingangssignaal een eindige verandering in het uitgangssignaal oplevert). De Box-Jenkins methode gaat in principe uit van stabiele systemen.

Analoog aan het bovenstaande kan een discreet lineair dynamisch systeem worden weergegeven met de lineaire differentie vergelijking:

$$(1+C_1\nabla+\dots+C_r\nabla^r)Y_t = g(1+D_1\nabla+\dots+D_s\nabla^s)X_t \quad [6]$$

met  $t$  de discrete tijdsindex,  $C_1\dots C_r$  en  $D_1\dots D_s$  modelparameters en  $\nabla$  de differentie-operator ( $\nabla Y_t = Y_t - Y_{t-1}$ ).

De oplossing hiervan kan worden uitgeschreven als:

$$\begin{aligned} Y_t &= \sum_0^{\infty} v_u \cdot X_{t-u} \\ &= v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots \\ &= (v_0 + v_1 B + v_2 B^2 + \dots) X_t \quad [7] \end{aligned}$$

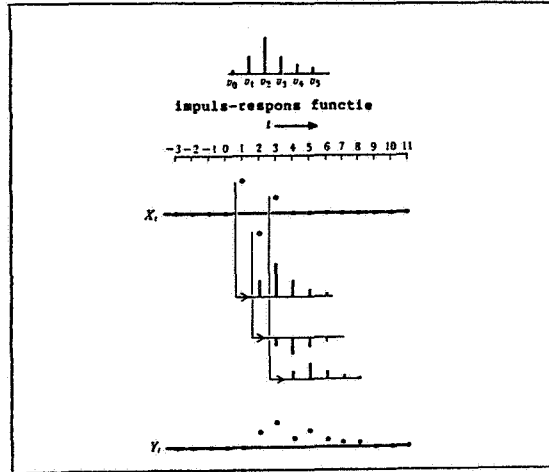
met  $v_0, v_1, v_2, \dots$  de impuls-respons gewichten (deze vormen de discrete impuls-respons functie) en  $B$  de backshift operator ( $BX_t = X_{t-1}$ , let wel:  $B$  is geen parameter, maar slechts een operator die dient om de notatie te vereenvoudigen).

De impuls-respons functie vertegenwoordigt de dynamica van het systeem en beschrijft in feite het uitgangssignaal  $Y$  bij een eenheidspuls in het ingangssignaal  $X$ . En als  $X$  bestaat uit een opeenvolging van verschillende pulsen, wordt  $Y$  beschreven als de som van door  $X$  geschaalde impuls-respons gewichten (zie [7]). Het principe van deze lineaire overdracht (transfer) van  $X$  naar  $Y$  is schematisch weergegeven in figuur 4, voor een hypothetische impuls-respons functie (boven weergegeven als staafdiagram).

In Box-Jenkins notatie wordt [6] uitgeschreven als:

$$(1-\delta_1 B - \dots - \delta_r B^r) Y_t = (\omega_0 - \omega_1 B - \dots - \omega_s B^s) X_t \quad [8]$$

met  $\delta_1 \dots \delta_r$  de autoregressieve parameters,  $\omega_0 \dots \omega_s$  de moving-average parameters en B weer de backshift-operator (hier gebruikt als substituut voor  $\nabla$ , aangezien  $B = 1 - \nabla$ ). Dit noemt men de transfer functie van de orde (r,s).



Figuur 4 De lineaire transfer van ingangssignaal X naar uitgangssignaal Y

#### 4.4 Het Arima model

Een bijzondere vorm van een transfer functie kan worden weergegeven met het Arima (AutoRegressive Integrated Moving-Average) model. Het uitgangssignaal wordt daarbij beschreven als een lineaire functie van witte ruis (zie figuur 2). De algemene vorm van het Arima model vertoont zeer grote overeenkomsten met die van de transfer functie [8], namelijk:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(\nabla^d Y_t^\lambda - c) = (1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad [9]$$

met  $\phi_1 \dots \phi_p$  de autoregressieve parameters,  $\theta_1 \dots \theta_q$  de moving-average parameters,  $a$  de witte ruis,  $\nabla$  de differentie-operator,  $d$  het aantal differenties,  $\lambda$  de transformatieparameter en  $c$  een constante. Dit noemt men het Arima model van de orde (p,d,q).

De bewerkingen in het tweede deel van de linkerkant van de vergelijking zijn om  $Y$  (indien nodig) te herleiden tot een (statistisch) stationair uitgangssignaal, oftewel een signaal waarvan de statistische eigenschappen (zoals gemiddelde en variantie) niet veranderen in de tijd. Want bij een stabiel systeem levert een stationair ingangssignaal (zoals a) ook een stationair uitgangssignaal.

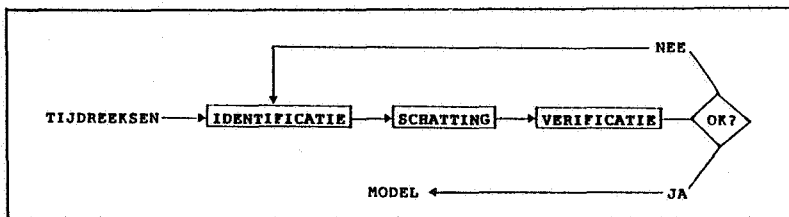
Tijdreeksen van de stijghoogte vertonen vaak duidelijke seizoensschommelingen. Zo'n tijdreeks kan beschreven worden met een seizoenaal Arima model, dat in feite een eenvoudige uitbreiding is van het bovengenoemde model.

## 5 HET BOUWEN VAN EEN BOX-JENKINS MODEL

### 5.1 Algemeen

Het bouwen van een Box-Jenkins model is, net als het bouwen van de meeste modellen, een iteratief proces. Per ronde worden drie fasen doorlopen (zie figuur 5). We onderscheiden namelijk:

- a) identificatie van het model;
- b) schatting van de modelparameters;
- c) verificatie van het model.

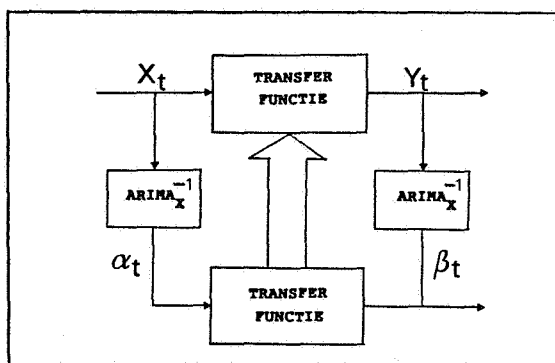


Figuur 5 Schema van de Box-Jenkins modelbouw

De fasen zullen hier worden toegelicht voor het transfer-/ruis model. De inhoud van de fasen voor het transfer-/ruis en het Arima model wijken weliswaar enigszins af, maar daar zal hier verder niet op worden ingegaan.

## 5.2 Identificatie transfer functie

De identificatie van een transfer functie bestaat uit het afleiden van de orde  $(r,s)$  van de functie aan de hand van de empirische samenhang tussen het ingangs- en uitgangssignaal van het systeem. Zo'n identificatie zou aanzienlijk kunnen worden vergemakkelijkt als het ingangssignaal zou kunnen worden gegenereerd in één of andere speciale vorm, zoals een stap, puls, sinus of anders als witte ruis. Maar bij veel systemen is zo'n proefopzet niet mogelijk en kan alleen worden uitgegaan van de beschikbare tijdreeksen van het ingangssignaal en uitgangssignaal. Dankzij de procedure van "prewhitening" (Box and Jenkins, 1970) kan hier echter aan worden tegemoet gekomen. Daartoe worden het ingangssignaal  $X$  en het uitgangssignaal  $Y$  aan de hand van het (inverse) Arima model van  $X$  getransformeerd tot nieuwe reeksen  $\alpha$  en  $\beta$  (zie figuur 6).



Figuur 6 Het principe van "prewhitening"

Omdat het nieuwe ingangssignaal ( $\alpha$ ) nu gevormd wordt door witte ruis is de identificatie eenvoudiger uit te voeren. En vanwege de lineaire en identieke transformatie geldt de afgeleide relatie tussen  $\alpha$  en  $\beta$  evenzeer voor de oorspronkelijke reeksen  $X$  en  $Y$ .

Centraal bij de identificatie staat het empirische kruiscorrelogram van  $\alpha$  en  $\beta$ , een plot van de geschatte kruiscorrelatiefunctie  $r_{\alpha\beta,k}$ , berekend volgens:

$$r_{\alpha\beta,k} = \frac{\frac{1}{n} \sum_{t=1}^{n-k} (\alpha_t - \bar{\alpha})(\beta_{t+k} - \bar{\beta})}{s_{\alpha} \cdot s_{\beta}} \quad \text{voor } k = 0, 1, 2, \dots \quad [10]$$

met  $k$  het tijdsverschil,  $\bar{\alpha}$  respectievelijk  $\bar{\beta}$  het gemiddelde en  $s_{\alpha}$  respectievelijk  $s_{\beta}$  de geschatte standaardafwijking van  $\alpha$  respectievelijk  $\beta$  en  $n$  het aantal waarnemingen.

De orde  $(r,s)$  van de transfer functie van  $\alpha$  en  $\beta$  kan namelijk visueel worden afgeleid uit de vorm van het kruiscorrelogram. Want als we er van uitgaan dat  $\alpha$  en  $\beta$  beide stationair zijn, kan worden aangetoond dat de kruiscorrelatie functie rechtevenredig is aan de impuls-respons functie, volgens:

$$\bar{v}_k = \frac{r_{\alpha\beta,k} \cdot s_{\beta}}{s_{\alpha}} \quad \text{voor } k = 0, 1, 2, \dots \quad [11]$$

met  $\bar{v}_k$  de geschatte impuls-respons functie en de overige symbolen als boven.

En de relatie tussen de vorm van de impuls-respons functie en de orde  $(r,s)$  van de transfer functie volgt uit de combinatie van [7] en [8]:

$$(v_0 + v_1 B + v_2 B^2 + \dots) = \frac{(\omega_0 - \omega_1 B - \dots - \omega_s B^s)}{(1 - \delta_1 B - \dots - \delta_r B^r)} \quad [12]$$

Praktische handleidingen voor Box-Jenkins modelbouw (o.a. McLeod, 1983) bevatten voorbeelden van kruiscorrelogrammen voor de meest gangbare orden van transfer functies, zoals  $(0,0)$ ,  $(0,1)$ ,  $(1,0)$  en  $(1,1)$ . Deze zijn theoretisch berekend aan de hand van [11] en [12]. Door het empirische kruiscorrelogram te vergelijken met deze voorbeelden kan de orde worden bepaald.

### 5.3 Identificatie transfer-/ruis model

Bij de identificatie van een transfer-/ruis model wordt de orde  $(r,s)$  van elke afzonderlijke transfer functie en de orde  $(p,d,q)$  van het Arima model voor de ruis vastgesteld. De identificatie van een afzonderlijke transfer functie kan bemoeilijkt worden door de ruis en/of de invloed van andere ingangssignalen. Zo zou transfer functie 1 in figuur 3 moeten worden geïdentificeerd aan de hand van  $X_1$  en  $U_1$ , maar deze kan slechts worden afgeleid aan de hand van  $X_1$  en  $Y (=U_1+U_2+N)$ , omdat  $U_1$  niet beschikbaar is. Als een directe identificatie hierdoor onmogelijk wordt, zijn er soms wel manieren om hieraan tegemoet te komen:

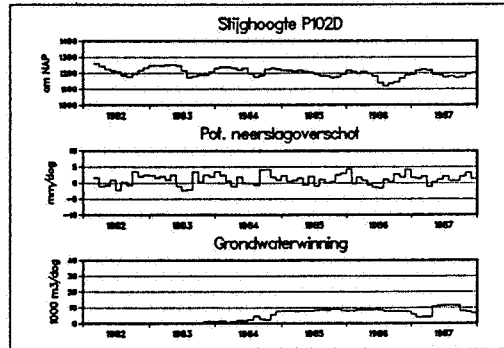
- . als er bijvoorbeeld één ingangssignaal is dat een duidelijke trend in het uitgangssignaal veroorzaakt, kan deze trend verwijderd worden alvorens een transfer functie voor een ander ingangssignaal te identificeren;
- . aan de hand van globaal inzicht in de dynamica van het (deelsysteem) kan een transfer functie gepostuleerd worden, die vervolgens in de verificatiefase geëvalueerd en daarna eventueel bijgesteld wordt.

Het Arima model voor de ruis kan bij de eerste identificatie slechts gepostuleerd worden, we weten immers niet hoe de ruis er uitziet. Maar als alle relevante ingangssignalen bij de modellering zijn betrokken kan er van worden uitgegaan dat het ruismodel stationair is. Als zo'n model in de verificatiefase overigens niet voldoet is er blijkbaar een relevant ingangssignaal over het hoofd gezien.

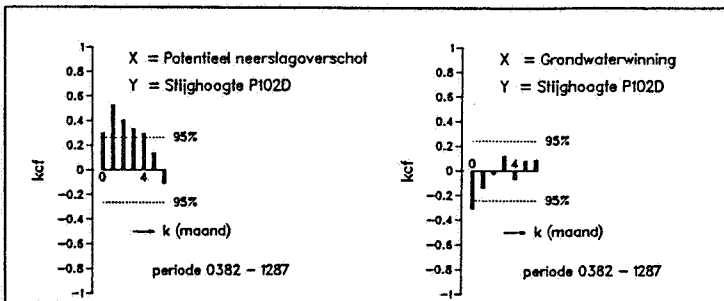
Het bovenstaande kan kort worden toegelicht aan de hand van het transfer-/ruis model van de stijghoogte in waarnemingsput P102D. Uit een gebiedsinventarisatie kwamen het neerslagoverschot en een grondwaterwinning als vermoedelijk relevante ingangssignalen naar voren. Het schema van het te bouwen transfer-/ruis model is dan als in figuur 3. Omdat er geen gegevens over het neerslagoverschot beschikbaar waren,



wordt deze bij de modellering vertegenwoordigd door het potentieel neerslagoverschot. De tijdreeksen zijn op maandbasis weergegeven in figuur 7 en de kruiscorrelogrammen in figuur 8.



Figuur 7 Tijdreeksen op maandbasis van de stijghoogte in waarnemingsput P102D, het potentieel neerslagoverschot en een nabijgelegen grondwaterwinning



Figuur 8 Relevante kruiscorrelogrammen (na "prewhitening") voor de identificatie van het transfer-/ruis model voor de stijghoogte in P102D

Het linker kruiscorrelogram suggereert dat de stijghoogte reageert op een puls van het neerslagoverschot met een korte stijging, gevolgd door een exponentieel verlopende terugkeer naar de evenwichtssituatie. Het rechter kruiscorrelogram suggereert een onmiddellijke daling van de stijghoogte als reactie op een puls van de grondwaterwinning, weer gevolgd

door een exponentieel verlopende terugkeer naar de evenwichtssituatie. Ze duiden op transfer functies van de orde (1,1) respectievelijk (1,0). Omdat alle (vermoedelijk) relevante ingangssignalen bij de modellering zijn betrokken, is het Arima model voor de ruis gepostuleerd als een stationair model (1,0,0).

#### 5.4 Schatting parameters transfer-/ruis model

De schatting van de parameters geschiedt aan de hand van de beschikbare tijdreeksen en de gepostuleerde orden van de afzonderlijke transfer functies en het Arima model. Belangrijk aspect hierbij is dat álle parameters simultaan worden geschat.

Voor de schattingsprocedure staan verschillende mathematische methoden ter beschikking, die alle gebaseerd zijn op een optimalisatie ten aanzien van één of ander criterium.

In theorie levert de maximum likelihood methode de meest verantwoorde schatters, maar deze geeft soms rekentechnische problemen. De kleinste kwadraten methode geeft minder problemen en levert bovendien in de meeste gevallen schatters die een zeer goede benadering geven van de maximum likelihood schatters. De methode schat de parameters zódanig, dat de variantie van het stochastische ingangssignaal minimaal is. Hierbij wordt voor elke geschatte parameter de standaardafwijking bepaald en wordt de correlatie tussen de geschatte parameters bepaald.

Aan de hand van de geschatte parameters van een afzonderlijke transfer functie kan ook het evenwichtseffect van een bepaald ingangssignaal worden geschat. Voor het bovengenoemde voorbeeld levert dit een stationaire verlaging van de stijghoogte in P102D van 41,6 cm (standaardafwijking 4,6 cm) als gevolg van een grondwaterwinning van 3 miljoen m<sup>3</sup> per jaar.

## 5.5 Verificatie transfer-/ruis model

De zeer grondige kritische beoordeling van het model in de verificatiefase is een sterk punt in de Box-Jenkins modelbouw. In deze fase wordt het model getoetst aan de centrale vooronderstelling die er aan ten grondslag ligt, namelijk dat het stochastische ingangssignaal witte ruis vormt. Daarnaast wordt het model geëvalueerd aan de hand van:

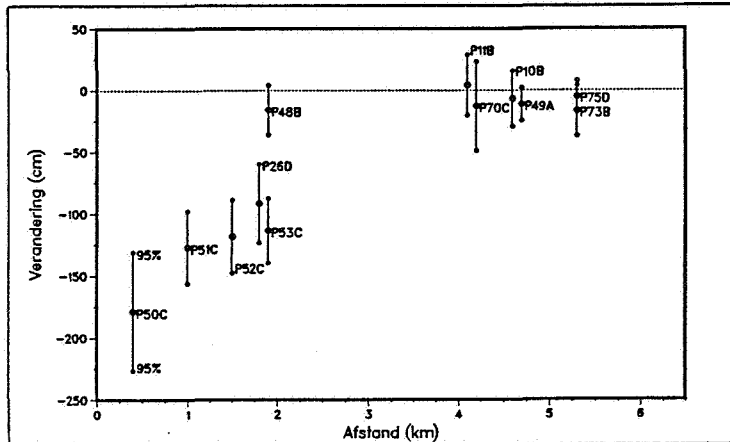
- . de statistische significantie van elke geschatte parameter;
- . de correlatie tussen geschatte parameters (waaruit eventueel de overbodigheid van een bepaalde parameter kan worden vastgesteld);
- . het kruiscorrelogram van elk afzonderlijk ingangssignaal met het stochastische ingangssignaal (waaruit eventueel een onvolledig geformuleerde afzonderlijke transfer functie kan worden vastgesteld).

Met behulp van geconstateerde discrepanties kan vervolgens een nieuw model worden geïdentificeerd.

## 6 TOEPASSING

Er is al gebleken dat de benadering onder bepaalde randvoorwaarden met succes toegepast kan worden bij empirisch verlagingsonderzoek (zie o.a. Van Geer, 1988 en Baggelaar, 1988). Er wordt daarbij dankbaar gebruik gemaakt van het feit dat een stationaire verlaging kan worden geschat, zonder dat deze hoeft te zijn opgetreden. Als het wordt toegepast voor een aantal waarnemingsfilters op verschillende diepten rond de winplaats, levert het per onderscheiden watervoerend pakket een discontinu beeld van de stationaire verlaging van de stijghoogte als gevolg van de winning (zie bijvoorbeeld figuur 9). Dit kan een doel op zich zijn, bijvoorbeeld voor een vergunningsverlenende instantie, maar het kan natuurlijk ook gebruikt worden om bodemconstanten te schatten met

behulp van stationaire pompproefformules, of om een drie-dimensionaal (mechanistisch) grondwatermodel te ijken. De methode kan ook gebruikt worden om een beter inzicht te verkrijgen in de reactie van de stijghoogte op bepaalde invloedsvariabelen, zoals het neerslagoverschot.



Figuur 9 Voorbeeld van de resultaten van een empirisch verlagingsonderzoek: de geschatte stationaire verlagingen in het 1<sup>e</sup> watervoerend pakket (en hun 95% betrouwbaarheidsintervallen) als gevolg van een winning van 5 miljoen m<sup>3</sup> per jaar (uit: Baggelaar, 1987)

## 7 ENKELE PRAKTISCHE SLOTOPMERKINGEN

Als er een transfer-/ruis model wordt gebouwd met twee sterk gecorreleerde ingangssignalen, zal de methode nauwelijks onderscheid weten te maken tussen hun invloeden op het uitgangssignaal (hetgeen tot uiting komt in sterke correlaties tussen de betreffende parameters). Men moet hier altijd op bedacht zijn, omdat het soms tot bizarre resultaten kan leiden.

Met enige vindingrijkheid kan de Box-Jenkins methode ook worden gebruikt voor niet-lineaire situaties. Als bijvoorbeeld wordt verondersteld dat een verlaging seizoensafhankelijk is, kan er een transfer-/ruis model worden gebouwd met

de zomerwinning en de winterwinning als afzonderlijke ingangssignalen (dit zijn dan tijdreeksen die een half jaar de waarde nul hebben).

Gezien de vele rekenintensieve handelingen spreekt het vanzelf dat de modelbouw geschiedt met behulp van software. Er is al software op de markt die deze modelbouw vrijwel volledig automatisch uitvoert. Voor hydrologische doeleinden moet ik deze echter afraden, er zijn namelijk verschillende beslismomenten in de modelbouw, waarbij geohydrologische inzichten net zo goed dienen mee te wegen als mathematische indicaties.

#### LITERATUUR

- BAGGELAAR, P.K. (1987): 'Invloed van de grondwaterwinningen op de winplaatsen De Mondaf en Lievensberg (Bergen Op Zoom) op stijghoogten. Tijdreeksanalyse van stijghoogten.' KIWA N.V., SWO-87.325, Nieuwegein.
- BAGGELAAR, P.K. (1988): 'Tijdreeksanalyse bij verlagingsonderzoek: principe en voorbeeld.' H<sub>2</sub>O (21), 1988, nr. 16, pp. 443 - 450.
- BOX, G.E.P. and JENKINS, G.M. (1970): 'Time Series Analysis, Forecasting and Control.' Holden-Day, San Francisco.
- GEER, F.C. VAN (1988): 'Verlagingsberekening met transfermodellen rond de winplaats Spannenburg.' H<sub>2</sub>O (21) 1988, nr. 16, pp. 451 - 454.
- MCLEOD, G. (1983): 'Box Jenkins in Practice. Volume I: Univariate Stochastic and Transfer Function/Intervention Analysis.' Gwilym Jenkins & Partners Ltd, Time Series Library, Lancaster.
- YULE, G.U. (1927): 'On a method of investigating periodicities in disturbed series, with special reference to Wolfer sunspot numbers.' Phil. Trans. Roy. Soc. London, Ser. A, 226, pp. 267 - 298, 1927.

TIJDREEKSANALYSE VAN VERANDERINGEN IN CHEMISCHE SAMENSTELLING VAN DE  
BODEMOPLOSSING IN DE STUDIELOCATIE HACKFORT GEDURENDE 6 JAAR

A. Stein, T. Feijtel, T. Pape en H. Booltink

**Abstract**

Space-time variability plays a central role when data monitored in spatially inhomogeneous areas are analyzed. In this study, space-time models are utilized which are closely linked to models describing spatial variation. Methods are applied to soil solution data, collected in two plots (B and C) of the Hackfort estate in the Eastern part of the Netherlands. Interest focusses on changes in the chemical composition of the soil solution. Plot B showed a more intense pH decrease due to a higher N-loading, leading to acidification over the whole profile. In the topsoil of plot C acidification was also observed, but at a more moderate rate. Seasonal variation indicated a pH-decrease in the summer and early fall due to an increase of mineralization and nitrification.

**1 Inleiding**

Atmosferische depositie van  $\text{SO}_x$ ,  $\text{NO}_x$  en  $\text{NH}_x$  wordt voor de bodem een externe bron van ionen nadat deze componenten op de vegetatie of in de bodem zijn geoxideerd tot zwavelzuur en salpeterzuur. Wanneer deze sterke zuren met het regenwater in de bodem infiltreren verandert de chemische samenstelling van de bodemoplossing. Deze verandering hangt af van de aard en de omvang van chemische en biologische interacties tussen bodemoplossing enerzijds en (a-)biotische bodemcomponenten anderzijds. Mineraalverwerking, kationenuitwisseling, nitrificatie en afbraak van organische stof zijn de sleutelprocessen in de buffering van extern aangevoerd zuur en zij reguleren de samenstelling van de bodemoplossing. Formuleringen van deze processen en transporten zijn uitvoerig behandeld in Mulder (1988) en Van Grinsven (1988). Simulatie van transport en chemische interacties door een waterverzuringmodel (ILWAS) blijkt

mogelijk te zijn door calibratie van snelheidsconstanten voor nitrificatie, afbraak van organische stof en oplossing van gibbsiet (Van Grinsven, 1988).

De doelstelling van deze studie is om inzicht te krijgen in de temporele veranderingen van de bodemoplossing in een bodemprofiel t.g.v. zure depositie. Hierbij gaat de aandacht in eerste instantie uit naar mogelijke veranderingen in zuurgraad (pH) t.g.v. van aantasting en permanente veranderingen in het zuurbufferend vermogen van de bodem. Een eventuele pH-verlaging in deze systemen kan mogelijk gerelateerd worden aan een uitputting van aluminium houdende mineralen of een verhoogde nitrificatie snelheid. Met name het detecteren van veranderingen van deze effecten bij toenemende diepte is het uitgangspunt van deze studie.

In de literatuur wordt tot op heden weinig aandacht besteed aan ruimte-tijd modellen (zie echter Bottcher en Strebel, 1988a, 1988b en 1988c en Bennett, 1979). De opzet die in deze studie gekozen wordt is echter nieuw en is gebaseerd op de theorie van de intrinsieke stochastische functies. Deze theorie wordt op grote schaal toegepast in ruimtelijke voorspellings methoden zoals Kriging en Cokriging (zie bijv. Stein et al., 1988). Nieuw in de benadering die hier gekozen is, is dat we niet langer spreken over een punt in een 1-, 2- of 3-dimensionale ruimte, maar over een punt in een 2-, 3- of 4-dimensionale ruimte-tijd.

## 2 Studielocatie

In een 3.2 ha groot bosgebied op het landgoed Hackfort in Oost-Gelderland is in het kader van een 6-jarig bodemverzuringsonderzoek (1981-1987) van de vakgroep Bodemkunde & Geologie van de Landbouw Universiteit Wageningen een uitgebreid hydrogeochemisch monitoring-programma uitgevoerd (Fig. 1). De bodems in dit gebied zijn gevormd in



*Fig. 1 Plaats van de studie locatie Hackfort in Oost-Nederland*

zandige tot lemige, Pleistocene rivierafzettingen. Behalve textuurverschillen zijn er verschillen in hoogteligging, in hydrologie en in de dieptes waarop kalk voorkomt. Uit de 4 bemonsterde locaties (plot A, B, C en D) zijn voor deze studie plot B en C gekozen. Plot B ligt t.o.v. plot C  $\pm$  30 cm hoger, heeft een wat lemiger bovengrond en er komen naast eiken ook enkele berken voor. In plot B komt kalk voor binnen 2m diepte wat van invloed is op de samenstelling van het grondwater en op het bodemvocht op 90 cm diepte.

De samenstelling van depositie en bodemvocht werd maandelijks geanalyseerd. Bodemvocht werd bemonsterd op de dieptes 0 cm (vanaf 1983), 10 cm (in duplo), 20 cm (in enkelvoud), 40 cm (plot B in duplo, plot C in enkelvoud), 60 cm (plot B in enkelvoud, plot C in duplo) en 90 cm (in duplo, vanaf september 1983 in viervoud). Door het afwijkende gedrag van de strooisellaag (0 cm) en de door de aanwezigheid van kalk veroorzaakte verschillen tussen de plots B en C op 90 cm is voor deze studie alleen gebruik gemaakt van de monsters op 10/20 cm (de bovengrond) en op 40/60 cm (de ondergrond). Tevens werd gebruik gemaakt van de gemeten bodemtemperatuur (STEMP, Soil Temperature) op dieptes 10, 20 en 50 cm en de gesimuleerde vochtgehalten (MCONT, Moisture Content) en vochtspanningen (PRHEAD, Pressure Heads) op de verschillende dieptes. Simulaties werden uitgevoerd met het SWATRE model (Feddes et al. 1978; Belmans et al. 1983), een non-stationair flow model dat gebaseerd is op de 'Darcy' transport- en de continuïteits-vergelijking. Details m.b.t. SWATRE simulaties van de Hackfort locaties worden beschreven door Van Grinsven (1988).

Voor gedetailleerde gegevens over de meetlocaties, het monitoring programma, de analyse-methoden, de volledige analyse-resultaten en diverse conclusies wordt verwezen naar Van Breemen et al. (1988) en de aanvullende rapportage t.b.v. het nationale programma Zure Regen (rapport 02-01, 1988).

### **3 Statistische analyses**

#### **3.1 De tijdreeks**

In deze analyse wordt uitgegaan van een multipele tijdreeks, bestaande uit waarnemingen aan de concentraties van de  $H^+$ ,  $Al^{3+}$ ,  $NO_3^-$  en



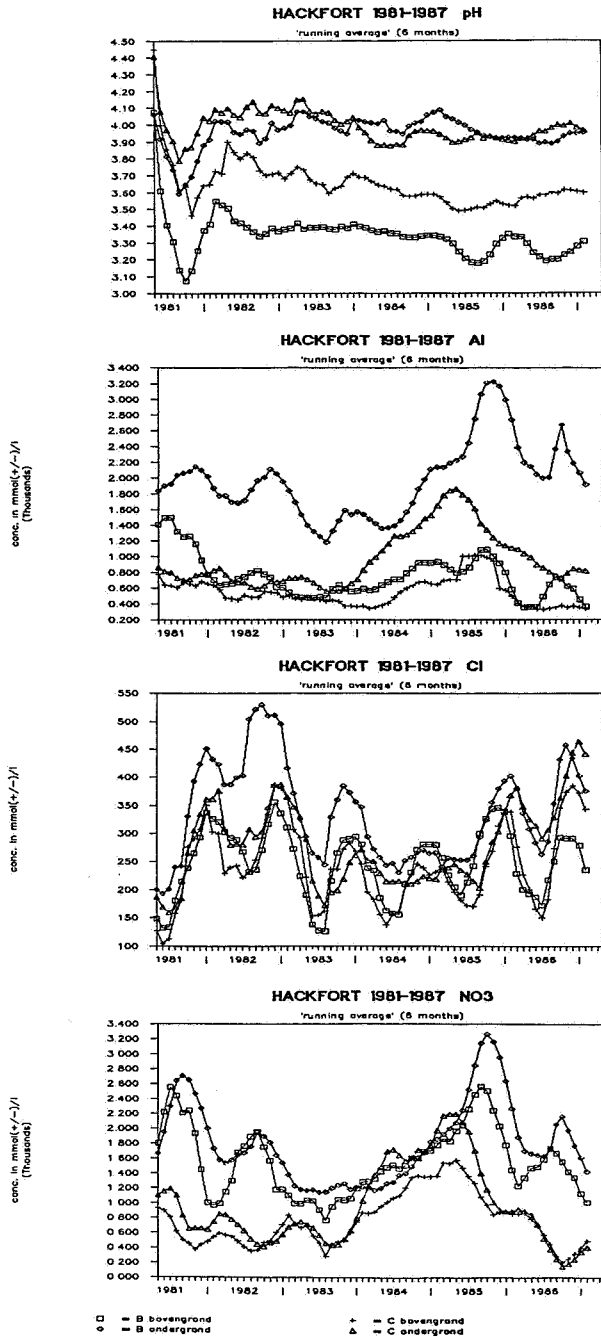


Fig. 2 Tijdreeksen voor pH, Al<sup>3+</sup>, Cl<sup>-</sup> en NO<sub>3</sub><sup>-</sup> voor de boven- en ondergrond van plots B en C. Smoothing is uitgevoerd op grond van een 6-maands moving average.

- Cl<sup>-</sup>-ionen die over een periode van 6 jaar zijn verkregen (Fig. 2). Deze tijdreeks wordt onder andere gekarakteriseerd door de volgende facetten:
- missing values in de zomermaanden, veroorzaakt door het regelmatig ontbreken van voldoende bodemvocht;
  - en sterke seizoensfluctuatie en een verschil tussen 'natte' en 'droge' jaren;
  - een vermoedelijke correlatie met bodemtemperatuur, vochtgehalte en vochtspanning, die op hun beurt weer gerelateerd zijn aan seizoensveranderingen;
  - een uitgesproken diepte-effect, dat samenhangt met de variatie in het bufferende vermogen van de bodem op verschillende dieptes;
  - adsorptie/desorptie problemen in de cups in de beginperiode van de monitoring;
  - een mogelijke trendmatige verandering van de bodemkenmerken in de tijd, waarbij sprake kan zijn van ernstige verzuring van de bodem;
  - systematische verschillen tussen de plots B en C;
  - een toevals-component die veroorzaakt wordt door meeton nauwkeurigheid, verschil in meetmateriaal, etc.

Het ontbreken van waarnemingen, soms gedurende meerdere opeenvolgende maanden, is een structureel probleem dat met lineaire interpolatie, zoals dat gangbaar is in bepaalde tijdreeksanalyses, niet adequaat opgelost kon worden. Het uitvoeren van een spectraal-analyse verviel hiermee, omdat de standaardpakketten uitgaan van volledige tijdreeksen. Voor het analyseren van deze multi-pele tijdreeks is daarom een model-formulering gekozen waarin de waarnemingen niet regelmatig verdeeld hoefden te zijn in de ruimte-tijd. Er is met name gestreefd naar het verkrijgen van inzicht in de temporele factoren, het diepte-effect en het verschil tussen de plots. Ook werd de invloed van de externe factoren op de concentraties van ionen in de bodem onderzocht.

### 3.2 Ruimte-tijd modellering

Voor verschillende doeleinden worden relaties onderzocht die zowel ruimtelijke als temporele variabiliteit kennen: veranderende landkarakteristieken in de loop der jaren in een bepaald gebied, het temperatuurverloop in Nederland in 1988, de geomorfologische wordingsgeschiedenis

van een terrassensysteem, etc. Al deze onderwerpen hebben gemeen dat er (in de tijd) veranderende kenmerken binnen een regionale context onderzocht worden.

Vaak wordt een aantal, zeg  $m$ , kenmerken tegelijk bestudeerd. Het  $j^e$  kenmerk  $y_j$ ,  $j=1, \dots, m$ , wordt genoteerd als  $y_j(x, t)$ , waarbij onderstreping aangeeft dat de kenmerken stochastisch zijn. De plaatsvector  $x$  is een vector in 1, 2 of 3 dimensies,  $t$  representeert de tijd. De modelvorming is nu als volgt:

$$y_j(x, t) = X(x, t)\beta_j + \varepsilon_j(x, t). \quad (1)$$

Hierin is  $X(x, t)$  een  $n$  bij  $p$  matrix met in de  $i^e$  rij waarnemingen aan  $p$  functies van de plaatscoördinaten, in drie dimensies:  $\zeta_1$ ,  $\zeta_2$  en  $\zeta_3$ , en van de tijdscoördinaat  $\zeta_t$  in het  $i^e$  punt in de ruimte-tijd.  $\beta_j$  is een, gewoonlijk stochastische,  $p$ -vector van te schatten parameters en  $\varepsilon_j(x, t)$  is een stochastische vector met storingstermen, die in het algemeen afhangen van  $x$  en  $t$ . In navolging van Matheron (1973) veronderstellen we dat de stochastische functies  $y_j$  ofwel stationair zijn, ofwel dat ze stationaire aangroeiingen van orde  $k$  hebben. Laat  $r$  een vector in de ruimte-tijd zijn met projecties  $r_x$  in de ruimte en  $r_t$  in de tijd. Een stochastische functie heet stationair als de verwachting van deze functie bestaat, d.w.z.  $E(y_j(x, t)) = \mu_j$ , en de covariantie tussen de waarnemingspunten met de plaatsvectoren  $(x, t)$  en  $(x+r_x, t+r_t)$  bestaat, d.w.z.  $\text{cov}(y_j(x, t), y_j(x+r_x, t+r_t)) = c_j(r)$ , waarbij beide onafhankelijk zijn van de vector  $(x, t)$ . We spreken van stochastische functies met stationaire aangroeiingen of incrementen van orde  $k$  indien  $E(\lambda' y_j(x, t)) = 0$  voor alle  $\lambda$  met  $\lambda' X(x, t) = 0$ . Een covariantie functie voor zo'n kenmerk hoeft niet te bestaan, maar wel geldt:

$$\text{cov}(\lambda' y_j(x, t), \lambda' y_j(x+r_x, t+r_t)) = \lambda' C_j(r) \lambda \quad (2)$$

waarin de elementen van  $C_j$  worden gegeven door een gegeneraliseerde (of pseudo-) covariantie functie. In het laatste geval, dat correspondeert met een graad van de polynomiale trend gelijk aan  $k$ , worden de functies van de coördinaten in de ruimte-tijd beschreven door monomen, d.w.z. functies van de gedaante  $\zeta_1^{k_1} \zeta_2^{k_2} \zeta_3^{k_3} \zeta_t^{k_t}$  met  $k_1+k_2+k_3+k_t \leq k$ . Nu hoeft niet te gelden dat  $c_j(0) = \text{var}(y_j)$ , omdat deze variantie mogelijk niet

bestaat.

In de huidige modelformulering wordt uitgegaan van isotropie, d.w.z.  $c_j(r)$  hangt voor alle  $j$  alleen af van de lengte van de vector  $r$  in de ruimte-tijd en niet van de richting, m.a.w.  $c_j(r) = c_j(|r|)$ , waarbij  $|r|^2 = \zeta_1^2 + \zeta_2^2 + \zeta_3^2 + \zeta_4^2$ , de (Euclidische) lengte van  $r$ . Hoewel er meerdere functies  $c(r)$  voldoen aan de eis dat  $c(r)$  een (gegeneraliseerde) covariantie functie is, voldoet bij een polynomiale trend van graad  $k$  een  $c(r)$  van de gedaante

$$c(r) = \alpha_0 \delta(|r|) + \sum_{l=1}^{k+1} \alpha_l |r|^{2l-1} \quad (3)$$

waar  $\delta(r)$  de Delta-functie van Dirac is, d.w.z.  $\delta(|r|) = 1$  als  $|r| = 0$  en  $\delta(|r|) = 0$  als  $|r| \neq 0$ . De coëfficiënten  $\alpha_l$  moeten aan zekere positief-definietheids eisen voldoen. Voor  $k = 2$  en de dimensie van de ruimte-tijd gelijk aan 4, kan m.b.v. Fourier getransformeerden aangetoond worden dat er moet gelden:  $\alpha_0 \geq 0$ ,  $\alpha_1 \leq 0$ ,  $\alpha_3 \leq 0$  en  $\alpha_2 \geq -\sqrt{(28/3)}\alpha_1\alpha_3$ . Voor lagere dimensies gelden analoge eisen (Delfiner, 1976).

Naast een mogelijke trend in de kenmerken kan er sprake zijn van temporele fluctuaties en van belangrijke invloeden van andere kenmerken. Model (1) kan worden uitgebreid (Kendall et al., 1983) met een aantal covariabelen op de volgende wijze:

$$y_j(x,t) = X(x,t)\beta_j + Z_1 s_{1j} + Z_2 s_{2j} + e_j \quad (4)$$

Hierin is  $Z_1$  een matrix met voor de  $q^e$  kolom ( $q = 1, \dots, 12$ ) op de  $i^e$  plaats een 1 als de  $i^e$  waarneming uit de  $q^e$  maand stamt en 0 elders.  $Z_2$  is een matrix met waarnemingen aan de co-variabelen en  $e$  is een vector met residuen. Een schatting voor de elementen van de vectoren  $s_{1j}$  en  $s_{2j}$  wordt gevonden door een matrix  $L$  te kiezen zo dat  $LX(x,t) = 0$ , waarmee de trend wordt uitgefilterd. Als we daarnaast de matrices  $Z_1$  en  $Z_2$  samenvoegen tot een matrix  $Z$  en de vectoren  $s_{1j}$  en  $s_{2j}$  tot een vector  $s_j$ , en de waarnemingsvectoren  $y_j(x,t)$  tot een matrix van waarnemingen  $\underline{Y}$ , dan kan model (4) herschreven worden als:

$$L\underline{Y}_j = LZs_j + L\underline{e}_j \quad (5)$$

Schattingen voor de coëfficiënten van de vector  $s$  vinden we dan via een gegeneraliseerd lineaire regressie model. Als de matrix  $Z$  niet van volle rang is dienen we ons te richten op schatbare functies van de parameters, waarbij b.v. de waarden in verschillende maanden worden vergeleken met die in één van te voren vastgestelde maand. De oplossingen voor de coëfficiënten van de vector  $s_j$  worden daarmee gevonden in een deelruimte van de oplossingsruimte die loodrecht staat op de ruimte van alle oplossingen. Hiervoor moeten enkele vrijheidsgraden worden ingeleverd, wat bij voldoende waarnemingen als regel geen bezwaar is.

Het probleem van ontbrekende waarnemingen wordt opgelost door een matrix  $L$  van contrasten te kiezen loodrecht op de deel-matrix van  $X$  die correspondeert met aanwezige waarnemingen.

Als regel zijn de graad van de polynomiale trend (en daarmee het aantal kolommen van de matrix  $X$ ) evenals de coëfficiënten van de (gegeneraliseerde) covariantiefunctie niet bekend terwijl het van belang is om hier inzicht in te krijgen. Een van de methoden, die daarbij gevolgd kan worden en aansluit op Delfiner (1976), komt neer op het volgende. Op grond van waarnemingen in de ruimte-tijd aan het te voorspellen kenmerk, een keuze voor de graad van de trend en een keuze voor de gegeneraliseerde covariantie functie worden voorspellingen uitgevoerd in alle locaties waar waarnemingen beschikbaar zijn. Bij het uitvoeren van de voorspellingen worden de waarnemingen in de voorspelpunten niet gebruikt, omdat de voorspeller exact is. Als voorspeller  $\underline{t}$  voor de waarde  $\underline{y}_0$  van het kenmerk in zo'n locatie  $z_0$  wordt gebruik gemaakt van:

$$\underline{t} = x_0' \hat{\beta} + c_0' C^{-1} (y - X \hat{\beta}) \quad (6)$$

waarbij  $\hat{\beta}$  een schatting is voor  $\beta$ ,  $x_0$  een vector bestaande uit waarden aan dezelfde monomen als de matrix  $X$ , geevalueerd in  $z_0$ ,  $C$  en  $c_0$  respectievelijk een matrix en een vector die afhangen van de gegeneraliseerde covariantie functie. Er kan bewezen worden dat  $\underline{t}$  de beste lineaire zuivere voorspeller voor  $\underline{y}_0$  is (Stein & Corsten, 1988). De verschillen tussen waarneming  $y_i$  en voorspelling  $t_i$ , de voorspelfouten, zijn aangroeiingen. Omdat de verwachting van zo'n aangroeiing gelijk is aan 0, is een gekwadrateerde aangroeiing een

schatting voor de variantie van de voorspelfout. Omdat uit (3) volgt dat (2) lineair is in de coëfficiënten  $\alpha_1$  kunnen m.b.v. een lineair regressie model de gekwadraterde aangroeiingen  $(t_i - y_i)^2$  in de voorspel-locaties zo goed mogelijk verklaard worden door (2). Dit levert nieuwe schattingen voor de coëfficiënten  $\alpha_1$  op. Dit proces convergeert als regel betrekkelijk snel, hetgeen bij de gekozen graad van de trend leidt tot een covariantiefunctie. Door deze procedure te herhalen voor verschillende graden van trends en de gekwadraterde aangroeiingen te vergelijken, b.v. met behulp van een verdelingsvrije toets, wordt de graad van de trend verkregen.

Indien er echter sprake is van een trend, dan kunnen wel schattingen voor de coëfficiënten van de vector  $\beta$  verkregen worden, ze zijn echter formeel inhoudsloos omdat nu niet meer geldt dat  $E\mathbf{y} = \mathbf{X}(x,t)\beta$ , maar alleen  $E(\lambda'\mathbf{y}) = \lambda'\mathbf{X}\beta = 0$ . We moeten dan volstaan met de aan-/afwezigheid van een trend en de richting (stijgend/dalend).

#### 4 Resultaten en Discussie

##### 4.1 Aannames en keuzes

De modelformulering zoals hierboven is beschreven is toegepast voor concentraties van opgeloste bestanddelen in het bodemvocht op de studielocatie Hackfort. De ruimte-tijd kent hier twee dimensies, één dimensie voor de tijd en één dimensie voor de diepte. Hoewel het in principe mogelijk is om alle gegevens in één ruimte-tijd model op te nemen, is er gekozen voor toepassing van een model in meerdere situaties, omdat we met name geïnteresseerd waren in verschillen tussen bovengrond en ondergrond en in verschillen tussen de plots B en C. In totaal is een model vier maal toegepast en zijn de resultaten met elkaar vergeleken: toepassing 1 betreft de bovengrond (dieptes 10cm en 20cm) voor plot B; toepassing 2 betreft de ondergrond (dieptes 40cm en 60cm) voor plot B; toepassing 3 betreft de bovengrond voor plot C; toepassing 4 betreft de ondergrond voor plot C. Voor de covariantie functies is verondersteld dat deze de gedaante hebben:

$$c_j(r) = c_j(0), \text{ als } r = 0$$

$$c_j(r) = 0 \text{ als } r \neq 0.$$

d.w.z.  $\alpha_0 = c(0)$  en  $\alpha_1 = \alpha_2 = \alpha_3 = 0$  voor alle  $j$ . Deze keuze impliceert dat de kenmerken een bijzondere afhankelijkheidsrelatie kennen, nl. dat de waarnemingen onderling onafhankelijk zijn.

Een cruciaal punt in bovenstaande modelformulering vormt de isotropie van de kenmerken in de ruimte-tijd. Op grond van de overweging dat de infiltratiesnelheid in de profielen ca. 1m per jaar bedraagt, is het duidelijk dat de  $x=t$  as een hoek van ongeveer  $45^\circ$  maakt indien een periode van 1 jaar overeenkomt met een diepte van 1 meter.

De maand april of het begin van het hydrologische jaar werd gekozen als referentie (goed uitgespoeld profiel).

Aandacht gaat in eerste instantie uit naar significantie, graad en grootte van de trend voor de verschillende kenmerken op de verschillende dieptes. Vervolgens wordt aandacht besteed aan de grootte van de seizoenseffecten en aan de invloeden van de externe factoren STEMP en MCONT. Omdat PRHEAD waarden via een pF relatie verkregen zijn uit MCONT is dit verklarende kenmerk in de verdere analyse niet meegenomen. M.b.t. de residuen is het verder van belang om inzicht te krijgen in de onderlinge samenhang van de residuen voor de verschillende te verklaren kenmerken en hun seriële correlatie.

Alle analyses zijn uitgevoerd m.b.v. het statistische pakket SAS.

## 4.2 Resultaten

Voor een overzicht van de kenmerken in boven- en ondergrond en de eenheden en hun beschrijvende statistiek wordt verwezen naar Appendix A. De ontwikkelingen van de concentraties gedurende de onderzoeksperiode waren met name van belang om permanente veranderingen in de bodemoplossing op te sporen (Tabel 1).

Opvallend zijn de verschillen in verzuring tussen de plots B en C in de bovengrond. In beide gevallen wordt er een sterk significante verzuring geconstateerd, die echter sneller verloopt in plot B. Dit komt enigzins als een verrassing aangezien plot B (gemiddelde pH = 3.34) in de bovengrond reeds zuurder is dan plot C (gemiddelde pH = 3.65). Deze sterkere pH daling in plot B gaat gepaard met een significante toename van  $Al^{3+}$ , hetgeen zou wijzen op  $Al^{3+}$ -buffering in de bovengrond.

Consistent lagere N-depositie in plot C gedurende de laatste decennia

		pH	Al <sup>3+</sup>	Cl <sup>-</sup>	NO <sub>3</sub> <sup>-</sup>
Plot B	b.g.	-.020**	+28.5**	+8.06*	+62.9**
	o.g.	-.011**	+61.7**	+8.82**	+105.7**
Plot C	b.g.	-.011**	+14.3	+8.54*	-10.5
	o.g.	-.001	-17.1	-5.00	-38.3*

Tabel 1 Grootte, richting (+/-) en significantie (\*:  $\alpha = .05$ , \*\*:  $\alpha = .01$ ) van de trend in eenheden per maand.

heeft ertoe geleid dat zowel de biota als de vegetatie nog in staat is om deze input te assimileren. Plot B heeft (en had waarschijnlijk ook) een grotere N-aanvoer vanuit de atmosfeer, en kan daarom een groot deel niet meer assimileren. De N uitspoeling loopt op tot 28 kg/ha.jr in plot B bij een aanvoer van 56.2 kg/ha.jr (Van Breemen et al. 1987). T.g.v. N-transformaties ( $\text{NH}_4^+ \rightarrow \text{NO}_3^- + 2\text{H}^+$ ) is de actuele  $\text{H}^+$  produktie in plot B ongeveer 20% hoger dan in plot C, als gevolg van verschil in N-input. De pH-daling in plot B blijkt dan ook in sterke mate samen te gaan met een significante toename in  $\text{NO}_3^-$  (Tabel 1). Deze trend zet zich voort tot op grotere diepte in plot B waar zowel  $\text{NO}_3^-$  als  $\text{Al}^{3+}$  significant toenemen.

Kolom experimenten (materiaal uit plot B en C) hebben aangeven dat de  $\text{H}^+$  consumptie snelheid ongeveer gelijk is aan de  $\text{Al}^{3+}$  oplosnelheid (Van Grinsven, 1988; Kooge, 1988). Verschillen tussen plot B en C kunnen echter niet verklaard worden aan de hand van mogelijke mineralogische en chemische discrepanties tussen beide plots.

Uit Tabel 2 blijkt dat de verhouding van de concentratie van  $\text{Al}^{3+}$  tot de totale hoeveelheid zuur ( $\text{H}^+ + \text{Al}^{3+}$ ) in de oplossing weinig verschilt van de verandering in deze ratio met de tijd t.g.v. verzuringseffecten.

		Gemiddelde		Verandering		% Al <sup>3+</sup> van de	
		H <sup>+</sup>	Al <sup>3+</sup>	H <sup>+</sup>	Al <sup>3+</sup>	Zuurgraad	
Plot B	b.g.	457	763	21.5	28.5	57.0	62.5
	o.g.	224	548	5.74	14.3	71.5	71.0

Tabel 2 Gemiddelde concentraties ( $\mu\text{mol}(+)/\text{l}$ ) en gemiddelde jaarlijkse veranderingen ( $\mu\text{mol}(+)/\text{l.yr}$ ) van  $\text{H}^+$  en  $\text{Al}^{3+}$ .



De bijdrage van  $\text{Al}^{3+}$  tot total zuur bedraagt ongeveer 60% in plot B en 70% in plot C. In plot C is er kennelijk sprake van een effectievere consumptie van  $\text{H}^+$  en produktie van  $\text{Al}^{3+}$  dan in plot B. De minder efficiënte buffering in plot B heeft wellicht te maken met mogelijke verschillen in reaktiviteit of oploskinetiek van vrij  $\text{Al}_2\text{O}_3$  en Al-houdende silicaten. Dat zou betekenen dat de hoeveelheid aan vrije  $\text{Al}_2\text{O}_3$  (0.5% in plot B en 0.4% in plot C) geen maat is voor de efficiëntie van de  $\text{H}^+$  buffering van de bodem. Als de verzuring nog eens zes jaar met dezelfde snelheid zou doorlopen, zou dit resulteren in een pH van respectievelijk 2.70 en 3.06 voor plots B en C. Deze waarden lijken onredelijk laag, en de grootte van de coëfficiënten lijkt ongeschikt om voorspellingen te doen.

Nadat de trends bepaald zijn is gekeken naar de maand-effecten. Een aantal interessante facetten is te bespeuren in het parallelle verloop van de pH in beide plots (Fig. 3). De pH daling van de bodemoplossing blijkt het sterkst te zijn in de zomer en herfst. Dit loopt samen met een hogere mineralisatie- en nitrificatie snelheid in beide plots. Het geproduceerde  $\text{NO}_3^-$  wordt effectief geconsumeerd door vegetatie en biota in de bovengrond van plot C. In de bovengrond van plot B en ondergrond van beide plots is er duidelijk een toename in  $\text{NO}_3^-$  waar te nemen aangezien produktie en remobilizatie duidelijk hoger liggen dan de assimilatie door biota en gewas. Tegen het eind van het seizoen spoelt het profiel door en verschillen in  $\text{NO}_3^-$  concentraties zijn gering. Temporele variaties in  $\text{Al}^{3+}$  laten een concentratie verhoging zien in de herfst en wintermaanden in de ondergrond van plot B. Uitspoeling van  $\text{Al}^{3+}$  en verzuring in de herfst en winter in plot B komt hier duidelijk tot uiting.

De kwaliteit van het voorspellende model is in het algemeen erg goed (Tabel 3) en het is geen verrassing dat opname van de co-variabelen STEMP en MCONT een betrekkelijk geringe verhoging van  $R^2$  laat zien. Uitzondering vormen de concentraties van  $\text{Al}^{3+}$  en  $\text{Cl}^-$  in de ondergrond van plot B.

Seizoensvariaties zoals stijging van STEMP en geleidelijke daling van MCONT in de zomermaanden kunnen leiden tot inzoutingseffecten met mogelijke concentratie verhogingen. Uit een analyse uitgevoerd zonder maandeffecten blijkt dat er geen significante relatie bestaat tussen de bodemtemperatuur en het  $\text{Cl}^-$  gehalte. De toename van de  $R^2$  waarde van het

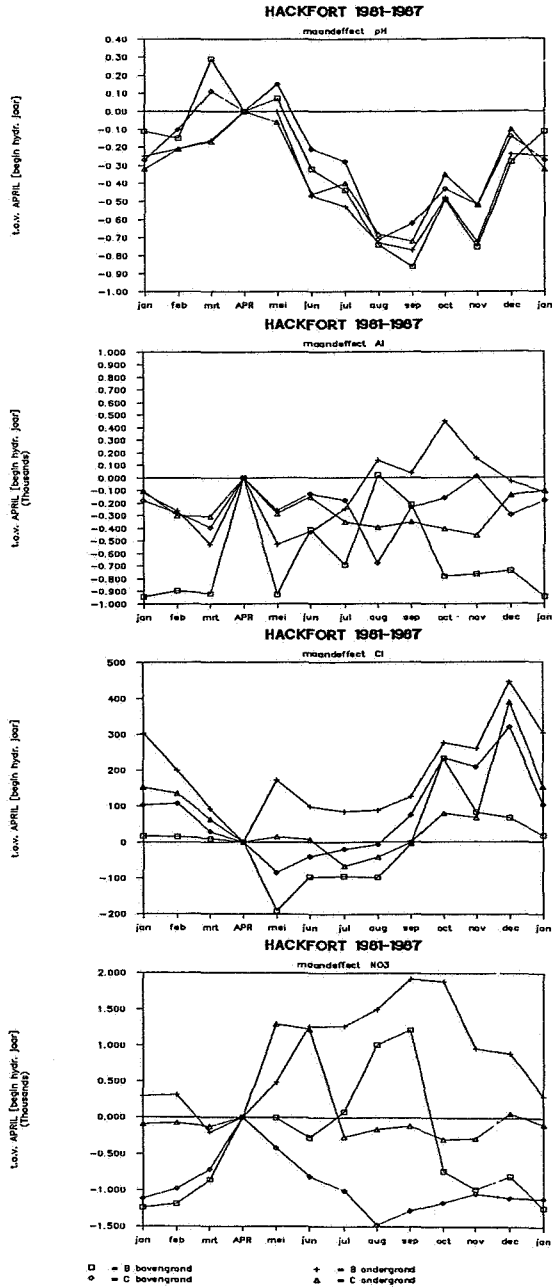


Fig. 3 De maandeffecten in vergelijking met het begin van het hydrologisch jaar (april) voor de verschillende elementen voor boven- en ondergrond van plots B en C.

		pH	Al <sup>3+</sup>	Cl <sup>-</sup>	NO <sub>3</sub> <sup>-</sup>
Plot B	b.g.	93.0/93.8	76.2/77.5	89.3/90.2	79.3/86.3
	o.g.	89.0/93.8	46.0/69.5	48.1/61.5	92.0/92.8
Plot C	b.g.	88.2/88.5	70.0/70.0	78.2/80.6	92.9/94.0
	o.g.	81.9/82.9	69.9/70.5	73.2/73.6	72.6/75.5

Tabel 3. Percentage verklaarde variantie bij de verschillende plots, boven- en ondergrond zonder (links) resp. met (rechts) de co-variabelen STEMP en MCONT.

model met co-variabelen voor Cl<sup>-</sup> en Al<sup>3+</sup> in de ondergrond van plot B kan toegeschreven worden aan mogelijke variaties in depositie en aan verdunningseffecten. Seizoensvariaties in Cl<sup>-</sup> worden beïnvloed door hoge depositie in de winter en indamping in de zomer, hetgeen een mogelijke seizoensvariatie zou kunnen opheffen. De waarden voor de maandeffecten worden lastiger te interpreteren indien beide co-variabelen in het model worden opgenomen. Er is ongetwijfeld sprake van zoiets als collineariteit tussen de maandeffecten en de co-variabelen voor zover dit van toepassing is op dummy variabelen en continue variabelen.

Seizoensvariaties in bodemconcentraties zijn niet altijd te interpretern in termen van processen. Interacties tussen bodemoplossing en minerale vaste fase, mineralisatie, opname door vegetatie, evaporatieve indamping in de zomer, en verdunning door infiltratie in de winter bemoeilijkt de interpretatie. Netto produktie van vrij H<sup>+</sup> in de bodemoplossing is het grootst in de bovengrond door verhoogde mineralisatie en nitrificatie tijdens de zomermaanden. Dit blijkt ook uit de sterke samenhang van NO<sub>3</sub><sup>-</sup> met de temperatuur. Uitspoeling en verlaagde N-opname door vegetatie in de herfst en wintermaanden zou verantwoordelijk kunnen zijn voor een uitgesproken pH daling in de wintermaanden. Het contrasterende verzuringsgedrag in boven- en ondergrond is consistent met het NO<sub>3</sub><sup>-</sup> en Al<sup>3+</sup> patroon die beide sterk beïnvloed worden door veranderingen in STEMP en MCONT. Concentraties van Al<sup>3+</sup> en NO<sub>3</sub><sup>-</sup> in de bodemoplossing blijken respectievelijk significant toe te nemen tijdens de zomermaanden in de bovengrond en af te nemen in de ondergrond.

Er zijn verschillende evaluaties toegepast op de residuen die overblijven na deze analyses. Zo is er gekeken naar de eerste orde

		pH	Al <sup>3+</sup>	Cl <sup>-</sup>	NO <sub>3</sub> <sup>-</sup>
Plot B	b.g.	.08	-.11	.28	-.53
	o.g.	-.13	-.27	-.21	.13
Plot C	b.g.	.24	.06	.18	.08
	o.g.	-.05	.07	.03	-.53

Tabel 4 Geschatte eerste orde autocorrelaties op grond van de Durbin-Watson stochast.

autocorrelaties, verkregen m.b.v. de Durbin-Watson stochast (Tabel 4). Na opname van de co-variabelen STEMP en MCONT veranderen de eerste orde autocorrelaties voor Al<sup>3+</sup> en Cl<sup>-</sup> in de ondergrond van plot B in resp. -.02 en -.09. Op grond van deze eerste orde autocorrelaties lijken we aan te kunnen nemen dat de residuen zich als min of meer stationaire tijdreeksen gedragen, al zijn de waarden voor NO<sub>3</sub><sup>-</sup> in de bovengrond van plot B en in de ondergrond van plot C aan de hoge kant en al moeten we de Durbin-Watson stochast met de nodige voorzichtigheid interpreteren. Hoewel de R<sup>2</sup>-waarde weinig verandert bij opname van de co-variabelen STEMP en MCONT in het model, veranderen dez eerste orde autocorrelaties voor NO<sub>3</sub><sup>-</sup> in de bovengrond van plot B en in de ondergrond van plot C in resp. -.26 en -.43.

#### 4.3 Discussie

In deze analyse is gepostuleerd dat de covariantie functie bekend is, en zodanig is dat de waarnemingen onderling onafhankelijk zijn. Dit is als een eerste benadering gedaan. Het biedt in de toekomst de mogelijkheid om een vergelijking met meer geavanceerde modellen te kunnen maken. Mede gezien de Durbin-Watson waarden lijkt de gehanteerde benadering in deze studie reëel te zijn met name voor pH, Al<sup>3+</sup> en Cl<sup>-</sup>. Dit laat toe dat uitspraken gedaan worden met behulp van standaard beschikbare statistische software.

In plot B is er sprake van een significante pH daling in zowel boven- als ondergrond, terwijl er in plot C enkel verzuring optreedt in de bovengrond. Dit verschil is toe te schrijven aan een hogere N-depositie in plot B, zoals aangegeven door de sterke samenhang tussen H<sup>+</sup> en NO<sub>3</sub><sup>-</sup>.

Zowel in plot B als in plot C wordt de protonen aanvoer gebufferd door Al-mineralen. Buffering in plot B gebeurt minder efficiënt dan in plot C, mogelijk toe te schrijven aan verschillen in oplosbaarheid van secundaire Al-oxyden.

In deze studie wordt een belangrijk probleem gevonden in de collineariteit (Belsley et al., 1980) van de verschillende kenmerken, m.n. tussen maandeffecten en de co-variabelen STEMP en MCONT. Met de analyses die zijn uitgevoerd om deze collineariteit te detecteren zijn weinig resultaten bereikt. De belangrijkste reden is dat de  $Z'Z$  matrix singulier is en dat we zoeken naar collineariteit tussen factoren, nl. maandeffecten, en variabelen. Zo vonden we een hoge waarde voor de grootste singuliere waarde ( $\geq 5000$ ) waarvoor de proportionele variantie van alle maandeffecten gelijk is aan 1.000. De proportionele variantie van de co-variabelen zijn hiervoor echter erg laag. Voor de op één na grootste singuliere waarde (10-15) zien we een hoge proportionele variantie voor STEMP, maar hierop zijn de proportionele varianties van de maandeffecten uiteraard gelijk aan 0.000. De indruk bestaat dat deze analyse methode uitbreiding behoeft naar een situatie van singulariteit van  $Z'Z$  matrices.

#### Acknowledgement

De auteurs willen Prof. Dr. Ir. N. van Breemen bedanken voor het verzamelen van de gegevens voor zijn kritisch commentaar op dit manuscript.

## Literatuur

- Belmans, C., J.G. Wesseling, and R.A. Feddes. 1983. Simulation of the waterbalance of a cropped soil: SWATRE. *J. Hydrol.* 63, 271-286.
- Belsley, D.A., E. Kuh and R.E. Welsch. 1980. Regression diagnostics: Identifying influential data and sources of collinearity. Wiley, New York.
- Bennett, R.J. 1979. Spatial time series. Pion ltd. London.
- Bottcher, J. and O. Strebel. 1988a. Spatial variability of groundwater solute concentrations at the water table under arable land and coniferous forest. Part 1: Methods for quantifying spatial variability (geostatistics, time series analysis, Fourier transform smoothing. *Z. Pflanzenernähr. Bodenk.*, 151, 185-190.
- Bottcher, J. and O. Strebel. 1988b. Spatial variability of groundwater solute concentrations at the water table under arable land and coniferous forest. Part 2: Field data for arable land and statistical analysis. *Z. Pflanzenernähr. Bodenk.*, 151, 191-195.
- Bottcher, J. and O. Strebel. 1988c. Spatial variability of groundwater solute concentrations at the water table under arable land and coniferous forest. Part 3: Field data for a coniferous forest and statistical analysis. *Z. Pflanzenernähr. Bodenk.*, 151, 197-203.
- Delfiner, P. 1976. Linear estimation of non stationary spatial phenomena. In: M. Guarascio et al. (eds.), *Advanced geostatistics in the mining industry*, 49-68. Reidel, Dordrecht.
- Feddes, R.A., P.J. Kowalik, and H. Zaradny, 1978. Simulation of field water use and crop yield. *Simulation Monographs*. Pudoc. Wageningen.
- Jorgenson, D.W. 1964. Minimum variance linear unbiased seasonal adjustment of economic time series. *J. Amer. Statist. Ass.*, 59, 681.
- Kendall, M., A. Stuart and J.K. Ord. 1983. *The advanced theory of statistics*. Volume 3, Design and Analysis, and Time-series. Griffith & Co. London.
- Kooge, C. 1988. *Verzuringsonderzoek in Hackfort*. Doctoraal scriptie. Bodemkunde en Geologie, Landbouw Universiteit Wageningen.
- Matheron, G. 1973. The intrinsic random functions and their applications. *Adv. Appl. Prob.*, 5, 439-468.
- Mulder, J. 1988. Impact of acid atmospheric deposition on soils. Field monitoring and aluminium chemistry. PhD thesis.

- Stein, A. and L.C.A. Corsten. 1988. Universal Kriging and Cokriging as a regression procedure. Submitted to Biometrics.
- Stein, A., W. van Dooremolen, J. Bouma and A.K. Bregt. 1988. (Co-) kriging point data on moisture deficit. Soil Sci. Soc. Am. J. In press.
- Van Breemen, N., J. Mulder, and J.J.M. van Grinsven 1987. Impact of acid atmospheric deposition on woodland soils in the Netherlands. II. N-transformations. Soil Sci. Soc. Am. J. 51, 1634-1640.
- Van Breemen, N., W.F.J. Visser and Th. Pape. 1988. Biochemistry of an oak-woodland ecosystem in the Netherlands affected by acid atmospheric deposition. Agricultural Research Reports 930, Pudoc, Wageningen.
- Van Grinsven, J.J.M. 1988. Impact of acid atmospheric deposition on soils. Quantification of chemical and hydrologic processes. PhD thesis.

Appendix

Overzicht van de onderzochte kenmerken.

Eenheid		pH	Al <sup>3+</sup> -----	Cl <sup>-</sup> μeq./l	NO <sub>3</sub> <sup>-</sup> -----	STEMP °C	MCONT %	PRHEAD mbar	
Plot B	b.g.	μ	3.34	763	257	1580	8.9	0.27	-149
		σ	0.16	499	122	742	4.3	0.05	187
		N	131	113	131	126	120	129	129
	o.g.	μ	3.96	1972	335	1789	8.9	0.22	-202
		σ	0.16	666	134	806	3.2	0.07	254
		N	135	122	135	133	121	133	133
Plot C	b.g.	μ	3.65	548	250	797	8.8	0.16	-222
		σ	0.17	427	135	505	4.3	0.06	703
		N	138	126	138	135	123	136	136
	o.g.	μ	3.98	984	288	969	9.0	0.15	-389
		σ	0.18	418	140	683	3.5	0.08	1150
		N	140	130	141	137	123	139	139

Gemiddelde ( $\mu$ ), standaardafwijking ( $\sigma$ ) en aantal observaties ( $N$ ) voor de verschillende kenmerken over de gehele tijdsperiode, uitgesplitst naar boven- en ondergrond en naar plot.





## SPECTRAALANALYSE

A.W. Heemink

### SAMENVATTING

Het doel van de tijdreeksanalyse in het algemeen is het onderkennen van systematische verbanden die mogelijk binnen één meetreeks of tussen meerdere reeksen bestaan en deze verbanden te onderscheiden van toevallig opgetreden verstoringen of ruis. De tijdreeksanalyse laat zich onderverdelen in parametrische en niet-parametrische methodieken. De eerste beschrijven de systematische verbanden in het tijdsdomein met behulp van een tijdreeksmodel (bijvoorbeeld Box en Jenkins modellen). De modelbouwer kiest in dit geval een model met een aantal onbekende parameters en vervolgens worden de meetreeksen gebruikt om deze parameters te bepalen. Bij de niet-parametrische techniek of spectraalanalyse wordt de analyse van de tijdreeks in het frekwentie domein uitgevoerd en worden o.a. energie spectra en frekwentie-overdrachtsfuncties voor het beschrijven van de systematische verbanden tussen verschillende meetreeksen bepaald. Het gebruik van spectraalanalyse is aanzienlijk eenvoudiger dan een Box en Jenkins benadering.

In dit artikel wordt een korte beschrijving van de spectraalanalyse gegeven en wordt verder vooral ingegaan op de verschillen en overeenkomsten met de Box en Jenkins aanpak.

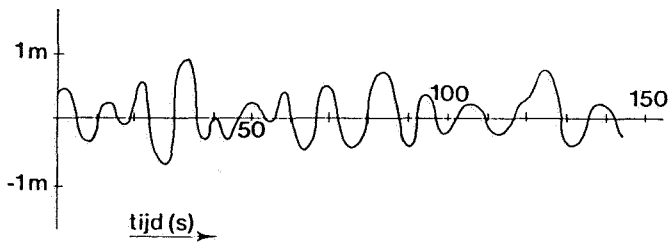
## 1 INLEIDING

Bij de meeste statistische methodieken, zoals onder andere regressie-analyse, beschouwt men de waarnemingen van een bepaald verschijnsel als onderling onafhankelijk. De volgorde waarin de metingen worden verricht is hier onbelangrijk. Het wezenlijke verschil tussen deze statistische technieken en de tijdreeksanalyse is dat in het laatste geval de onderlinge samenhang van de metingen niet wordt verwaarloosd. Sterker nog, er wordt van de systematische verbanden die in een meetreeks bestaan gebruik gemaakt om nauwkeuriger voorspellingen te kunnen berekenen. In dit geval is de volgorde waarin metingen worden verricht dus juist van wezenlijk belang. Vooral sinds het verschijnen van de inmiddels als standaardwerken beschouwde boeken "Spectral Analysis and Its Applications" van Jenkins en Watts in 1968 en "Time Series Analysis: Forecasting and Control" van Box en Jenkins in 1970 is de tijdreeksanalyse bijzonder populair geworden. Het eerste boek behandelt de analyse van een reeks metingen m.b.v. de spectraalanalyse, het tweede boek is gewijd aan ARIMA-modellen.

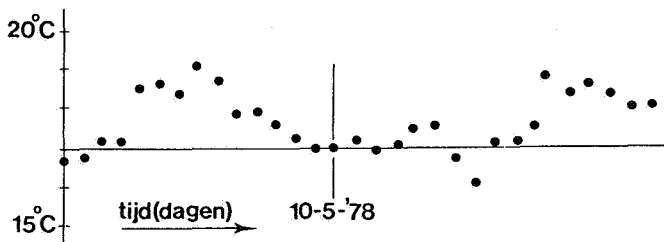
In dit artikel wordt een beschrijving van de spectraalanalyse gegeven. Na een korte inleiding over stochastische processen in paragraaf 2 wordt in paragraaf 3 nader ingegaan op de tijdreeksanalyse in het algemeen en, vooral, op de overeenkomsten en verschillen tussen de spectraalanalyse en de aanpak van Box en Jenkins. Paragraaf 4 behandelt in het kort de klassieke spectraalanalyse en tenslotte worden in paragraaf 5 enkele, meer recente, ontwikkelingen beschreven.

## 2 STOCHASTISCHE PROCESSEN

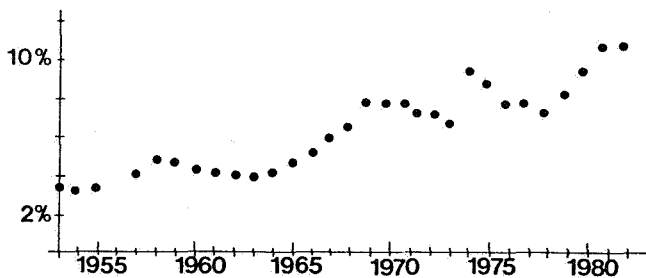
Een stochastisch proces is een wiskundig model voor een fluctuerend verschijnsel waarvan het verloop in de tijd alleen in termen van waarschijnlijkheid te beschrijven is. In de figuren 1 t/m 4 zijn ter illustratie registraties van een aantal van zulke fluctuerende verschijnselen gegeven.



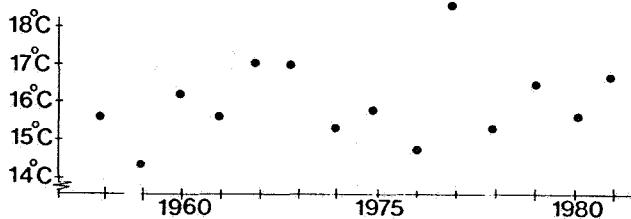
Figuur 1 Golfhoogte in volle zee



Figuur 2 Daggemiddelde van de temperatuur in De Bilt



Figuur 3 Jaarlijkse gemiddelde van de hypotheekrente



Figuur 4 Gemiddelde temperatuur in de zomer

Ondanks de onvoorspelbaarheid van de fluctuaties in de verschillende signalen valt op dat de fluctuaties steeds een specifiek patroon vertonen, dat duidelijk te onderscheiden is van de patronen die bij de andere signalen optreden. De theorie van de stochastische processen heeft nu als doel de wetmatigheden die aan het verschijnsel ten grondslag liggen wiskundig te karakteriseren. Dit kan door het definiëren van de volgende eigenschappen van een stochastisch verschijnsel:

- het **gemiddelde** waaromheen het verschijnsel fluctueert;
- de **spreiding** van de fluctuaties rond het gemiddelde;
- de **autokovariantiefunctie**, die aangeeft hoe sterk een waarde van het proces samenhangt met waarden van datzelfde proces in het verleden. Zo zullen, als deze samenhang klein is, opeenvolgende waarnemingen van een proces weinig met elkaar samenhangen en zal het proces heel snel fluctueren (zie figuur 4).

Voor de tijdreeksanalyse, waar in het volgende hoofdstuk aandacht aan zal worden geschonken is één speciaal type van proces van belang: het **stationaire** proces. Bij dit proces geldt dat de bovengenoemde drie statistische eigenschappen onafhankelijk zijn van de tijd. Dit houdt in dat verondersteld wordt dat een proces zich in de toekomst statistisch gezien hetzelfde zal gedragen als in het verleden. Voorbeelden hiervan zijn gegeven in de figuren 1 en 2, terwijl in figuur 3 een proces is getekend dat duidelijk niet stationair is: Zowel het gemiddelde als de spreiding van het proces nemen toe terwijl het proces sneller gaat fluctueren.

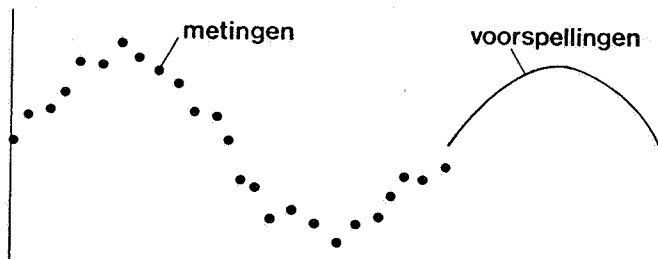
### 3 TIJDREEKSANALYSE

Het doel van de tijdreeksanalyse in het algemeen is het onderkennen van systematische verbanden die mogelijk binnen één meetreeks of tussen meerdere reeksen bestaan en deze verbanden te onderscheiden van toevallig opgetreden verstoringen of ruis. Deze verbanden kunnen bijvoorbeeld gebruikt worden om voorspellingen te berekenen. Deze werkwijze kan geïllustreerd worden met figuur 5, waar de meetreeks een sinusbeweging voorstelt met daarop gesuperponeerd een ruis proces, bijvoorbeeld veroor-

zaakt door meetfouten. Is men nu in staat om met de metingen de periodieke beweging te scheiden van de ruis, dan kunnen er optimale voorspellingen worden berekend (zie figuur 5).

Er is één wezenlijke voorwaarde waaraan de reeksen moeten voldoen wil de tijdreeksanalyse met succes kunnen worden toegepast: de metingen moeten afkomstig zijn van een **stationair** stochastisch verschijnsel. De reeks wordt immers voorspeld met behulp van metingen die in het verleden zijn verricht. De voorspellingen zijn in dit geval slechts optimaal als het verschijnsel zich in de toekomst statistisch gezien hetzelfde gedraagt als in het verleden. Zo zal in figuur 5 de voorspelling niet optimaal zijn als de frekwentie van het verschijnsel in de toekomst verandert.

Een beperking van de tijdreeksanalyse is dat in het algemeen de verbanden lineair worden verondersteld. Bij de aanwezigheid van sterke niet-lineaire verbanden wordt de theorie aanzienlijk gecompliceerder. De praktijk heeft echter uitgeezen dat de lineaire theorie bij veruit de meeste toepassingen toereikend is.



Figuur 5 Voorspellen van een sinusbeweging met ruis

De tijdreeksanalyse laat zich onderverdelen in **parametrische** en **niet-parametrische** methodieken. De laatste, waar de spectraalanalyse een voorbeeld van is, beschrijven de systematische verbanden met behulp van overdrachtsfuncties. De functies worden uitsluitend uit de ingevoerde meetreeksen bepaald en er hoeft geen extra informatie van het gemeten verschijnsel, zoals bijvoorbeeld de vorm van de overdrachtsfuncties, te worden toegevoegd.

Bij parametrische methodieken, waartoe ook de ARIMA-model-aanpak van Box en Jenkins behoort, worden de systematische verbanden met behulp van een model beschreven. De modelbouwer kiest in dit geval één ARIMA-model met een aantal onbekende parameters en vervolgens worden de meetreeksen gebruikt om deze parameters te bepalen. Dit heeft ten opzichte van de niet-parametrische methode een aantal belangrijke voordelen:

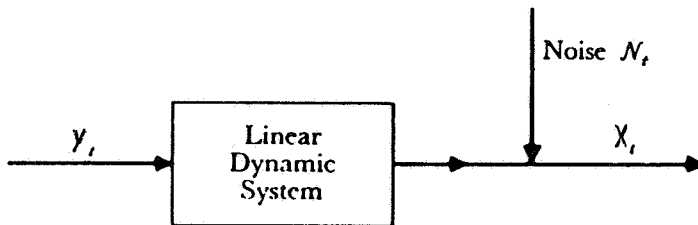
- met de parametrische methode is het veel eenvoudiger om optimale voorspellingen te berekenen. De reden hiervan zal in paragraaf 4 worden behandeld;
- met de parametrische methode is het mogelijk, zij het in beperkte mate, gebruik te maken van bestaande theorieën of wetten waaraan de betrokken tijdreeksen voldoen;
- gebruik makend van de parametrische methode behoeven alleen enkele modelparameters met behulp van de metingen te worden geschat. De betrokken reeksen kunnen hierdoor vrij kort zijn. Bij de niet-parametrische methoden moeten gehele overdrachtsfuncties worden geschat. Hiervoor zijn veel langere reeksen noodzakelijk.

Het gebruik van de spectraalanalyse is echter aanzienlijk eenvoudiger dan de Box en Jenkins aanpak doordat de invloed die de modelbouwer op het eindresultaat kan uitoefenen relatief gering is. Dit resultaat is dan ook minder subjectief dan bij het gebruik van ARIMA-modellen.

De spectraalanalyse houdt zich bezig met problemen waarbij men aan de hand van waarnemingen van een stochastisch proces uitspraken doet over een ander proces dat statistisch met het waargenomen proces samenhangt. We beschouwen hiertoe de volgende situatie. Gegevens twee processen  $X_t$  en  $Y_t$ , welke statistisch afhankelijk zijn volgens figuur 6. We willen nu een schatting van  $X_t$  berekenen door gebruik te maken van een waargenomen realisatie van  $Y_t$ ,  $-\infty < t < \infty$ , en gebruiken hiervoor een lineair tijdsinvariant filter:

$$\hat{X}_t = \int_{-\infty}^{\infty} h(s) Y_{t-s} ds \quad (1)$$

Het bovenstaande filter is niet-causaal, d.w.z. we laten toe dat voor het berekenen van de schatting van  $X_t$  ook toekomstige waarnemingen van  $Y_t$  gebruikt mogen worden.



Figuur 6 Samenhang van de processen  $X_t$  en  $Y_t$



De weekfunctie of impulsresponsie  $h(s)$  is optimaal, in de zin dat

$$E\left\{\left[X_t - \int_{-\infty}^{\infty} h(s)Y_{t-s} ds\right]^2\right\} \quad (2)$$

minimaal is, als  $h(s)$  voldoet aan de volgende integraalvergelijking:

$$r_{xy}(\tau) = \int_{-\infty}^{\infty} h(s)r_y(\tau-s)ds \quad (3)$$

met  $r_y(\tau) = E\{Y_t Y_{t+\tau}\}$  als de autokovariantiefunctie van  $Y_t$  en  $r_{xy}(\tau) = E\{X_t Y_{t+\tau}\}$  als de kruiskovariantie tussen  $X_t$  en  $Y_t$ . De functie geeft de statistische samenhang van de twee processen weer.

De vergelijking (3) is m.b.v. de Fourieranalyse eenvoudig op te lossen. Stel hiertoe:

$$P_y(f) = \int_{-\infty}^{\infty} r_y(\tau) e^{-i2\pi f\tau} d\tau \quad (4)$$

$$P_{xy}(f) = \int_{-\infty}^{\infty} r_{xy}(\tau) e^{-i2\pi f\tau} d\tau \quad (5)$$

Hierbij is  $P_y(f)$  het energiespectrum van  $Y_t$  en wordt  $P_{xy}(f)$  het kruisspectrum van  $X_t$  en  $Y_t$  genoemd. M.b.v. de convolutie stelling uit de Fourieranalyse volgt uit vergelijking (3):

$$P_{xy}(f) = H(f)P_y(f) \quad (6)$$

met

$$H(f) = \int_{-\infty}^{\infty} h(s) e^{-i2\pi fs} ds \quad (7)$$

zodat een schatting van de frekwentieresponsiefunctie  $H(f)$  verkregen kan worden d.m.v.:

$$H(f) = \frac{P_{xy}(f)}{P_y(f)} \quad (8)$$

We hebben hiermee de vergelijking (3) naar het frekwentiedomein getransformeerd. Hierbij dient opgemerkt te worden dat  $P_{xy}(f)$  een complex spectrum is, en dus beschreven wordt door een reëel en een imaginair spectrum (resp. co-spectrum en quadrature spectrum) of door het amplitude en het fase spectrum. Hetzelfde geldt voor  $H(f)$ . Het spectrum  $P_y(f)$  is reëel. De impulsresponsie  $h(s)$  volgt direct uit (7) via een inverse Fouriertransformatie:

$$h(s) = \int_{-\infty}^{\infty} H(f) e^{i2\pi fs} df \quad (9)$$

Met behulp van de bovenstaande methode is het vinden van het optimale niet-causale filter (2) gereduceerd tot het verkrijgen van schatting van  $P_{xy}(f)$  en  $P_y(f)$ . Voor het berekenen van de laatste worden in de klassieke theorie eerst een schatting  $R_\theta(\tau)$  berekend van de kovariantiefunctie  $r_y(\tau)$  uitgaande van een waargenomen realisatie van het proces  $Y_t$  op een interval  $[0, \theta]$ :

$$R_\theta(\tau) = \frac{1}{\theta - \tau} \int_0^{\theta - \tau} Y_{t+\tau} Y_t dt, \quad 0 \leq \tau \leq \theta \quad (10)$$

Vervolgens wordt de schatting  $\Phi_\theta(f)$  voor het spectrum  $P_y(f)$  berekend met behulp van vergelijking (4):

$$\Phi_\theta(f) = \int_{-\theta}^{\theta} R_\theta(\tau) e^{-i2\pi f \tau} d\tau \quad (11)$$

Zonder ons in rekenwerk te storen vermelden we hier dat de schatter (11) niet consistent is (dit houdt ruwweg in dat de nauwkeurigheid bij het gebruik van meer waarnemingen niet groter hoeft te worden). De reden hiervoor kan als volgt worden aangegeven: Naarmate de verschuiving  $\tau$  groter wordt, komen steeds minder gegevens beschikbaar om  $r(\tau)$  te schatten (immers het bruikbare deel van de registratie is steeds  $\theta - \tau$ ). Het gevolg daarvan is dat voor toenemende  $\tau$  de uitkomst van  $R_\theta(\tau)$  een steeds slechtere benadering wordt voor  $r(\tau)$ . De slechte uitkomsten krijgen in de Fourierintegraal (11) een even zwaar gewicht als de nauwkeuriger uitkomsten voor kleine  $\tau$ . Hoe groot  $\theta$  ook gekozen worden, er blijven steeds onnauwkeurige bijdragen in de integraal zodat van consistentie geen sprake is. Deze moeilijkheid kan als volgt worden opgelost. Voor  $\tau \rightarrow \infty$

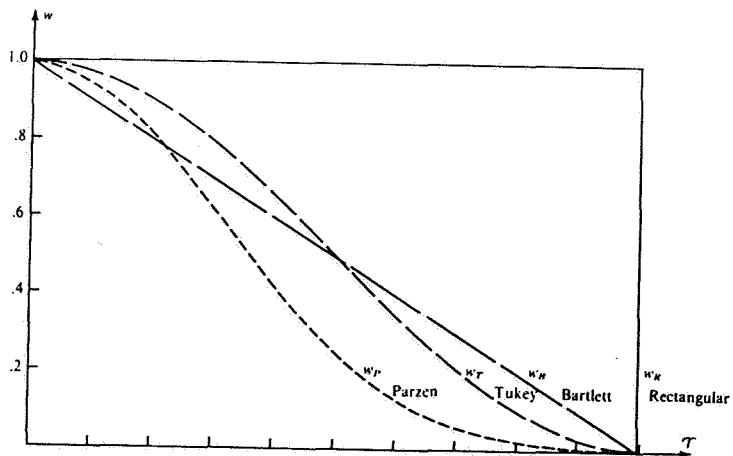
zal de kovariantiefunctie in de meeste gevallen tot nul naderen. Hierdoor kunnen we voor grote waarden van  $\tau$  de bijdragen tot de integraal (11) verwaarlozen en zo de nauwkeurigheid vergroten. We beschouwen daarom schatters van de vorm

$$\Phi_{\theta}(f) = \int_{-\infty}^{\infty} w(\tau) R_{\theta}(\tau) e^{-i2\pi f \tau} d\tau \quad (12)$$

waarin  $w(\tau)$ ,  $-\theta \leq \tau \leq \theta$  een gegeven symmetrische weegfunctie is met:

$$w(\tau) = 0, \quad |\tau| > T, \quad T \ll \theta$$

Enkele van deze weeg- of vensterfuncties zijn getekend in figuur 7.



Figuur 7 Vensterfuncties

We zullen ons hier niet verder verdiepen in de vele consequenties die het gebruik van vensters met zich meebrengt. Hiervoor wordt verwezen naar Jenkins en Watts (1968).

Het filter (1) is niet causaal zodat ook toekomstige waarnemingen van  $Y_t$  bekend moeten zijn voor het bepalen van een schatting van  $X_t$ . Daarom zijn deze filters voor het berekenen van optimale voorspellingen niet geschikt. Hiervoor zou een causaal filter wel gebruikt kunnen worden:

$$\hat{X}_t = \int_{-\infty}^0 g(s) Y_{t-s} ds \quad (13)$$

Bij dit filter worden de toekomstige waarden van  $Y_t$  niet gebruikt voor het bepalen van een schatting van  $X_t$ . De integraalvergelijking waaraan de optimale impulsresponsiefunctie  $g(s)$  voldoet luidt in dit geval:

$$r_{xy}(t) = \int_{-\infty}^0 g(s) r_Y(t-s) ds \quad (14)$$

Deze vergelijking, die de Wiener-Hopf vergelijking wordt genoemd, is veel moeilijker oplosbaar dan vergelijking (3). Daarom is het voor het berekenen van voorspellingen veel eenvoudiger gebruik te maken van de Box en Jenkins aanpak waar de causaliteit gewaarborgd is.

## 5 SLOTOPMERKINGEN

Bij de spectraalanalyse zoals beschreven in paragraaf 4 worden eerst de auto- en kruiskovariantiefuncties bepaald en vervolgens hieruit de spectra. Door de introductie van het venster vòòr het uitvoeren van de Fouriertransformatie kost dit weinig rekentijd. Echter door de vooruitgang die de laatste jaren is gemaakt op het gebied van de numerieke Fouriertransformatie ("Fast Fourier Transform") is het tegenwoordig ook mogelijk om direct de spectra uit de beschikbare meetreeksen te schatten. Deze aanpak wordt daarom de directe methode genoemd, in tegenstelling tot de klassieke aanpak die ook wel de indirecte methode wordt genoemd. Het venster, ook bij de directe methode noodzakelijk om de consistentie van de schatter te waarborgen, wordt nu echter niet in het tijdsdomein

vòòr de Fouriertransformatie, maar in het frekwentiedomein nà het uitvoeren van deze transformatie toegepast.

In dit artikel wordt de spectraalanalyse geïllustreerd aan de hand van het bepalen van de overdrachtsfunctie tussen één invoer- en één uitvoerreeks. Generalisatie van de aanpak naar meerder invoer- en meerdere uitvoerreeksen is echter mogelijk. Hiervoor wordt verwezen naar Jenkins en Watts (1968).

#### LITERATUUR

BOX, G.E.P. en JENKINS, G.M., 1970. Time series analysis: forecasting and control, Holden-Day, 575 p.

JENKINS, G.M. en WATTS, D.G., 1968. Spectral analysis and its applications, Holden-Day, 525 p.

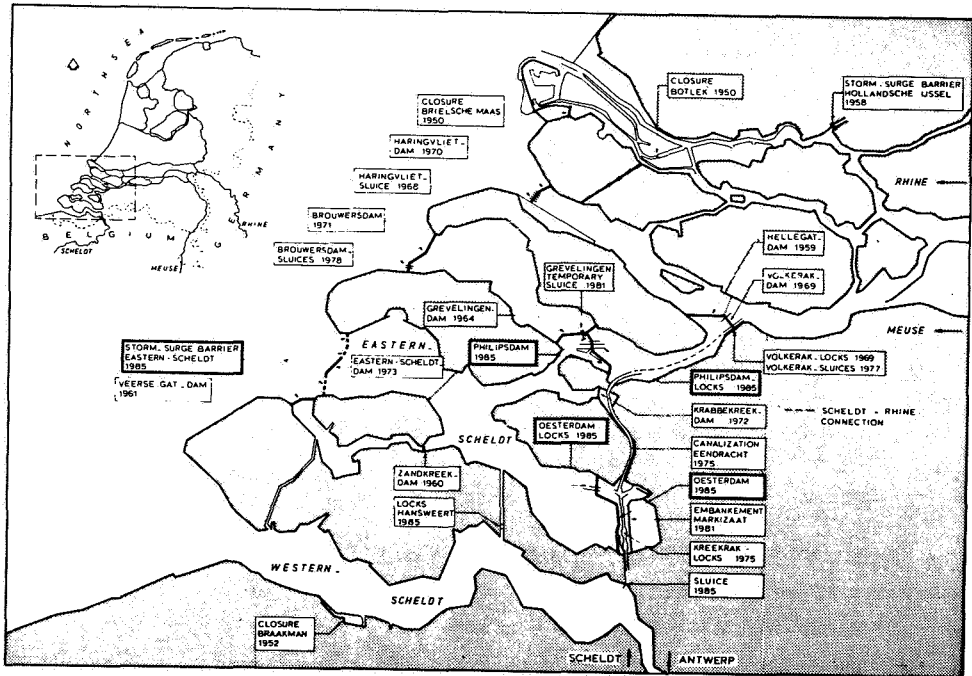
A. Langerak en J.J. Leendertse

### SAMENVATTING

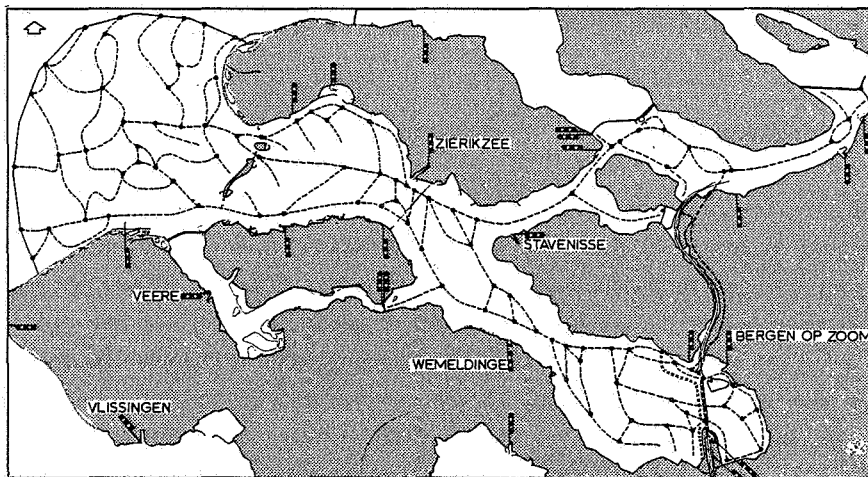
Met de bouw van de getijmodellen voor de Oosterscheldewerken kwam ook de spectraal-analyse in gebruik. Naarmate de modellen complexer werden (van één- naar twee-dimensionaal) groeide de behoefte aan inzicht in het getijverschijnsel om de afregeling van de dure twee-dimensionale modellen te vergemakkelijken. Verder stelden de twee-dimensionale modellen strengere eisen aan de kwaliteit van randvoorwaarden. Bij het krijgen van inzicht in het getijverschijnsel en bij de kwaliteitsverbetering van randvoorwaarden is spectraal-analyse een essentiële hulpmiddel geweest.

### 1 DE OOSTERSCHELDEWERKEN EN DE MODELTECHNIEK

De Oosterscheldewerken vormen in meer dan één opzicht een mijlpaal in de Deltawerken. Na de stormramp van 1953 viel het besluit om een groot aantal zee-gaten te sluiten om zo de kustlijn te bekorten (zie fig. 1). De plannen, die overigens al in de 2<sup>e</sup> wereldoorlog voorbereid werden en die voor een groot deel al voor 1953 gereed waren, voorzagen in een logische opbouw. De Deltawerken zouden van klein naar groot en van eenvoudig naar complex uitgevoerd worden. Lerenderwijs zou men naar het laatste en meest complexe werk, de Oosterscheldewerken, toegroeien. De modeltechniek voor getijden heeft onbewust dezelfde weg gevolgd, waarbij de spectraal-analyse alleen bij de Oosterscheldewerken een ondersteunende rol heeft gespeeld.



Figuur 1 Overzicht van de Deltawerken



Figuur 2 De schematisatie van de Oosterschelde voor het één-dimensionale model

De eerste getijmodellen dateren uit 1920 en vanaf deze tijd is er een gestage ontwikkeling geweest. Lorentz vereenvoudigde de wetten van behoud van momentum en massa door het gebied te schematiseren tot een één-dimensionaal netwerk van kanalen (zie ook fig. 2), waarmee hij ze toegankelijk maakte voor "numerieke" integratie. Een zaal vol ambtenaren, een soort menselijke computer, bepaalde in een maanden durende Sysiphus-arbeid vervolgens hoe de waterstanden en stromingen langs de Noordhollandse en Friese kust zouden veranderen als de Zuiderzee gesloten werd. Hij koos voor deze kostbare aanpak, omdat hij de beschrijving van het getij (de fysica) zo compleet mogelijk wilde houden. Een votum van wantrouwen in de tot toen gebruikte benaderende oplossingen van het probleem. De Rijkswaterstaat heeft deze lijn in de daarop volgende jaren gevolgd. Men bleef vasthouden aan een integrale aanpak van het getij.

De getijmodellen hebben vanaf 1920 twee ontwikkelingen doorgemaakt, nl. een ontwikkeling in de gebruikte apparatuur en een ontwikkeling in de fysische formulering. De apparatuur toonde een ontwikkeling van de menselijke computer van de jaren '30 - '50 via een zelfgebouwde analoge computer, de zgn. Delta-analoge Rekenmachine, naar de digitale machines. Daarnaast werkte men als "apparaat" veel met schaalmodellen (een Madurodam met stromend water). De fysische formulering toonde een ontwikkeling naar een numeriek twee-dimensionaal model dat niet de stroming door een één-dimensionaal stelsel van geulen (netwerk kanalen) berekent, maar dat over het hele gebied op vaste intervallen de diepte-gemiddelde stroming en de getijstanden als functie van de tijd berekent.

Toen de regering besloot om de Oosterschelde niet af te sluiten maar om een halfdoorlatende kering te bouwen, die alleen gesloten zou worden bij storm, (zie fig. 2) voldeden de gebruikte één-dimensionale modellen niet volledig meer. De fysica van de modellen was daarvoor te beperkt. De twee-dimensionale modellen voldeden echter wel en de Rijkswaterstaat heeft een heel scala van twee-dimensionale overzichtsmoellen en detailmodellen voor de Oosterscheldewerken gebouwd (zie fig. 3 en 4), die een aanvulling waren op de bestaande één-dimensionale modellen, omdat ze aangaven hoe het stroompatroon veranderde als een geul geblokkeerd werd.

Met de komst van de twee-dimensionale modellen kwam ook de spectraal-analyse,





als hulpmiddel, in gebruik. Naarmate de modellen complexer werden (van één- naar twee-dimensionaal) groeide de behoefte aan inzicht in het getijverschijnsel om de afregeling van de dure twee-dimensionale modellen te vergemakkelijken. Verder stelden de twee-dimensionale modellen strengere eisen aan de kwaliteit van randvoorwaarden. Bij het krijgen van inzicht in het getijverschijnsel en bij de kwaliteitsverbetering van randvoorwaarden is spectraal-analyse een essentiële hulpmiddel geweest.

## 2 DE FYSICA VAN HET GETIJ

De wisselende aantrekkingskracht van zon en maan en de geometrie van de aarde bepalen de getijverschijnselen langs de Nederlandse kust.

Laplace was de eerste die het getij analyseerde en zijn uitgangspunten zijn nog steeds de grondslagen van de huidige getijanalyse. Hij stelde terecht dat de frekwenties (en de sterkte) van de aandrijvende krachten teruggevonden zouden worden in het getijverschijnsel. De stelling van Laplace heeft twee belangrijke implicaties: de regelmatige baan van zon en maan impliceert dat we alleen diskrete frekwenties in het getijspectrum mogen verwachten; het verschil in aantrekkingskracht tussen maan en zon impliceert dat de zon aanzienlijk minder invloed heeft op het getij dan de maan.

Het aantal astronomische frekwenties in het getij is overigens zeer groot en niet elke frekwentie is in de praktijk even belangrijk. De beweging van zon en maan en de invloed, die zon en maan op de beweging van de aardbaan hebben is zeer gecompliceerd. De bewegingen van de drie hemellichamen veroorzaken daardoor honderden discrete getijfrekwenties, die voor de praktijk een periodiciteit hebben die ligt tussen 18,5 jaar en 1 uur en waarvan de amplitude sterk wisselt. Een harmonische getijanalyse is daardoor een ambachtelijk gebeuren. De analist schrapt zoveel mogelijk frekwenties zonder de weergave van het getij te verstoren.

Niet alleen zon en maan, maar ook de geometrie blijkt frekwenties op te wekken. Dit is te zien aan de hand van de diepte-gemiddelde vergelijkingen voor het behoud van massa en momentum in twee-dimensies, nl.:

$$\frac{\delta U}{\delta t} + U \frac{\delta U}{\delta x} + V \frac{\delta U}{\delta y} + g \frac{\delta h}{\delta x} - fV + g \frac{U \sqrt{U^2 + V^2}}{C^2 H} = 0 \quad \text{Momentum in x-richting}$$

$$\frac{\delta V}{\delta t} + U \frac{\delta V}{\delta x} + V \frac{\delta V}{\delta y} + g \frac{\delta h}{\delta y} + fU + g \frac{V \sqrt{U^2 + V^2}}{C^2 H} = 0 \quad \text{Momentum in y-richting}$$

$$\frac{\delta h}{\delta t} + \frac{\delta HU}{\delta x} + \frac{\delta HV}{\delta y} = 0 \quad \text{Behoud van massa}$$

t : tijd

x,y : Carthesische coördinaten

U,V : snelheid in resp. x- en y-richting

H : waterdiepte

h : waterstand t.o.v. een vast referentievlak

f : Coriolis-parameter

C : Chézy-parameter

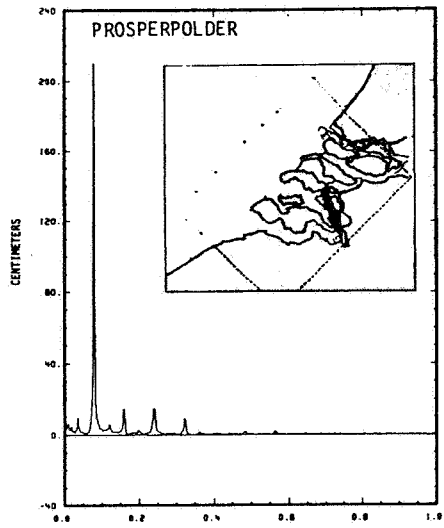
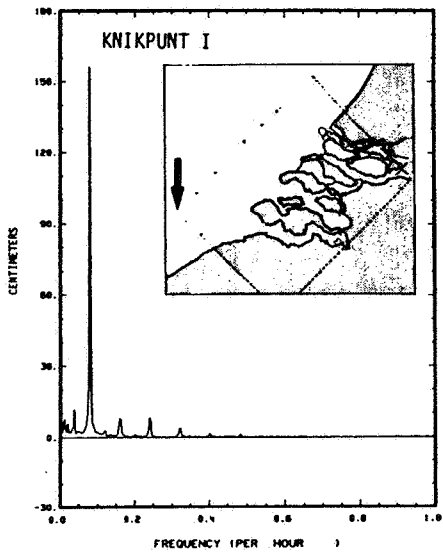
De eerste twee vergelijkingen zijn niet lineair in de 2e, 3e en laatste term. Dit zorgt er voor dat het astronomische zonne- en maangetij som- en verschiltermen opwekken. Uit de praktijk blijkt dat de 2e en 3e term uit het 2x daagse maangetij M2 vooral het 4x daagse M4-getij opwekken en dat de laatste term, de bodemweerstand, uit het 2x daagse M2-getij vooral het 6x daagse M6-getij opwekt (zie ook fig. 5).

### 3 TOEPASSING SPECTRAAL-ANALYSE BIJ VOORBEREIDEND GETIJONDERZOEK

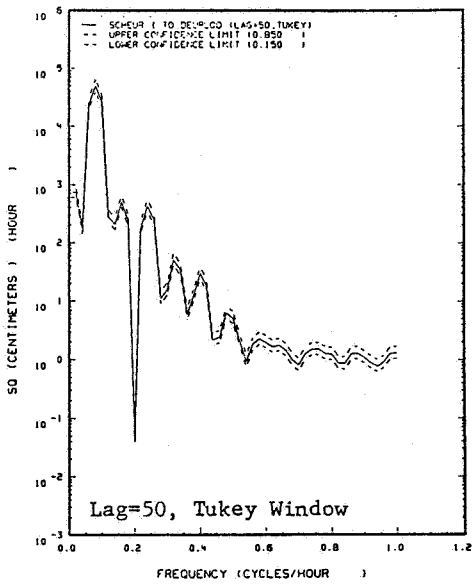
Om een goed getijmodel van een gebied te kunnen bouwen moeten we inzicht in de getijbeweging hebben. Aan de bouw van het eerste getijmodel voor de Oosterscheldewerken ging dan ook een grondige getijanalyse vooraf, waarin we de voortplanting en vervorming van het getij met spectraal-analyse en Fourier-analyse bepaald hebben.

Voor de analyse waren er over 16 dagen metingen in 55 stations beschikbaar. Van de 16 dagen zijn er 15 dagen gebruikt, omdat een periode van 15 dagen veelvouden van de belangrijkste synoptische getijkomponenten bevat.

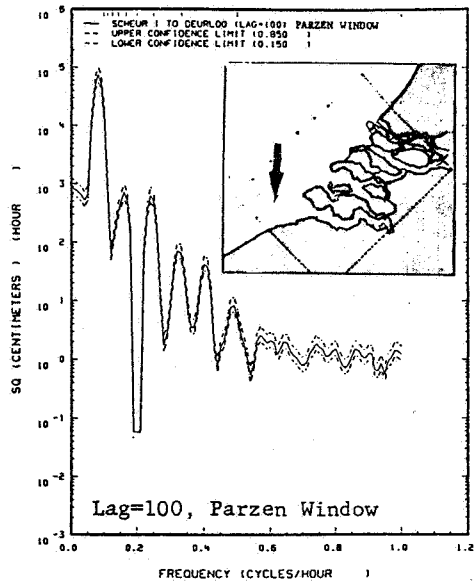
Als eerste stap in de getijanalyse is met een Fourier-analyse een frekwentie/



Figuur 5 De verdeling van de getijcomponenten over de frekwentieband



COMPUTED POWER SPECTRAL ESTIMATES OF THE Y SERIES



COMPUTED POWER SPECTRAL ESTIMATES OF THE Y SERIES

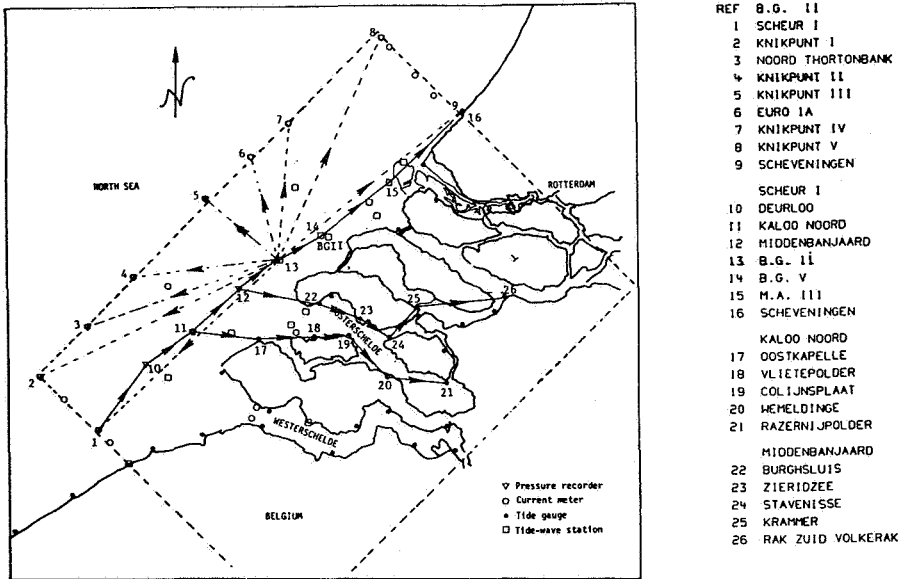
Figuur 6 Verdeling spectrale dichtheid over de frekwentieband bij Tukey-venster en Lag=50 en bij Parzen-venster en Lag=100

amplitudespectrum gemaakt voor de 55 stations. De Fourier-analyse leert (figuur 5) dat het getij maar één dominante frekwentie kent die opgewekt wordt door de maan. Het primaire maangetij (M2) met een frekwentie van 0,08 (periode ca. 12,5 uur) is vaak een orde groter dan de andere componenten. Toch zijn deze kleine componenten, doordat ze sommeren, niet onbelangrijk. Over de cyclus van doodtij (nieuwe maan), springtij (volle maan) veroorzaken ze getijvariatië van 20% en gedurende de dag zorgt de zonnebeweging voor een dagelijkse ongelijkheid in de getij-amplitude van 15%.

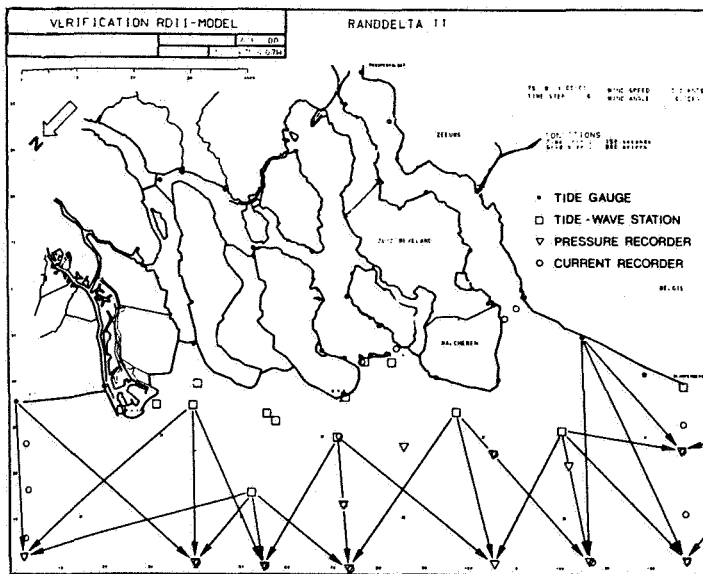
De Fourier-analyse toonde bijv. aan dat het zonnegetij met een golflengte van orde 1000 km (frekwentie  $f=0,04$ ) nauwelijks over het gebied varieert, terwijl wijl de kortere component, het M2-getij met golflengte van 500 km, door resonantieverschijnselen en reflectie wel sterk varieert. Verder zijn de door de niet-lineaire termen opgewekte getijden, zoals M4 ( $f=0,16$ ) en M6 ( $f=0,24$ ) in de estuaria beduidend groter dan op open zee. Dat is op zich niet verwonderlijk. Het M4-getij wordt opgewekt uit het M2 getij door termen als  $U \delta U / \delta x$  die de snelheidsvariatie over het gebied beschrijven en deze variaties zijn groot in een estuarium en klein in de vlakke Noordzee. Het M6 getij wordt opgewekt uit het M2 getij door een term als  $gU \sqrt{U^2 + v^2} / C^2 H$ , en in een estuarium is de diepte H veel geringer dan in de Noordzee terwijl de snelheden veel groter zijn. Deze term neemt daardoor sterk in grootte toe en wekt daarmee een sterker M6-getij op.

Als volgende stap in de getijanalyse is voor opeenvolgende stations de amplificatie en faseverschuiving in het getij bepaald. Hiervoor is spectraal-analyse gebruikt. De voortplanting en vervorming van het getij is vastgelegd door over het gebied een netwerk te leggen (zie fig.7) en door daarna tussen opeenvolgende meetstations de overdrachtsfuncties per frekwentieband te bepalen. Dit legt enerzijds het getijgedrag in de natuur vast en levert anderzijds de kenmerken, waaraan een goed model moet voldoen.

De analyse is niet zonder problemen verlopen. Het getij heeft geen continue min of meer vlakverlopend bandspectrum, maar heeft een discontinue lijnspectrum. Hierdoor luisteren de instelparameters in de spectraal-analyse veel nauwer. Via trial en error is er toch een goed compromis gevonden tussen het oplossend vermogen, de stabiliteit en de nauwkeurigheid van de spectraal-analyse. Het compromis is gevonden door de parameters "lag" en "venster" te optimaliseren.



Figuur 7 Netwerk gebruikt bij het bepalen van de overdrachtsfuncties tussen twee opeenvolgende meetstations



Figuur 8 Overzicht van de correlaties gelegd tussen de meetstations op de rand en de vaste meetstations 10 km uit de rand

Eerst is een lag van 50 gebruikt. Dit gaf goede resultaten, d.w.z. een hoge nauwkeurigheid voor alles behalve het 1x daagse zonnegetij ( $f=0,04$ ). Omdat we vermoedden dat er ook lek optrad van het 2x daags M2-getij naar de andere frekwentiebanden, is de lag verhoogd naar 100. Dit zou samen met een zgn. "Parzen-venster" voldoende beperking van de lek van het M2-getij naar lagere frekwenties geven en bij het zonnegetij zou het resultaat aanmerkelijk verbeteren. Lek van het M2 naar de hogere frekwenties ( $f \gg \pm 0,16$ ) werd door de grotere frekwentie-scheiding niet verwacht. Figuur 6 geeft een vergelijk van een Parzen-venster met 100 lags en een Tukey-venster met 50 lags. Tukey geeft daarbij al een scherpere piek in het dichtheidspectrum, maar mist deels de energie in de band van het zonnegetij ( $f=0,04$ ). Met de gekozen parameters (lag 100, venster="Parzen") is langs diverse lijnen de versterking en faseverschuiving in de diverse getijcomponenten bepaald en vastgelegd.

Uit de spectraal-analyse hebben we twee dingen geleerd:

- Ten eerste waar we de randen van onze detailmodellen niet moesten leggen. In de componenten met een frekwentie hoger dan 0,08 komen in de Oosterschelde knopen en buiken voor. Als we de rand van zo'n model op zo'n knoop leggen dan worden de meetfouten, die altijd in de randvoorwaarde zitten, versterkt waardoor een dergelijke component nooit goed in het model weergegeven zal worden.
- Ten tweede, dat we erg veel aandacht aan de bodemligging in het model moesten besteden om de M4- en in mindere mate ook de M6-komponent goed weer te geven.

#### 4 TOEPASSING SPECTRAAL-ANALYSE BIJ RANDVOORWAARDEN MODELRAND

Randvoorwaarden zijn een essentieel onderdeel van een model. Het getij wordt in feite in de oceanen opgewekt en komt via het Kanaal en de Noordelijke Noordzee onze Zuidelijke Noordzee binnen. Het grootste deel van de getij-energie in de Zuidelijke Noordzee is indirect afkomstig uit het getij in de Atlantische Oceaan en minder dan 1% is afkomstig uit de directe aantrekkingskracht van zon en maan. De randvoorwaarden bepalen daarom samen met de geografie van het gebied het volledige getij in onze modellen.

De randvoorwaarden voor de open modelranden hebben twee problemen opgele-

verd. Ze waren of niet goed genoeg of ze waren er niet. Hoe we spectraal-analyse toegepast hebben om deze problemen op te lossen, zullen we met twee voorbeelden illustreren.

#### Voorbeeld 1

Voor het grootste model voor de Oosterscheldewerken, Randdelta-II, waarin de interactie tussen de Ooster- en de Westerschelde onderzocht is, waren de beschikbare randvoorwaarden niet nauwkeurig genoeg. Rondom een kuststrook van 140 km lang en 30 km breed (fig. 4) waren met onderwaterdrukdozen op een aantal punten metingen uitgevoerd. De drukdozen maten om het halve uur de druk van de waterkolom op de zeebodem, waaruit, na correctie voor de variërende dichtheid van water, de waterstand bepaald kon worden. Helaas maten de drukdozen na zorgvuldige uitwerking de waterstand wel op 5 cm (2 promille) nauwkeurig, maar deze nauwkeurigheid was niet voldoende voor het model. De relatief grote diepte van de kuststrook (25 m), en de geringe bodemwrijving zorgde bij fouten in de waterstand van orde 5 cm al voor grote vereffeningsstromen. De stroming die volgens stroommetingen evenwijdig aan de lange noordwest rand van het model moest lopen vertoonde allerlei rotaties, die in het model soms tot ver in het te onderzoeken gebied doordrongen.

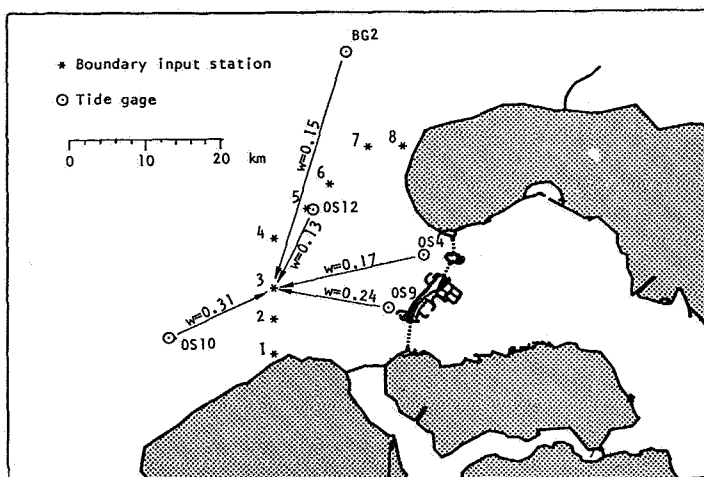
Het probleem is opgelost door correlaties te leggen tussen vaste stations die ca. 10 km van de rand van het model lagen en die met een hoge nauwkeurigheid gemeten waren, en de onnauwkeurige randstations. Het getijsignaal correleert dan wel en de meetfouten niet. Hierdoor kunnen de meetfouten uit het signaal gehaald worden. Elke randpunt is bij de correctie-procedure gecorreleerd met drie vaste stations (zie figuur 8). De terugvoorspelde randvoorwaarde uit elk van de drie vaste stations is bij het regenereren gewogen naar de afstand tussen het vaste station en het randstation. De correctie-procedure heeft uitstekend voldaan.

#### Voorbeeld 2

Voor het overzichtsmodel van de hele Oosterschelde, Oost-II, ontbraken de randvoorwaarden voor het springtij van 11 januari 1982, terwijl er in het tracé van de toekomstige stormvloedkering in de Oosterschelde op die dag een uitgebreide stroommeting en waterstandsmeting uitgevoerd was, die gebruikt zou worden om het model te ijken. Langs de rand van het model waren er op die dag geen metingen en dus geen randvoorwaarden beschikbaar. Wel was het



overzichtsmodel Randdelta-II, dat inmiddels geijkt was, beschikbaar. Dit model voorspelde de waterstanden in het mondingsgebied van de Oosterschelde en dus ook de eventuele randvoorwaarden voor het Oosterschelde-model goed. Helaas waren er voor het Randdelta-model alleen voor de periode 2 t/m 5 september 1975 berekeningen mogelijk en niet voor de periode 11 januari 1982, zodat de resultaten van het Randdelta-model niet rechtstreeks gebruikt konden worden als randvoorwaarden voor het Oosterschelde-model Oost-II. Wel waren er voor de periode van 11 januari 1982 resultaten van vaste meetopstellingen in de buurt van de rand beschikbaar.



Figuur 9 Overzicht van de correlaties gelegd tussen een randstation en de vijf vaste stations ( $w$  = weegfactor)

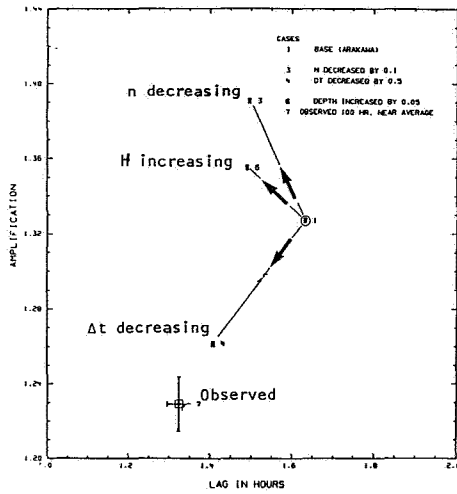
De procedure voor het maken van randvoorwaarden is verder in globale termen gelijk aan de procedure voor Randdelta-II. Eerst zijn er op basis van de waterstanden, die berekend zijn in het Randdelta-II-model voor 2 t/m 5 september, relaties gelegd tussen de waterstanden t.p.v. de toekomstige randpunten van het Oost-II en de waterstanden t.p.v. de vaste opstellingen. Teltkens zijn er tussen de vijf vaste stations en één randpunt correlaties gelegd (fig. 9). Met behulp van de beschikbare correlaties en de metingen die in januari 1982 op de vaste meetstations uitgevoerd zijn, zijn er randvoorwaarden gemaakt voor 11 januari, die goed bleken te voldoen.

Een getijmodel afregelen is, bij gebrek aan voldoende afregelmiddelen, een kwestie van kiezen tussen meerdere kwaden. Het getij bestaat uit een groot aantal diskrete frekwenties, die elk hun eigen fase en amplitude hebben en elke fase en amplitude heeft weer een eigen respons bij variatie van de afregelparameters. Verder bestrijkt een model meestal een groot gebied, waarbinnen het getij sterk kan variëren. Een model zal deze variatie zelden volledig volgen. Met één verantwoorde afregelknop, de bodemwrijving, kan theoretisch van één komponent op één plaats één aspect, n.l. of de fase of de amplitude, tot reproductie gebracht worden.

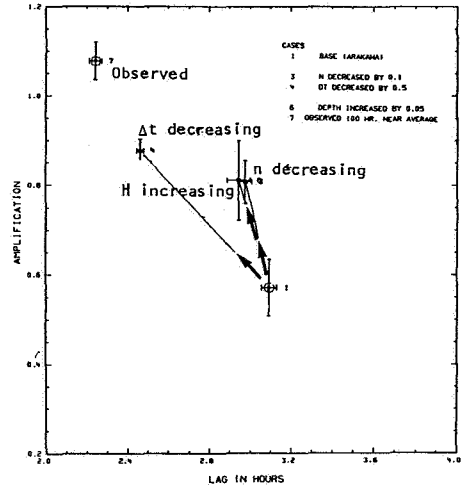
Meestal regelt men onbewust af op de globale fase van het maangetij M2 in het gebied. Dit komt, omdat men achterin het estuarium afregelt op de kleinste variantie tussen model en natuur. Nu blijkt de variantie in de praktijk het kleinst te zijn als de fase van de grootste (M2-)komponent reproduceert. En verder blijkt dat men in de praktijk via het station achterin het estuarium de fase van het maangetij over de rest van het estuarium gemiddeld redelijk afregelt.

Een dergelijke afregelprocedure levert geen informatie over de kwaliteit van het model. Dit geldt met name voor de kwaliteit van de geometrie en de bodemweerstand. Andere afregelparameters horen nauwelijks een rol te spelen. De gebruikte numerieke oplosmethode en de fysische benaderingen in het model zijn zodanig goed dat het model de natuur zeer goed zou moeten reproduceren. Het detail in de bodemweergave en de formulering van de bodemwrijving zijn dan de beperkende factoren.

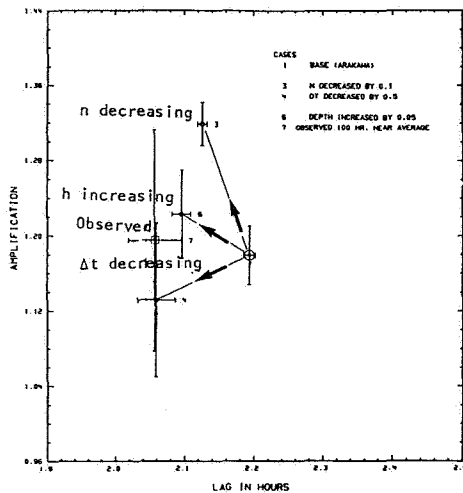
Spectraal-analyse geeft wel inzicht in de kwaliteit van het afgeregelde model. Door het signaal op te splitsen in componenten, waaruit de overdrachtsfuncties van het ene naar het andere meetstation bepaald worden, kan per gebied bepaald worden of de bodemligging niet voldoende gedetailleerd is (M4 reproduceert niet) of dat de bodemwrijving niet goed geformuleerd is (M6 reproduceert niet). De reproductie van het model kan dus met spectraal-analyse gedetailleerd gediagnostiseerd worden. Verder geeft de spectraal-analyse inzicht in de betrouwbaarheid van de metingen en het model, waardoor aan te geven is of verder afregelen van het model zin heeft.



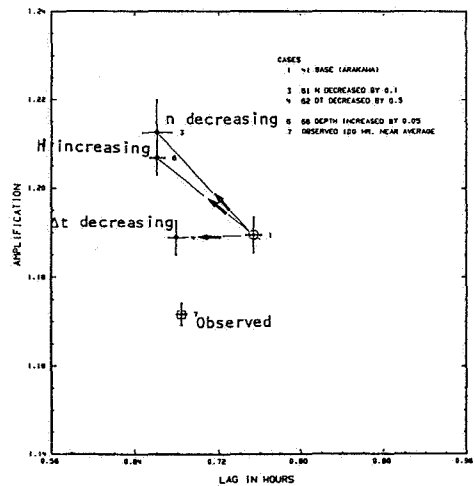
Figuur 10 Amplificatie en fasedraaiing van het M2-getij over de Oosterschelde bij variatie van de tijdstep  $\Delta t$ , diepte H, bodemweerstand n



Figuur 11 Amplificatie en fasedraaiing van het M4-getij over de Oosterschelde bij variatie van de tijdstep  $\Delta t$ , diepte H, bodemweerstand n



Figuur 12 Amplificatie en fasedraaiing van het M6-getij over de Oosterschelde bij variatie van de tijdstep  $\Delta t$ , diepte H, bodemweerstand n



Figuur 13 Amplificatie en fasedraaiing van het M2-getij in de noordelijke tak van de Oosterschelde bij variatie van  $\Delta t$ , H en n

Spectraal-analyse kan ook tijdens de afregeling van het model inzicht geven in de mogelijke verbeteringen. Door in een gevoeligheidsonderzoek alle parameters in het model zoals bijv. bodemligging, bodemweerstand, maar ook de numerieke parameters, zoals de tijdstap van de numerieke integratie, te variëren en door vervolgens via spectraal-analyse de overdrachtsfuncties te bepalen, kan per gebied en per component bepaald worden welke parameter-variëaties tot een best-fit leiden.

Figuur 10 t/m 12 geven een positief voorbeeld van de resultaten van een dergelijk gevoeligheidsonderzoek. Uit de amplitudeversterking en fase draaiing van de M2-komponent blijkt dat vooral de tijdstap van de numerieke integratie te groot is en dus omlaag moet en dat ook de bodemweerstand mogelijk iets verlaagd mag worden, omdat bij deze aanpassingen mede de reproductie van de M4- en de M6-komponent verbetert.

Figuur 13 geeft een negatief voorbeeld van de resultaten van een dergelijk gevoeligheidsonderzoek. Uit de resultaten blijkt dat als de  $\Delta t$  eenmaal volgens figuur 10 t/m 12 aangepast wordt er geen enkele parameter-variëatie tot een best-fit leidt. Dit is gezien het gebied te verwachten. In het gebied van het voorbeeld treden over de diepte grote variaties in de snelheid op, die het gevolg zijn van dichtheidsverschillen tussen het zoete (rivier)water en het zoute zeewater. Het model is diepte-gemiddeld en neemt deze variaties niet mee. Het is in dit gebied dan ook niet toepasbaar.

## 6 RESUMÉ EN SLOTOPMERKINGEN

Spectraal-analyse is een belangrijk middel geweest bij de bouw van de numerieke getijmodellen voor de Oosterscheldewerken. Spectraal-analyse heeft inzicht gegeven in het gedrag van het getij en de kwaliteit van de getijmodellen. Verder is het gebruikt bij het verbeteren en het aanmaken van randvoorwaarden voor de getijmodellen. Spectraal-analyse is daarbij altijd gebruikt ter ondersteuning van een deterministisch model. Door het ontbreken van een fysisch fundament kan de spectraal-analyse techniek vaak niet rechtstreeks worden toegepast bij het maken van voorspellingen. Hierin komt verandering. Momenteel werkt de Rijkswaterstaat (Dienst Informatieverwerking) aan een combinatie van een signaal-analyse-model met een deterministisch model. Met de signaal-analyse (Kalmanfilter) wordt een deterministisch model bijgestuurd om voor de korte termijn betere operationele voorspellingen te maken.

## REFERENTIES

1. Langerak, A, De Ras, M.A.M., Leendertse, J.J.; Adjustment and Verification of the Randdelta II model; Proceedings of the 16th International Conference on Coastal Engineering; Hamburg, 1978
2. Leendertse, J.J., Langerak, A., De Ras, M.A.M.; Two-dimensional Tidal Models for the Delta Works; Proceedings of a Symposium on Predictive Ability of Transport Models for Inland and Coastal Waters; Berkeley, 1981.
3. Leendertse, J.J.; Verification of a Model of the Eastern Scheldt; Randcorporation; Santa Monica, 1984.
4. Leendertse, J.J.; A summary of Experiments with a model of the Eastern Scheldt; Randcorporation, Santa Monica, 1988.
5. Liu, S.K., Leendertse, J.J. and Buckholz, G.; Field Data Analysis - Waterlevels for RDII model, Data Series August 29 - September 12, 1975; Working Note Randcorporation; Santa Monica, 1976.

M.A.J. VAN MONTFORT EN J.V.WITTER

### SAMENVATTING

Uitspraken over toekomstige jaarmaxima kunnen gebaseerd zijn op jaarmaxima en op overschrijdingen van een zekere drempelwaarde. Hierbij optredende limietverdelingen zijn respectievelijk de "Generalised Extreme-Value" (GEV) verdeling en de "Generalised Pareto" (GP) verdeling. Door niet alleen het jaarmaximum, maar jaarlijks de top-r waarden in beschouwing te nemen, dan wel door jaarlijks de twaalf maandmaxima te beschouwen, kan de verdeling van het jaarmaximum nauwkeuriger worden geschat. Hier worden "Maximum Likelihood" (ML) en "Probability-Weighted Moments" (PWM) schatters gegeven voor de GEV verdeling, uitgaande van jaarmaxima, top-r data en maandmaxima, en voor de GP verdeling uitgaande van overschrijdingen. Er wordt ingegaan op veiligheid en herhalingstijd, en op het schatten van de T-jaarherhalingswaarde (kwantiel). Bovendien wordt een kort overzicht gegeven van beschikbare informatie met betrekking tot Nederlandse neerslag-extrema.

## 1 INLEIDING

In de hydrologie is men vaak geïnteresseerd in zelden voorkomende gebeurtenissen die ongewenste gevolgen kunnen hebben voor de mens (te warm: wegen smelten; te nat: wateroverlast; te laag debiet van rivier: te hoge chemicaliën concentratie; etc.). Hier wordt de aandacht gericht op maxima.

De algemene aanpak in de statistiek is de volgende. Maak een model; dit is een abstractie van de werkelijkheid om zo eenvoudig mogelijk het praktisch essentiële te beschrijven. Zo'n model bevat meestal enige onbekende parameters. Onderzoek vervolgens met behulp van data de adequaatheid van het gekozen model, en schat - na het modelfiat van de data - de onbekende parameters, waarmee dan uitspraken voor de toekomst mogelijk zijn.

Zulk model is een mathematische constructie, die zijn waarde ontleent aan de bruikbaarheid van de uitspraken die men ermee kan doen. Een model bevat meestal enige wezenlijk onjuiste of niet-verifieerbare vooronderstellingen, die het in theoretisch opzicht hanteerbaar maken en die hopelijk geen hinderlijke gevolgen hebben bij de toepassing. Voorbeeld: de Gumbel-verdeling is gedefinieerd met als waardenbereik  $-\infty$  tot  $+\infty$ , terwijl in vele correcte toepassingen negatieve waarden onmogelijk zijn.

De eenvoudigste aanpak op weg naar maxima voor de theoreticus ziet er als volgt uit. Beschouw een rij  $x_1, \dots, x_n$  van onderling onafhankelijke trekkingen uit een en dezelfde verdeling  $F(x)$ ; neem het maximum en bestudeer hoe dit zich gedraagt als  $n \rightarrow \infty$ . In de afleidingen speelt de limiet van  $Q(x) = F'(x) \{1-F(x)\} / (F''(x))^2$  voor  $x \rightarrow \infty$  een belangrijke rol. Aangezien voor grote  $x$  geen data voorradig zijn, is het asymptotische gedrag van  $Q(x)$  voer voor theoretici en is het uiteindelijke model hopelijk van nut voor practici. Bovenstaand model rammelt voor jaarmaxima aan alle kanten: onafhankelijkheid?, zelfde verdeling (dus geen binnen-jaar-verloop)?, hoe groot moet  $n$  zijn opdat limietgedrag voldoende goed bereikt is? Het resultaat kan echter bevredigend zijn.

Bij het traditioneel bekend stel uitgangsverdelingen  $F$  behoort - afhankelijk van het gedrag van  $Q(x)$  - één van drie limietverdelingen voor het maximum. Deze verdelingen hebben een plaatsparameter ( $\mu$ ) en een positieve schaalparameter ( $\sigma$ ). Deze drie limietverdelingen kunnen tot één verdeling, de GEV verdeling (Generalised Extreme-Value verdeling), samengevoegd worden door introductie van een vormparameter ( $\theta$ ). Aangezien over  $Q(x)$  voor  $x \rightarrow \infty$  empirisch weinig zinnigs gezegd kan worden, lijkt het zinvol voor beschrijving van extrema uit te gaan van de paraplu-verdeling  $GEV(\mu, \sigma, \theta)$ . Op grond van historische reeksen moet informatie verkregen worden over het type verdeling en de bijbehorende parameters.

Bovenvermelde eenvoudigste aanpak op weg naar jaarmaxima kan op velerlei wijze realistischer gemaakt worden, en steeds zal men bij de data steun zoeken voor het gekozen model. Statistisch gezien komt dit neer op het zoeken naar steun voor modelfouten. Maar bij interesse voor het gedrag van een verdeling in de verre staarten geeft een kleine steekproef weinig informatie.

Stel dat men geïnteresseerd is in de verdeling van het jaarmaximum van een rivierwaterstand en dat men beschikt over metingen op elke dag van 100 jaar, dus 36524 data. Hoe kan men dit databestand zinvol reduceren zonder al te veel verlies aan informatie of, anders gezegd, welke zijn zinvolle extractie-methoden? Een viertal methoden wordt nader uitgewerkt.

(i) Zoek voor elk kalenderjaar het maximum (top-1 genoemd) op. Als nadeel hiervan wordt gevoeld dat de sub-top van het ene jaar hogere waarden kan hebben dan de top (ofwel top-1) van een ander jaar.

(ii) Zoek voor elk kalenderjaar de top- $r$  op (bv. de top-5).

(iii) Zoek voor elke maand het maximum op met als doel om, op basis van de analyses van de 12 afzonderlijke maanden en bij onafhankelijkheid van de maxima in opeenvolgende maanden, de informatie over het maandmaximum in jan, feb, ..., dec te combineren tot informatie over het jaarmaximum.



(iv) Beschouw slechts "grote" waarnemingen. Dit kan op verschillende manieren:

\*1 Kies een drempelwaarde, en noteer de overschrijdingen van die drempel. Van deze overschrijdingen is van belang het aantal per jaar en de grootte van de overschrijdingen. Men kan bovendien uit in de tijd aangesloten groepjes van overschrijdingen slechts de grootste nemen (elk cluster geeft één piek). Deze werkwijze behoort bij het "peaks over threshold" model met vaste drempelwaarde.

\*2 Zet (in gedachten) alle waarnemingen op volgorde van klein naar groot en selecteer daaruit de top-nr bij n jaren en gemiddeld r waarnemingen (overschrijdingen of pieken) per jaar.

Merk op dat bij \*1 het aantal waarnemingen vooraf niet bekend is, en dat bij \*2 de drempelwaarde vooraf niet bekend is.

Bij (iv) is een omweg nodig om tot een maximumverdeling te komen. Bij een gekozen drempelwaarde maakt men een kansmodel voor het jaarlijks aantal periodes met overschrijding (bijv. Poisson verdeling) enerzijds, en voor de groottes van de overschrijdingen (bijv. Exponentiële verdeling) anderzijds. De combinatie Poisson-Exponentieel leidt tot de Gumbel verdeling (d.i. GEV met  $\theta = 0$ ). De combinatie Poisson-GP leidt tot de GEV (hierin staat GP voor Generalised Pareto verdeling). Men kan nu niet alleen het theoretische resultaat aanvaarden, maar ook de opbouw. In dat laatste geval worden de aantallen-verdeling en de groottes-verdeling afzonderlijk bestudeerd.

Met behulp van een model en data komt men vervolgens tot schattingen van de T-jaar-herhalingswaarde (kwantiel) of van kansen, met bijbehorende betrouwbaarheidsuitspraken. Modellen met te veel parameters zullenodeloos lange betrouwbaarheidsintervallen leveren, terwijl te stringente modellen systematische fouten geven. Daarnaast is er het verschijnsel voorspellingsinterval. Als de kans op een rampjaar 0.01 is, is de verwachte wachttijd tot een rampjaar 100 jaar, en zal de wachttijd  $w$  tot het volgende rampjaar hoogstwaarschijnlijk (d.i. met kans ongeveer 0.95) liggen tussen 2 en 368 {met als toelichting  $P[w \leq 2] = 1 - 0.99^2 = 0.020$  en  $P[w \geq 368] = 0.99^{367} = 0.025$ }, of minstens 6 zijn, of hoogstens 299 zijn.

Hier wordt niet ingegaan op onderzoek naar het bestaan van cycli en trends in extrema. Daarvoor wordt verwezen naar Smith (1986), die voor de hoogwaterstanden van Venetië werkt met een model met een verloop binnen het jaar, een cyclische component met een periode van vele jaren en een tweedegraads trend over de jaren.

De opbouw van dit verhaal is als volgt:

Onderdeel 2 koppelt het begrip (verwachte) herhalingstijd, dit is een functie van de te schatten parameters, aan het begrip veiligheid. Als alle parameters bekend zouden zijn, zou de verdeling van de herhalingstijd en daarmee de verwachte herhalingstijd bekend zijn, en kan een voorspellingsinterval voor de herhalingstijd gegeven worden. Er treden nu twee problemen op, te weten: de parameters zijn niet bekend (maar moeten geschat worden) en de verdelingen waarbij die parameters behoren, zijn slechts modellen.

In onderdeel 3 wordt een overzicht gegeven van panklare informatie over Nederlandse neerslagmaxima. Dergelijke informatie over andere hydrologische variabelen is schaarser. Voor dikte van het sneeuwdek kan verwezen worden naar Buishand (1985), voor windsnelheid en -richting naar Wieringa en Rijkoort (1983), voor temperatuur naar Buishand (1988), voor Rijn-afvoeren naar DWW (1984), en voor zeehoogten naar DGW (1986).

In onderdeel 4 wordt bij elk van de vier bovenvermelde extractiemethoden een bijbehorend waarschijnlijkheidsmodel (of kansmodel) gegeven, opdat met behulp van de geselecteerde gegevens schattingen over de verdeling van het jaarmaximum te verkrijgen zijn. Elk kansmodel bevat onbekende parameters, die geschat moeten worden. Gaarne bedient men zich van "beste" schattingsmethoden, waarbij men een criterium moet hebben voor wat "best" is, en waarbij de gecompliceerdheid van het benodigde rekenwerk eventueel nog een rol speelt.

Onderdeel 5 vermeldt bij elk van de extractiemethoden enige momenteel voor de hand liggende schattingswijzen van de onbekende parameters.

In onderdeel 6 wordt ingegaan op het geven van een schatting en van een betrouwbaarheidsinterval voor de T-jaar-herhalingswaarde (kwantiel), overigens zonder in te gaan op modelfouten.

In dit verhaal worden kansvariabelen onderstreept. Om de leesbaarheid te vergroten is de volgende symbolen- en begrippenlijst toegevoegd.

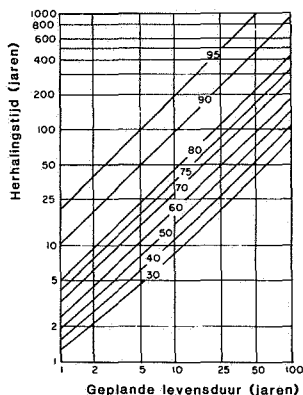
$\mu$	plaatsparameter
$\sigma$	schaalparameter
$\theta$	vormparameter
$\nu$	ondergrens van drager van verdeling
$\lambda$	parameter-vector
GEV	Generalised Extreme-Value verdeling
GEV( $\theta = 0$ )	Gumbel verdeling
GP	Generalised Pareto verdeling
ML	Maximum Likelihood (methode)
$\mathcal{L}$	Likelihood
L	$\log(\text{Likelihood})$
PWM	Probability-Weighted Moment(s)
M	Moment bij PWM methode
P,p	Probability, kans
Pr[...]	De kans op ...
T	(verwachte) herhalingsstijd
$x_p$	linker p-punt (= kwantiel) van verdeling van $\underline{x}$ ; voor $p = 1 - 1/T$ staat hier de T-jaar-herhalings- waarde als $\underline{x}$ het jaarmaximum is
E(.)	verwachting(swaarde)
$\gamma$	constante van Euler ( $\approx 0.577216$ )
$\Gamma(.)$	gamma-functie
n, N	steekproefomvang
$X_{[i:n]}$	i-de grootste waarneming uit steekproef van omvang n
$X_{(i:n)}$	i-de kleinste waarneming uit steekproef van omvang n
F,G	verdelingsfunctie van kansvariabele
f,g	kansdichtheid van kansvariabele
u	drempelwaarde
bias	onzuiverheid van een schattingsmethode (= schatter)
var(.)	variantie

eff(.)                    efficiëntie (quotiënt van schattingsvarianties) van een schattingsmethode ten opzichte van een andere schattingsmethode ter schatting van een parameter (-combinatie).

## 2      VEILIGHEID EN HERHALINGSTIJD

Veiligheid is een relatief begrip; absolute veiligheid bestaat niet. Veiligheid houdt verband met aanvaard(baar) risico, en het risico dat men aanvaardbaar acht ontstaat uit een afweging van kans van optreden enerzijds en de consequenties anderzijds.

Het operationeel maken van het begrip veiligheid vereist het kwantificeren van risico. Hierbij zijn statistische technieken om kansen op extreme gebeurtenissen te schatten onmisbaar. Het praktische probleem wordt hierbij gestileerd en vereenvoudigd. Als men, slechts met uiterst kleine kans, een dijkoverschrijding langs de volle lengte van de dijk aanvaardt, zal men zich alras beperken tot enige punten langs die dijk of sterker nog tot één punt. Aangezien ingenieurs behoefte hebben aan één getal wordt het resultaat van risico-beschouwingen vaak samengevat in de "T-jaar-herhalingswaarde", gedefinieerd als "de grenswaarde, die elk jaar met kans  $1/T$  overschreden wordt". De T-jaar-herhalingswaarde (tot overschrijding) is het  $(1-1/T)$ -kwantiel van een verdeling, en wordt hier wel kortheidshalve aangeduid met kwantiel. Impliciet wordt hier reeds aangenomen dat het systeem in de loop van de tijd niet verandert. In T jaar zal de T-jaar-herhalingswaarde minstens 1 maal overschreden worden met kans  $1 - (1-1/T)^T \approx 1 - e^{-1} = 0.63$  (in  $\frac{1}{2}T$  jaren en  $2T$  jaren is dit resp. 0.39 en 0.86). Hier is tevens impliciet aangenomen, dat resultaten in opeenvolgende jaren onderling onafhankelijk zijn. Voor de kans (P) op geen overschrijding binnen de geplande operationele periode (lengte: n jaren) van een werk, ontworpen met een grenswaarde voor een herhalingsstijd T geldt dan  $P = (1 - 1/T)^n$ . Voor grote T staat hier in benadering  $\ln(T) = -\ln(-\ln P) + \ln(n)$ . Hierop berust de aan Haan (1977) ontleende Figuur 1.



Figuur 1 De kans (%) op geen overschrijding in een periode van gegeven lengte (horizontaal) als functie van de herhalingstijd (vertikaal).

De vraag blijft natuurlijk hoe men erachter komt dat bij een bepaalde grenswaarde een kans op overschrijding in elk jaar behoort van  $1/T$ . In de praktijk is óf de grenswaarde, óf de herhalingstijd onbekend. Stel dat een wegebouwer een wegdek kan produceren dat aan de gestelde eisen voldoet bij temperaturen  $\leq 35^\circ\text{C}$ . Uit KNMI-kringen verneemt hij dat de verdeling van het jaarmaximum van de temperatuur goed benaderd wordt door een Normale verdeling met verwachting  $\mu = 31.5^\circ\text{C}$ , en  $\sigma = 2.0^\circ\text{C}$ . Hieruit volgt voor de kans ( $p_{35+}$ ) op een jaarmaximum  $\geq 35^\circ\text{C}$  als waarde 0.040, dus behoort bij de grens  $35^\circ\text{C}$  naar schatting een herhalingstijd van  $1/0.040 = 25$  jaren. Echter de informatie  $\mu = 31.5^\circ\text{C}$  en  $\sigma = 2.0^\circ\text{C}$  berust op een waarnemingsreeks van 133 jaren, waardoor bij de geschatte kans ad 0.040 een 0.95-betrouwbaarheidsinterval behoort:  $0.02 < p_{35+} < 0.07$ , waaruit volgt  $14 < T < 50$  jaren, zie Levert (1959).

Het geheel wordt voor de ingenieur die slechts één getal wil, nog onaangenamer als de statisticus opmerkt dat de gebruikte Normale verdeling eigenlijk ook nog enigszins in discussie is. Het feit dat voor de jaar-maximale temperatuur een Normale verdeling aangenomen mag worden, is een recent verworven ervaringsresultaat. Vaak zal men de GEV verdeling (Generalised Extreme-Value verdeling) gebruiken waarbij grenswaarde ( $x$ ) en herhalingstijd ( $T$ ) als volgt verbonden zijn

$$x = \mu + \sigma\{1 - (\ln T)^\theta\} / \theta$$

Bij gebruik van de Gumbel-verdeling (dit is de GEV met  $\theta = 0$ ) geldt

$$x = \mu + \sigma\{-\ln(-\ln(1/T))\}$$

Bij niet al te grote T zullen de GEV en Gumbel verdeling niet al te sterk verschillende resultaten geven. Ook hier zullen de parameters door schattingen vervangen worden, en zijn de modellen slechts benaderingen. Bovendien blijven de eisen van onafhankelijkheid (positieve autocorrelatie geeft langere betrouwbaarheidsintervallen) en geen veranderingen in het systeem in de loop van de tijd gehandhaafd.

In bovenstaande is bij de invoering van het begrip T-jaar-herhalingswaarde impliciet overgegaan op jaarmaxima. Het nut van overgaan op het maximum wordt hier nader toegelicht. Als men geïnteresseerd zou zijn in overschrijding van, zeg, een rivierdijk, zou men in alle eenvoud door archiefstudie kunnen vaststellen dat zulks in de afgelopen eeuw op 10 dagen is geschied, derhalve gemiddeld 1 dag met overschrijding per 10 jaar. De archiefstudie zal overigens wel opleveren dat er in die eeuw één aaneengesloten periode van 10 dagen was met overschrijding, derhalve één periode met overschrijding per 100 jaar. Waarschijnlijk zullen de consequenties van "iedere 10 jaar 1 overschrijding" en van "iedere 100 jaar 1 periode van 10 overschrijdingen" verschillend zijn, waardoor de twee situaties duidelijk onderscheiden moeten worden. Dit kan door naar de jaarmaxima van de rivierwaterstand te kijken. In bovenvermeld voorbeeld zal dit maximum in 1, 2, ... óf 10 gevallen boven de hoogte van de rivierdijk liggen met als uitersten 1 jaar met 10 overschrijdingen enerzijds en 10 jaar met elk één overschrijding anderzijds. Interpretatie-problemen ten gevolge van clustering worden hier vermeden door naar jaarmaxima te kijken.

Herhalingstijden kleiner dan één jaar kan men als volgt invoeren. Stel een drempelwaarde, bepaal de beginpunten van clusters van overschrijdingen en beschouw de wachttijd vanaf het begin van een cluster tot het begin van het volgende cluster. De verwachting van zulke wachttijd kan men weer herhalingstijd noemen.

Omdat de situatie in Nederland niet zo gunstig is als in Engeland, waar in het Flood Studies Report (NERC, 1975) een grote hoeveelheid onderzoeksresultaten naar maxima van neerslagen en afvoeren is gebundeld, is het nuttig een overzicht te geven van beschikbare resultaten. Tabel 1, overgenomen uit Buishand (1983), geeft kwantielen van de neerslagverdeling voor verschillende duren, en voor zowel het gehele jaar als voor het winterhalfjaar (oktober t/m maart).

Voor herhalingstijden van 0.5 tot 100 jaar kan Figuur 2 worden gebruikt, overgenomen uit Buishand en Velds (1980). Voor duren korter dan een dag kan Figuur 3, eveneens overgenomen uit Buishand en Velds (1980), worden gebruikt.

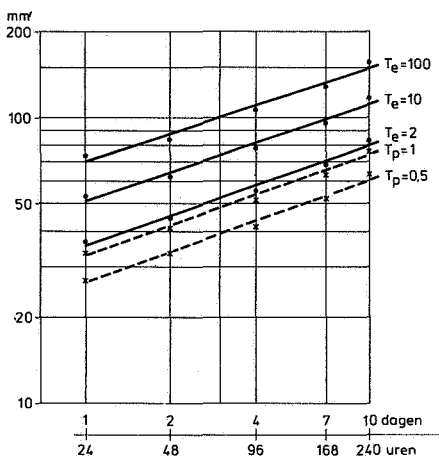
In de hieronder genoemde resultaten is uitgegaan van de GEV verdeling als verdeling voor jaarmaxima. Men kan ook historische reeksen ordenen, en dan zonder aanpassing van een verdeling procentpunten bepalen. Dit is voor dagneerslagen voor 140 neerslagstations met reeksen voor de periode 1951-1980 gebeurd in KNMI (1984). Tabel 2 geeft een overzicht van de resultaten.

Tabel 1 Maxima van de neerslaghoeveelheid (mm) die voor verschillende duren gemiddeld eens per T jaar worden overschreden (overgenomen uit Buishand, 1983).

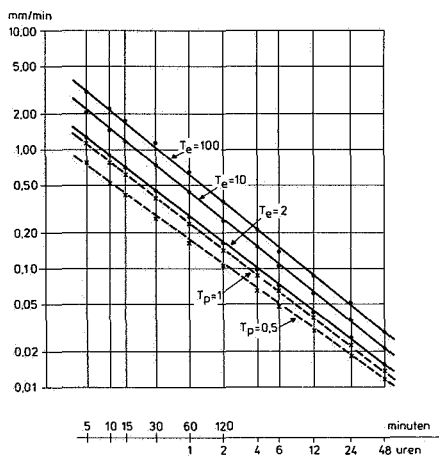
Herhalingstijd T (jaar)	Duur D (dagen)				
	1	2	4	7	10
Gehele jaar:					
100	68	79	97	118	140
200	75	86	104	125	148
500	86	96	113	134	158
1000	95	104	120	141	165
Winterhalfjaar:					
100	52	66	85	105	123
200	58	72	92	111	130
500	67	80	100	119	139
1000	73	86	106	125	145

Tabel 2 Procentpunten (95%  $\approx$  gemiddeld 18 keer per jaar, 98% 7 keer, en 99% 4 keer) van de dagneerslag (mm), overgenomen uit KNMI (1984). Tussen haakjes is de standaardafwijking vermeld, en tussen accolades het aantal beschouwde stations in een regio.

Procent- punt	Nederland	Regio 1 {28}	Regio 2 {24}	Regio 3 {88}	De Bilt
95%	10.4 (0.5)	11.1 (0.4)	9.9 (0.2)	10.3 (0.3)	10.7
98%	15.6 (0.7)	16.5 (0.5)	14.9 (0.4)	15.4 (0.4)	15.4
99%	19.7 (0.8)	20.7 (0.8)	18.9 (0.5)	19.6 (0.5)	20.4



Figuur 2 Extreem grote neerslagen voor De Bilt (1906-1977) voor duren van 1 tot en met 10 dagen (overgenomen uit Buishand en Velds (1980)).  $T_p$  is de herhalingstijd, gebaseerd op partiële reeksen;  $T_e$  is de herhalingstijd gebaseerd op reeksen van jaarmaxima.

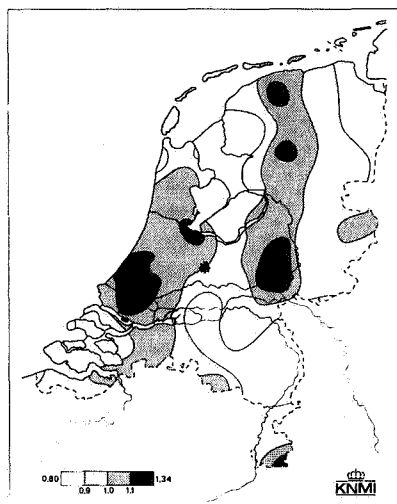


Figuur 3 Extreem hoge neerslagintensiteiten voor De Bilt (1906-1977) voor duren van 5 minuten tot en met 48 uren (overgenomen uit Buishand en Velds (1980)). Voor de symbolen  $T_e$  en  $T_p$  wordt verwezen naar Figuur 2.

In Tabel 2 is een regionale opsplitsing gemaakt. Deze opsplitsing, gebaseerd op de gemiddelde jaarneerslag, is statistisch significant (Witter, 1984). In Buishand (1987) wordt een factor  $f_p$  afgeleid waar-



mee de berekende overstortingsfrequentie uit een stippengrafiek voor De Bilt herleid kan worden tot een theoretische overstortingsfrequentie voor een plaats elders in Nederland, zie Figuur 4. In een stippengrafiek wordt voor een bepaalde reeks van pluviograafgegevens de neerslaghoeveelheid uitgezet tegen de regenduur. Met zulk een grafiek worden overstortingsfrequenties afgeleid voor gemengde rioolstelsels bij gegeven beschikbare berging en pompoevercapaciteit, zie bijvoorbeeld Koot (1977). Alle stations met  $f_p > 1.1$  behoren tot de natte stations, en de meeste met  $f_p < 0.9$  tot de droge stations volgens de opsplitsing in Witter (1984).



Figuur 4. Factor waarmee de berekende overstortingsfrequentie uit een stippengrafiek voor De Bilt herleid kan worden tot een stippengrafiek voor een plaats elders in Nederland (overgenomen uit Buishand (1987)). De factor geldt voor een riolering met een berging van 7 mm en een pompoevercapaciteit van 0.7 mm/h. Het station De Bilt is met een ster aangegeven.

Statistische gebiedsreductiefactoren, waarmee puntneerslagen met een bepaalde herhalingstijd kunnen worden herleid tot gebiedsneerslagen met dezelfde herhalingstijd, kunnen voor verschillende herhalingstijden voor uur- en dagneerslagen worden gevonden in Witter (1984).

## 4.1    Jaarmaxima (top-1)

Veronderstel dat de waarnemingen  $x_1, \dots, x_n$  beschouwd mogen worden als onderling onafhankelijke trekkingen uit een uitgangsverdeling  $F$ . Deze zal in het geval van neerslaghoeveelheden op natte dagen veelal redelijk benaderd kunnen worden met een verschoven gammaverdeling (Buishand, 1977). Als er een limietverdeling  $G$  van het maximum,  $\underline{x}_{[1:n]}$ , bestaat dan geldt dat deze een van de drie mogelijke vormen van de GEV verdeling is, volgens

$$G(x) = \begin{cases} \exp[-\exp(-z)], & -\infty < z < \infty; \theta = 0 \text{ (GEV-1)} & (4.1a) \\ \exp[-(1-\theta z)^{1/\theta}], & 1-\theta z > 0; \theta < 0 \text{ (GEV-2)} & (4.1b) \\ \exp[-(1-\theta z)^{1/\theta}], & 1-\theta z > 0; \theta > 0 \text{ (GEV-3)} & (4.1c) \end{cases}$$

met  $z = (x-\mu)/\sigma$ , waarbij  $\mu$  een plaatsparameter,  $\sigma$  een positieve schaalparameter, en  $\theta$  een vormparameter is. Voor  $\theta < 0$  heeft de GEV verdeling een ondergrens  $\mu+\sigma/\theta$ , voor  $\theta > 0$  een bovengrens  $\mu+\sigma/\theta$ . Op basis van  $m$  jaren waarnemingen kan nu per jaar het maximum worden bepaald. Vervolgens wordt aan deze maxima de verdelingsfunctie volgens Vergelijking 4.1 aangepast. De inverse functie luidt:

$$x = \begin{cases} \mu+\sigma z & ; \theta = 0 & (4.2a) \\ \mu+\sigma \frac{1-\exp(-\theta z)}{\theta} & ; \theta \neq 0 & (4.2b) \end{cases}$$

waarbij  $z$  de gereduceerde variabele is volgens:

$$z = -\ln\{-\ln G(x)\} = -\ln\{-\ln(1-1/T)\} \quad (4.3)$$

met  $T$  de verwachte herhalingstijd (tot overschrijding). De relatie tussen  $x$  en  $z$  is afgebeeld in Figuur 5 voor  $\theta < 0$ ,  $\theta = 0$ , en  $\theta > 0$ . De verschillen zijn vooral groot voor grote herhalingstijden.

Vergelijking 4.2 is van belang bij het schatten van kwantielen, ook wel  $T$ -jaar-herhalingswaarden genoemd.

## 4.2 Top-r

In plaats van alleen de jaarmaxima te bepalen, kan jaarlijks de top-r

$$x_{[1:n]}^j \geq \dots \geq x_{[r:n]}^j, \quad (\text{jaarindex: } j=1, \dots, m) \text{ worden bepaald. Als}$$

de maxima een limietverdeling  $G(x)$  hebben, is de limietverdeling van de r-de grootste waarde (Leadbetter et al., 1983):

$$G(x) = \sum_{s=0}^{r-1} \{-\ln G(x)\}^{s/\theta} \quad (4.4)$$

Voor de simultane kansdichtheid  $f(x_{[1:n]}, \dots, x_{[r:n]}; \mu, \sigma, \theta)$  geldt (Smith, 1986):

$$f(x_{[1:n]}, \dots, x_{[r:n]}; \mu, \sigma, \theta) = \sigma^{-r} \exp\left\{-\left[1-\theta\left(\frac{x_{[r:n]}^{-\mu}}{\sigma}\right)\right]^{1/\theta} + \left(\frac{1}{\theta} - 1\right) \sum_{\ell=1}^r \ln\left[1-\theta\left(\frac{x_{[\ell:n]}^{-\mu}}{\sigma}\right)\right]\right\} \quad (4.5)$$

## 4.3 Maandmaxima

De parameters van de GEV verdeling van het jaarmaximum kunnen ook worden geschat op grond van de gecombineerde maandmaxima van een jaar. Als de limietverdeling voor het maximum van maand  $i$  wordt aangeduid met  $G_i(x)$ , dan geldt voor de verdeling  $G(x)$  van het jaarmaximum bij onafhankelijkheid van de maandmaxima

$$G(x) = \prod_{i=1}^{12} G_i(x) \quad (4.6)$$

Deze verdeling is (slechts) een GEV verdeling indien  $\theta = 0$  en  $\sigma_i = \sigma$ , of indien  $\theta_i = \theta$ ,  $\sigma_i = \sigma$ , en  $\mu_i = \mu$ . In dat geval bestaat er een expliciete uitdrukking voor de inverse verdelingsfunctie; in het andere geval vindt men de inverse met behulp van numerieke methoden.

Om het aantal parameters enigszins in de hand te houden, kan voor het seizoensverloop een eenvoudig parametrisch model worden gekozen.

Bijvoorbeeld kunnen  $\sigma_i/\mu_i$  en  $\theta_i$  constant worden verondersteld.

#### 4.4 Overschrijdingen van een drempelwaarde

Een laatste beschouwingwijze gaat uit van een drempelwaarde ( $u$ ) en beschouwt het aantal en de grootte  $X-u$  van overschrijdingen van deze drempelwaarde. Het jaarlijks aantal van zulke overschrijdingen is dus een toevalsvariabele. Indien dit aantal Poisson verdeeld is, geldt voor de conditionele limietverdeling  $G_U(x)$  van  $X-u$ , gegeven  $X \geq u$ , dat deze, mits de limietverdeling bestaat, een van de drie mogelijke vormen van de GP verdeling is (Pickands, 1975)

$$G_U(x) = \begin{cases} 1 - \exp(-z), & 0 < z < \infty; \theta = 0 & \text{(GP-1)} & (4.7a) \\ 1 - (1 - \theta z)^{1/\theta}, & 0 < z < \infty; \theta < 0 & \text{(GP-2)} & (4.7b) \\ 1 - (1 - \theta z)^{1/\theta}, & 0 < z < 1/\theta; \theta > 0 & \text{(GP-3)} & (4.7c) \end{cases}$$

met  $z = (x - \nu)/\sigma$ , waarbij  $\nu$  een plaatsparameter,  $\sigma$  een positieve schaalparameter en  $\theta$  een vormparameter is. De verdeling heeft als ondergrens  $\nu$ , en voor  $\theta > 0$  een bovengrens  $\nu + \sigma/\theta$ .

De overschrijdingen komen meestal in clusters voor. In de hydrologie wordt vaak uitgegaan van de pieken van dergelijke clusters. Men noemt dit het 'Peaks over Threshold' model (NERC, 1975). Reeksen van dergelijke pieken noemt men wel partiële reeksen. Er zijn twee mogelijke benaderingen, c.q. extractiemodellen. Bij de ene wordt uitgegaan van een vaste drempelwaarde  $u$  (methode 1), waarbij het jaarlijks aantal overschrijdingen dus een toevalsvariabele is; deze kan gemodelleerd worden met bijvoorbeeld de Poisson verdeling. Bij de andere wordt uitgegaan van een vast aantal pieken (methode 2). Bij methode 2 is dus de drempelwaarde  $u$  vooraf niet bekend. De GP verdeling blijkt voor beide typen van partiële reeksen goed te voldoen, zie bijvoorbeeld Van Montfort en Witter (1986) voor een toepassing van de GP verdeling op neerslagreeksen. Merk op dat de GP verdeling een veralgemening is van de Exponentiële verdeling. In het Flood Studies Report (NERC, 1975) is veel gebruik gemaakt van partiële reeksen waaraan Exponentiële verdelingen zijn aangepast.

De inverse verdelingsfunctie van de GP verdeling luidt:

$$x = \begin{cases} \nu + \sigma z & ; \theta = 0 \\ \nu + \sigma \frac{1 - \exp(-\theta z)}{\theta} & ; \theta \neq 0 \end{cases} \quad (4.8a)$$

met  $z = -\ln(1-G_U(x))$ . Ook hier geldt, evenals bij de GEV verdeling, dat de relatie van  $x$  en  $z$  wordt weergegeven door een concave, lineaire of convexe curve voor  $\theta$  resp.  $<0$ ,  $=0$  en  $>0$ . Met vergelijking 4.8 kunnen kwantielen worden berekend.

## 5 SCHATTINGSMETHODEN VAN DE MODELPARAMETERS

Voor het aanpassen van een verdelingsfunctie aan een historische reeks van jaarmaxima (top-1 of top- $r$ ), maandmaxima, of pieken, bestaat een groot aantal methoden. Doel is om eenvoudig goede schattingen van de parameters van de verdelingsfunctie te verkrijgen. Goed wordt dan meestal opgevat in de zin van zuiverheid van de schattingsmethode met als aanvullende eis minimale schattingsvariantie. Berekeningsgemak is tegenwoordig nauwelijks meer een zinnig beoordelingscriterium van schattingsmethoden. Dientengevolge blijven hier grafische methoden en de momentenmethode buiten beschouwing. Grafische methoden blijven overigens nuttig om snel een indruk van het gegevensmateriaal te krijgen (parameter en kwantiel schatting) en om modelfouten op te sporen.

Hier worden twee schattingsmethoden beschouwd: de Methode van de Grootste Aannemelijkheid (ML: Maximum Likelihood), en de Methode van de Kansgewogen Momenten (PWM: Probability-Weighted Moments).

ML-schatting berust bij onafhankelijke waarnemingen uit één en dezelfde verdeling op het volgende. Ga uit van een steekproef  $x_1, x_2, \dots, x_n$  aan een kansvariabele  $\underline{x}$  met verdelingsfunctie  $G(x, \lambda)$  en kansdichtheid  $g(x, \lambda)$ , waarbij  $\lambda$  een parametervector is, bijvoorbeeld  $(\mu, \sigma, \theta)'$ . Dan is de aannemelijkheidsfunctie (Likelihood)  $\mathcal{L}$  gedefinieerd als het product van de kansdichtheden:

$$\mathcal{L} = \prod_{i=1}^n g(x_i, \lambda) \quad (5.1)$$

De waarde van  $\lambda$ , waarvoor  $\mathcal{L}$  (en dus ook  $L = \ln \mathcal{L}$ ) maximaal is, is de ML-schatting voor  $\lambda$ . Deze wordt doorgaans gevonden door  $\partial L / \partial \lambda$  nul te stellen. De ML-schattingsmethode is een standaardmethode in de statistiek, met een aantal aantrekkelijke asymptotische eigenschappen:

- (a) consistentie (asymptotisch raak), dat wil zeggen dat door opvoeren van de steekproefomvang het verschil (of de afstand) tussen  $\hat{\lambda}_n$  en de te schatten  $\lambda$  willekeurig klein te maken is.
- (b) Asymptotische efficiëntie, dat wil zeggen minimale schattingsvariantie voor  $n \rightarrow \infty$ .
- (c) Multivariate normale verdeling van de parameterschatters voor  $n \rightarrow \infty$ .

Het feit dat ML asymptotisch best is, houdt niet in dat er voor kleine steekproeven niet betere schattingsmethoden zouden bestaan.

PWM-schattingen zijn pas vrij recent geïntroduceerd vanuit de hydrologie (Greenwood et al., 1979); parameters worden geschat op basis van enige lineaire combinaties van order statistics. Bij zuivere schatting van de PWM-s wordt slechts gebruik gemaakt van  $\Sigma x_{(i)}$ ,  $\Sigma i x_{(i)}$ ,  $\Sigma i^2 x_{(i)}$ , ... met  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ ; dit zijn de zogenaamde order statistics. Een goed overzicht biedt Hosking (1986). De methode gaat uit van kansgewogen momenten  $M_{p,r,s}$ , gedefinieerd als:

$$M_{p,r,s} = E[\underline{x}^p \{G(\underline{x})\}^r \{1-G(\underline{x})\}^s] = \int_{G=0}^1 x^p \{G(x)\}^r \{1-G(x)\}^s dG(x) \quad (5.2)$$

Voor  $r = s = 0$  staan hier de traditionele momenten. Invullen van  $G$  geeft een uitdrukking voor  $M$  als functie van de parameters van  $G$ . Indien  $r$  en  $s$  positieve, gehele getallen zijn, geldt (David, 1981; p.34):

$$M_{p,r,s} = \frac{r!s!}{(r+s+1)!} E\left\{ \underline{x}_{(r+1:r+s+1)}^p \right\} \quad (5.3)$$

Het kansgewogen moment  $M_{p,r,s}$  kan dus uitgedrukt worden als de verwachting van een macht van een order statistic uit een steekproef van nader te specificeren omvang, te weten macht  $p$  van de  $(r+1)$ -de order statistic uit een steekproef van omvang  $r+s+1$ . Bij beschikbare

steekproeven van omvang groter dan  $r+s+1$ , beschouwt men voor schatting van  $M$  alle verschillende daaruit te trekken deelsteekproeven, en middelt daarover. Door vervolgens schattingen van  $M_{p,r,s}$  gelijk te stellen aan de bijbehorende functies van de parameters, kunnen de parameters worden geschat. Bij PWM-schatting neemt men veelal  $p=1$  waardoor slechts de eerste macht van de order statistics gebruikt wordt. Voorts kiest men, zonder dat dit een beperking blijkt te zijn,  $\delta r = 0$  of  $s = 0$ .

Uit de definitie van  $M_{prs} = E[x^p \underline{G}^r (1-\underline{G})^s]$  volgt o.a.

$$M_{100} = E[x]$$

$$M_{101} = E[x(1-\underline{G})] = E[x] - E[x\underline{G}] = M_{100} - M_{110}$$

$$M_{102} = E[x(1-\underline{G})^2] = E[x - 2x\underline{G} + x\underline{G}^2] = M_{100} - 2M_{110} + M_{120}$$

Vaak wordt gewerkt met  $M_{1,r,0}$ , genoteerd als  $\beta_r$ . Afgeleid kan worden dat

$$\underline{b}_0 = \bar{x} \tag{5.4a}$$

$$\underline{b}_r = 1/n \sum_{i=r+1}^n x_{(i)} \cdot \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} \quad \text{voor } r = 1, 2, \dots \tag{5.4b}$$

een zuivere schatter is voor  $\beta_r$  op basis van de geordende steekproef  $x_{(1)} \leq \dots \leq x_{(n)}$ . Deze schatter wordt in het vervolg gebruikt.

De onzuiverheid (= bias) van de schatter ( $\underline{S}$ ) werd hier nul gesteld. Noem de variantie van de schatter  $\text{var}_S$ , dan geldt voor het kwadraat van de onnauwkeurigheid van de schatter:  $\text{MSE} = \text{bias}^2 + \text{var}_S$  (met MSE voor mean squared error). Om MSE klein te krijgen heeft  $\text{bias}_S$  niet nul te zijn. Hosking et al. (1985) bereiken verkleining van de MSE door  $\beta_r = E(x\underline{G}^r)$  te schatten met behulp van de zogenaamde plotting positions. Op waarschijnlijkheidspapier kan men  $x_{(i)}$  op de variabele-schaal uitzetten tegen de plotting position  $\hat{\beta}_i = (i+\alpha)/(n+1+2\alpha)$  op de kansschaal. Hosking et al. (1985) tonen aan dat

$$\hat{\beta}_r = 1/n \sum_{i=1}^n x_{(i)} \hat{\beta}_i^r$$

met  $\hat{\beta}_i = (i-\frac{1}{2})/n$  (, dus  $\alpha = -\frac{1}{2}$ ,) een verbetering is ten opzichte van Vergelijking 5.4. Sinclair en Ahmad (1988) stellen  $\alpha = \{(n^2+1)^{\frac{1}{2}} - (n+1)\}/2$  voor. In dit verhaal wordt verder slechts aandacht besteed aan de op zuiverheid berustende PWM-schatters volgens Vergelijking 5.4.

Voor de verschillende extractiemethoden zullen nu ML- en PWM-schattingen van de parameters worden gegeven.

### 5.1 Jaarmaxima (top-1)

Voor  $GEV(\theta=0)$ , ook wel Gumbel verdeling genoemd, volgt uit  $G(x) =$

$\exp[-\exp(-\frac{x-\mu}{\sigma})]$  voor de kansdichtheid  $g(x)$ :

$$g(x) = \frac{1}{\sigma} \exp(-\frac{x-\mu}{\sigma}) \cdot \exp[-\exp(-\frac{x-\mu}{\sigma})] \quad (5.5)$$

Dus geldt voor de log likelihood L:

$$L = -n \ln \sigma - \sum_{i=1}^n \frac{x_i - \mu}{\sigma} - \sum_{i=1}^n \exp(-\frac{x_i - \mu}{\sigma})$$

Uit  $\partial L / \partial \sigma = \partial L / \partial \mu = 0$  volgt (van Montfort, 1969):

$$\phi(\sigma) = \sigma - \sum x_i / n + S_1 / S_0 = 0 \quad (5.6a)$$

$$\mu = \sigma \ln(n / S_0) \quad (5.6b)$$

waarin

$$S_k = \sum x_i^k \exp(-x_i / \sigma).$$

Vergelijking 5.6a is op te lossen met de iteratieve Newton-Raphson methode. Er geldt:

$$\phi'(\sigma) = 1 + (S_0 S_2 - S_1^2) / (\sigma^2 S_0^2) \quad (5.7)$$

Voor  $GEV(\theta \neq 0)$  volgt uit  $G(x) = \exp[-(1-\theta \frac{x-\mu}{\sigma})^{1/\theta}]$  voor de kansdichtheid  $g(x)$ :

$$g(x) = \frac{1}{\sigma} \exp\{-[1-\theta \frac{x-\mu}{\sigma}]^{1/\theta} + (\frac{1}{\theta} - 1) \ln[1-\theta \frac{x-\mu}{\sigma}]\} \quad (5.8)$$

Dus geldt voor de log likelihood L:

$$L = \sum_{i=1}^n \{-\ln \sigma - (1-\theta) y_i - e^{-y_i}\}$$

na de herparametrisering:  $z = (x-\mu)/\sigma$ ;  $t = \theta z$ ;  $y = 1 - \frac{1}{\theta} \ln(1-t)$



=  $z(1+t/2+t^2/3+\dots)$ . Voor de berekening van  $L'$  en  $L''$  zijn de volgende afgeleiden nuttig:

$$y_\mu = \frac{-1}{\sigma(1-t)}; \quad y_\sigma = \frac{-z}{\sigma(1-t)}$$

$$y_\theta = \frac{z}{\theta(1-t)} - \frac{y}{\theta} = z^2\left(\frac{1}{2} + \frac{2}{3}t + \frac{3}{4}t^2 + \dots\right)$$

$$y_{\mu\mu} = \frac{\theta}{\sigma^2(1-t)^2}; \quad y_{\mu\sigma} = \frac{1}{\sigma^2(1-t)^2}; \quad y_{\sigma\sigma} = \frac{z(2-t)}{\sigma^2(1-t)^2}$$

$$y_{\mu\theta} = \frac{-z}{\sigma(1-t)^2}; \quad y_{\sigma\theta} = \frac{-z^2}{\sigma(1-t)^2}$$

$$y_{\theta\theta} = \frac{1}{\theta^2} \left[ \frac{zt}{(1-t)^2} - \frac{2z}{1-t} + 2y \right] = \frac{z^2}{\theta(1-t)^2} - \frac{2}{\theta} y_\theta =$$

$$= z^3 \left[ \frac{2}{3} + \frac{6}{4}t + \frac{12}{5}t^2 + \frac{20}{6}t^3 + \dots \right]$$

Dan geldt (met weglating van het somteken):

$$L' = (-1 + \theta + e^{-y}) \begin{bmatrix} y_\mu \\ y_\sigma \\ y_\theta \end{bmatrix} + \begin{bmatrix} 0 \\ -1/\sigma \\ y \end{bmatrix}$$

en

$$L'' = (-1 + \theta + e^{-y})y'' - e^{-y} \begin{bmatrix} y_\mu \\ y_\sigma \\ y_\theta \end{bmatrix} (y_\mu y_\sigma y_\theta) + \begin{bmatrix} 0 & 0 & y_\mu \\ 0 & 1/\sigma^2 & y_\sigma \\ y_\mu & y_\sigma & 2y_\theta \end{bmatrix}$$

met

$$y'' = \begin{bmatrix} y_{\mu\mu} & & \text{symm} \\ y_{\mu\sigma} & y_{\sigma\sigma} & \\ y_{\mu\theta} & y_{\sigma\theta} & y_{\theta\theta} \end{bmatrix}$$

De parametervector  $\lambda = (\mu, \sigma, \theta)'$  wordt iteratief verbeterd volgens  $\lambda_{j+1} = \lambda_j + L_j'(-L_j'')^{-1}$ , waarin  $j$  het volgnummer van de iteratieronde

is. Als beginschatting  $\lambda_0$  kan gebruikt worden de ML-schatting voor  $\mu$ ,  $\sigma$  bij  $\theta = 0$  (Jenkinson, 1977). Als stopcriterium kan bijv. gebruikt

worden

$$\sum_p \{(\hat{x}_p)_{j+1} - (\hat{x}_p)_j\}^2 / (x_{(n)} - x_{(1)})^2$$

met  $p = 0.01(0.49) 0.99$  en  $\hat{x}_p$  het geschatte  $p$ -punt.

Daar  $\mu + \sigma/\theta$  de eindige grens van de drager is, ontstaan problemen als na 'correctie' van de parameters geldt  $x_{(1)} < (\hat{\mu} + \hat{\sigma}/\hat{\theta})_{j+1} < x_{(n)}$ . Dan dient de correctie verkleind te worden, zodat de uiterste waarnemingen binnen de geschatte drager blijven. Bovenstaande afleiding voor GEV is overgenomen uit Van Montfort (1982).

ML-schatters zijn doorgaans onzuiver. Biascorrectie als functie van de steekproefomvang  $n$  is te vinden in Hosking (1985) voor GEV ( $\theta = 0$ ); Hosking et al. (1985) kwantificeren de bias voor de GEV in een simulatiestudie.

PWM-schatters voor GEV ( $\theta = 0$ ) en GEV ( $\theta \neq 0$ ) zijn afgeleid in Hosking (1986). Voor GEV ( $\theta = 0$ ) geldt:

$$\sigma = (2\beta_1 - \beta_0)/\ln 2; \mu = \beta_0 - \gamma\sigma \quad (5.9)$$

waarin  $\gamma$  de constante van Euler ( $\approx 0.577216$ ) voorstelt. Vervolgens worden  $\beta_0$  en  $\beta_1$  zuiver geschat op grond van Vergelijking 5.4 en ingevuld in Vergelijking 5.9. Zo worden zuivere schatters voor  $\sigma$  en  $\mu$  verkregen, evenals voor alle lineaire combinaties ervan zoals T-jaarherhalingswaarden.

Voor de asymptotische efficiëntie van deze PWM-schatters ten opzichte van ML-schatters geldt  $\text{eff}(\hat{\mu}) \approx 0.9963$  en  $\text{eff}(\hat{\sigma}) \approx 0.7555$ .

Voor GEV ( $\theta \neq 0$ ) geldt dat de vormparameter  $\theta$  geschat kan worden uit de vergelijking

$$(3b_2 - b_0)/(2b_1 - b_0) = (1 - 3^{-\hat{\theta}})/(1 - 2^{-\hat{\theta}}) \quad (5.10)$$

Hosking et al. (1985) suggereren als benadering van de oplossing

$$\hat{\theta} = 7.8590 c + 2.9554 c^2 \quad (5.11)$$

waarin

$$c = (2b_1 - b_0)/(3b_2 - b_0) - \ln 2/\ln 3$$

waarbij  $|\hat{\theta}_{PVM}-\hat{\theta}| < 0.0009$  voor  $|\hat{\theta}_{PVM}| < \frac{1}{2}$ .

Als de vraag zou zijn of de data steun geven aan de opmerking dat een Gumbel verdeling, dus  $\theta = 0$ , een te eenvoudige veronderstelling is, ligt hier een toets voor het oprapen. Hosking et al. (1985) melden dat  $\underline{\theta}(1.775n)^{\frac{1}{2}}$  voor Gumbeldata nagenoeg standaard Normaal verdeeld is. Als  $\underline{\theta}(1.775n)^{\frac{1}{2}}$  bijvoorbeeld kleiner dan  $-1.645$  zou uitvallen, is er overtuigend steun in de data om  $\theta < 0$  te stellen.

Voor de twee andere parameters geldt vervolgens

$$\hat{\sigma} = (2b_1 - b_0)\hat{\theta} / \{(1 - 2^{-\hat{\theta}})\Gamma(1 + \hat{\theta})\} \quad (5.12a)$$

$$\hat{\mu} = b_0 - \hat{\sigma}\{1 - \Gamma(1 + \hat{\theta})\} / \hat{\theta} \quad (5.12b)$$

waarin  $\Gamma(\cdot)$  staat voor de gamma-functie. Voor  $n < 100$  blijken PWM-schatters te verkiezen te zijn boven ML-schatters (Hosking et al., 1985).

Een voorbeeld kan de toepassing van PWM-schattingen verduidelijken. Sinclair and Ahmad (1988) vermelden 16 jaarmaxima van het debiet ( $m^3s^{-1}$ ) van de Annan te Brijdekirk (1967-1982).

Van klein naar groot zijn de waarnemingen ( $x_1 \leq x_2 \leq \dots \leq x_{16}$ ):

150.5	214.5	250.6	256.1	256.3	257.4	260.9	263.6
268.4	285.7	306.5	307.4	308.5	309.2	453.3	474.0

Er geldt ( $n = 16$ ) voor de GEV-fit:

$$b_0 = \sum_{i=1}^n x_i / n = 288.93125$$

$$b_1 = \sum_{i=2}^n (i-1)x_i / \{n(n-1)\} = 164.78958 \quad ; \quad 2b_1 - b_0 = 40.64791$$

$$b_2 = \sum_{i=3}^n (i-1)(i-2)x_i / \{n(n-1)(n-2)\} = 118.20155; \quad 3b_2 - b_0 = 65.67340$$

$$c = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\ln(2)}{\ln(3)} = -0.01199; \hat{\theta} \approx -0.0938$$

$$(3b_2 - b_0)/(2b_1 - b_0) = (1 - 3^{-\hat{\theta}})/(1 - 2^{-\hat{\theta}}) \rightarrow \hat{\theta} = -0.094;$$

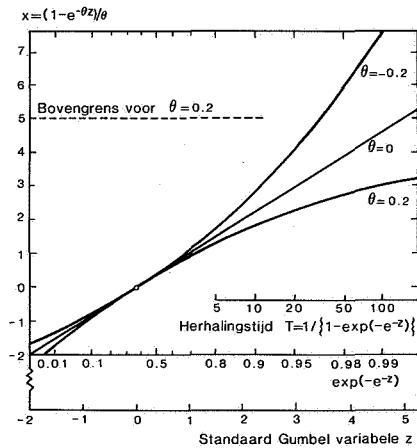
$$\hat{\theta}(1.775n)^{\frac{1}{2}} = -0.50 \quad (\text{met eenzijdige overschrijdingskans: } 0.31);$$

$$(2b_1 - b_0)/\{\Gamma(1 + \hat{\theta})(1 - 2^{-\hat{\theta}})/\hat{\theta}\} = \hat{\sigma} = 53.36 \text{ met } \Gamma(1 + \hat{\theta}) = 1.0638;$$

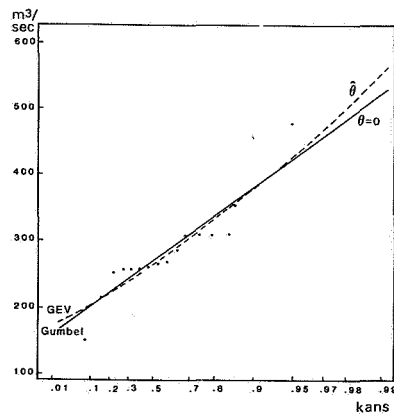
$$b_0 - \hat{\sigma}\{1 - \Gamma(1 + \hat{\theta})\}/\hat{\theta} = \mu = 252.71;$$

$$x_{0.99} = \mu + \sigma\{1 - (-\ln 0.99)^{\hat{\theta}}\}/\hat{\theta}, \text{ en } \hat{x}_{0.99} = 599.8.$$

Voor de Gumbel-fit wordt gevonden  $\hat{\sigma} = (2b_1 - b_0)/\ln 2 = 58.64$ ;  $\mu = b_0 - \gamma\hat{\sigma} = 255.08$  en  $\hat{x}_{0.99} = 524.9$ . Voor een grafische voorstelling van de aangepaste verdelingen wordt verwezen naar Figuur 6.



Figuur 5 Relatie tussen kwantielen  $x$  en de gereduceerde variabele  $z$  voor de GEV verdeling (overgenomen uit Buishand, 1986).



Figuur 6 Gumbelplot van Annan-data met aangepaste GEV ( $\mu = 252.71$ ;  $\sigma = 53.36$ ;  $\theta = -0.094$ ) en Gumbel verdeling = GEV ( $\mu = 255.08$ ;  $\sigma = 58.64$ ;  $\theta = 0$ ).

## 5.2 Top-r

Op grond van de simultane kansdichtheid volgens Vergelijking 4.5 voor de top-r waarnemingen  $x_{[1:n]} \geq \dots \geq x_{[r:n]}$  aan een GEV verdeling, kan voor de likelihood op grond van m jaren waarnemingen worden afgeleid (met  $x_{[r:n]}^j$  de r-de grootste in jaar j):

$$\begin{aligned} \mathcal{L}(\mu, \sigma, \theta) = & \sigma^{-mr} \exp\left\{-\sum_{j=1}^m \left[1-\theta \frac{x_{[r:n]}^j}{\sigma}\right]^{1/\theta}\right. \\ & \left.- \left(\frac{1}{\theta} - 1\right) \sum_{\ell=1}^r \ln\left[1-\theta \frac{x_{[\ell:n]}^j}{\sigma}\right]\right\} \end{aligned} \quad (5.13)$$

Maximalisatie van  $\mathcal{L}$  uit Vergelijking 5.13 met numerieke methoden leidt tot de ML-schatting van  $\mu, \sigma, \theta$ .

Voor  $\text{GEV}(\theta = 0)$  luidt de likelihood:

$$(\mu, \sigma) = \sigma^{-mr} \exp\left\{-\sum_{j=1}^m \left\{\exp\left(-\frac{x_{[r:n]}^j}{\sigma}\right) + \sum_{\ell=1}^r \frac{x_{[\ell:n]}^j}{\sigma}\right\}\right\} \quad (5.14)$$

Naast deze formule geeft Gomes (1981) ook de ML-schattingsformules. Met

$$\bar{x} = \frac{1}{mr} \sum_{j=1}^m \sum_{k=1}^r x_{[k:n]}^j$$

$$S_{\ell} = \sum_{j=1}^m z_{\ell}^j \exp(-z_j/\sigma) \text{ met } z_j = x_{[r:n]}^j$$

vindt zij voor de ML-vergelijkingen

$$\begin{aligned} \phi(\sigma) = S_1 - (\bar{x}-\sigma)S_0 = 0, \quad \rightarrow \phi'(\sigma) = S_0 + \{S_2 - (\bar{x}-\sigma)S_1\}/\sigma^2; \\ \mu = \sigma \ln(mr/S_0) \end{aligned}$$

Eerst wordt met behulp van de Newton-Raphson methode  $\sigma$  geschat (op basis van een beginschatting voor  $\sigma$ , bijvoorbeeld uit de top-1 data), en vervolgens  $\mu$ .

Buishand (1988) vergelijkt voor de Gumbel verdeling de asymptotische standaardfouten van ML-schatters van  $\mu$  en  $\sigma$  voor verschillende waarden van  $r$  in het interval  $[1,10]$ . Voor  $r = 10$  is de relatieve standaardfout (t.o.v.  $r=1$ ) van  $\underline{\mu}$  0.74, en van  $\underline{\sigma}$  0.39.

In de open literatuur zijn nog geen PWM-schatters voor de GEV verdeling met behulp van top- $r$  data beschreven.

### 5.3 Maandmaxima

In het geval van maandmaxima kunnen aan de afzonderlijke reeksen van maandmaxima met behulp van de in Hoofdstuk 5.1 of 5.2 beschreven methoden GEV verdelingen worden aangepast. De verdeling voor de jaarmaxima volgt uit Vergelijking 4.6. Als algemene opmerking kan vermeld worden, dat kijken naar maandmaxima meer loont naarmate meer maanden in het jaar in aanmerking komen om het jaarmaximum te leveren. Men zou het jaar kunnen opsplitsen in kwartalen, maanden of weken; naarmate de deelperiode korter is zal minder voldaan zijn aan de benodigde veronderstellingen voor asymptotisch gedrag, en kunnen opeenvolgende maxima sterker afhankelijk zijn.

### 5.4 Peaks over threshold model

Voor GP ( $\theta = 0$ ), dit is de gegeneraliseerde Pareto verdeling met  $\theta = 0$  ofwel de Exponentiële verdeling, is bij bekende drempelwaarde  $\nu$  (Methode 1) de voor bias gecorrigeerde ML-schatting van  $\sigma$  (NERC, 1975):

$$\hat{\sigma} = n(\bar{x} - \nu)/(n-1) \quad (5.15)$$

waarin  $\bar{x}$  het gemiddelde van de  $n$  pieken is.

Indien de drempelwaarde niet bekend is (Methode 2) geldt:

$$\hat{\sigma} = n(\bar{x} - x_{[n]})/(n-1) \text{ en } \hat{\nu} = x_{[n]} - \hat{\sigma}/n \quad (5.16)$$

waarin  $x_{[n]}$  de laagste piek in de steekproef is. De gebruikte schatters zijn eveneens gecorrigeerd voor bias.

Voor GP ( $\theta \neq 0$ ) volgt uit  $G(x) = 1 - (1 - \theta \frac{x-\nu}{\sigma})^{1/\theta}$  voor de kansdichtheid  $g(x)$ :

$$g(x) = \frac{1}{\sigma} \exp\left\{\left(\frac{1}{\theta} - 1\right) \ln\left[1 - \theta \frac{x-\nu}{\sigma}\right]\right\} \quad (5.17)$$

Dus geldt voor de log likelihood  $L$ :

$$L = \sum_{i=1}^n \{- \ln \sigma - (1-\theta)y_i\}$$

na de herparametrisering  $z = (x-\nu)/\sigma$ ,  $t = \theta z$ , en  $y = -\frac{1}{\theta} \ln(1-t)$

$= z(1+t/2+t^2/3+\dots)$ . Voor de berekening van  $L'$  en  $L''$  zijn de volgende afgeleiden nuttig (Van Montfort en Witter, 1985):

$$\begin{aligned} y_{\sigma} &= -z/[\sigma(1-t)], \\ y_{\theta} &= z/[\theta(1-t)] - y/\theta \quad (= z^2[\frac{1}{2} + 2t/3 + 3t^2/4 + \dots] \text{ voor kleine } t) \\ y_{\sigma\sigma} &= z(2-t)/[\sigma^2(1-t)^2] \\ y_{\theta\theta} &= z^2/[\theta(1-t)^2] - 2y_{\theta}/\theta \\ &= z^3[2/3 + 6t/4 + 12t^2/5 + 20t^3/6 + \dots] \text{ voor kleine } t) \\ y_{\sigma\theta} &= -z^2/[\sigma(1-t)^2] \end{aligned}$$

Dan geldt (met weglating van het somteken):

$$\begin{aligned} L' &= (\theta-1) \begin{pmatrix} y_{\sigma} \\ y_{\theta} \end{pmatrix} + \begin{pmatrix} -1/\sigma \\ y \end{pmatrix} \\ \text{en} \\ L'' &= (\theta-1) \begin{pmatrix} y_{\sigma\sigma} & y_{\sigma\theta} \\ y_{\sigma\theta} & y_{\theta\theta} \end{pmatrix} + \begin{pmatrix} 1/\sigma^2 & y_{\sigma} \\ y_{\sigma} & 2y_{\theta} \end{pmatrix} \end{aligned}$$

De parametervector  $\lambda = (\sigma, \theta)'$  wordt, iteratief verbeterd volgens

$$\lambda_{j+1} = \lambda_j + L_j' (-L_j'')^{-1}, \text{ waarin } j \text{ het volgnummer van de iteratieronde is.}$$

Als beginschatting  $\lambda_0$  kan de ML-schatting voor  $\sigma$  worden gebruikt bij  $\theta = 0$ . Evenals bij ML-schatting van GEV parameters kunnen hier problemen ontstaan bij een eindige grens van de drager. Correcties voor bias van de ML-schattingen zijn te vinden in Van Montfort en Witter (1986).

Indien wordt uitgegaan van een vast aantal pieken (Methode 2), leidt bovenstaande schattingsmethode, gegeneraliseerd tot het drie-parameter

geval met parameters  $\nu$ ,  $\sigma$  en  $\theta$  vaak tot numerieke problemen. Aangeraden wordt in dat geval  $\theta$  en  $\sigma$  te schatten zoals hierboven is aangegeven, en  $\nu$  te schatten met Vergelijking 5.16.

Voor GP ( $\theta = 0$ ) vallen de PWM-schatters samen met de voor bias gecorrigeerde ML-schatters (Hosking, 1986). Voor GP ( $\theta \neq 0$ ) zijn de PWM-schatters bij Methode 1 (Hosking, 1986):

$$\hat{\theta} = \underline{b}_0 / (2\underline{b}_1 - \underline{b}_0) - 2 \quad \text{en} \quad \hat{\sigma} = (1 + \hat{\theta})\underline{b}_0 \quad (5.18)$$

Asymptotische efficiënties worden in Hosking (1986) en Hosking and Wallis (1987) gegeven. Voor  $\theta > -0.1$  zijn conventionele momentenmethode schatters asymptotisch efficiënter dan de PWM-schatters.

Bij Methode 2 kunnen de parameters als volgt geschat worden (Hosking, 1986):

$$\hat{\theta} = \frac{n(b_0 - x_{[n]}) - 2(n-1)(2b_1 - b_0)}{(n-1)(2b_1 - b_0) - (b_0 - x_{[n]})} \quad (5.19a)$$

$$\hat{\sigma} = (1 + \hat{\theta})(2 + \hat{\theta})(2b_1 - b_0) \quad (5.19b)$$

$$\hat{\nu} = x_{[n]} - \hat{\sigma} / (n + \hat{\theta}) \quad (5.19c)$$

De asymptotische varianties van  $\hat{\sigma}$  en  $\hat{\theta}$  komen overeen met die bij Methode 1.

## 6 SCHATTEN VAN DE T-JAAR-HERHALINGSWAARDE

### 6.1 Bij top-1 data

Hier wordt bij top-1 data aandacht geschonken aan het schatten van de T-jaar-herhalingswaarde, genoteerd als  $x_p$ , en het geven van een betrouwbaarheidsinterval voor  $x_p$ .



### 6.1.1 PWM-schatting bij de GEV verdeling

Hosking et al. (1985) geven als benadering voor grote steekproefomvang voor de covariantiematrix van de PWM-schatters van de GEV parameters  $(\mu, \sigma, \theta)$

$$C = \frac{1}{n} \begin{bmatrix} \sigma^2 w_{11} & \sigma^2 w_{12} & \sigma w_{13} \\ \sigma^2 w_{12} & \sigma^2 w_{22} & \sigma w_{23} \\ \sigma w_{13} & \sigma w_{23} & w_{33} \end{bmatrix} \quad (6.1)$$

waarbij de w-waarden in tabelvorm gegeven zijn, zie Tabel 3.

De interesse gaat uit naar de T-jaar-herhalingswaarde, met linkerkans  $p = 1 - 1/T$ , waarvoor geldt

$$x_p = \mu + \sigma \frac{1 - (-\ln p)^\theta}{\theta}$$

Een benadering voor  $\text{var}(\hat{x}_p)$  wordt met behulp van eerste afgeleiden gevonden:

$$d_1 = \frac{\partial}{\partial \mu} x_p = 1; \quad d_2 = \frac{\partial}{\partial \sigma} x_p = \frac{1 - (-\ln p)^\theta}{\theta};$$

$$d_3 = \frac{\partial}{\partial \theta} x_p = \sigma \left\{ \frac{-\ln(-\ln p) \cdot (-\ln p)^\theta}{\theta} - \frac{1 - (-\ln p)^\theta}{\theta^2} \right\}$$

Tabel 3. Elementen van de asymptotische covariantiematrix in Vergelijking 6.1 van de PWM schatters van de GEV parameters.

$\theta$	$w_{11}$	$w_{12}$	$w_{13}$	$w_{22}$	$w_{23}$	$w_{33}$
-.4	1.6637	1.3355	1.1405	1.8461	1.1628	2.9092
-.3	1.4153	.8912	.5640	1.2574	.4442	1.4090
-.2	1.3322	.6727	.3926	1.0013	.2697	.9139
-.1	1.2915	.5104	.3245	.8440	.2240	.6815
.0	1.2686	.3704	.2992	.7390	.2247	.5633
.1	1.2551	.2411	.2966	.6708	.2447	.5103
.2	1.2474	.1177	.3081	.6330	.2728	.5021
.3	1.2438	-.0023	.3297	.6223	.3033	.5294
.4	1.2433	-.1205	.3592	.6368	.3329	.5880

De benadering is dan

$$\text{var}(\hat{x}_p) \approx \sum_{i=1}^3 \sum_{j=1}^3 d_i d_j c_{ij} = \sum_{i=1}^3 d_i^2 c_{ii} + 2 \sum_{i>j} d_i d_j c_{ij} \quad (6.2)$$

waarin  $c_{ij}$  de elementen zijn van bovengenoemde covariantiematrix. Vergelijking 6.2 volgt uit de benaderingsformules in Kendall en Stuart (1977) voor de variantie van een functie van kansvariabelen.

Een  $(1-\alpha)$ -betrouwbaarheidsinterval voor  $x_p$  volgt uit

$$|\hat{x}_p - x_p| < u_{\alpha/2} \cdot \{\hat{\text{var}}(\hat{x}_p)\}^{1/2}$$

waarbij  $u_{\alpha/2}$  het rechter  $\alpha/2$ -punt is uit de standaard Normale verdeling; in  $\hat{\text{var}}(\hat{x}_p)$  moeten de schattingen voor  $\sigma$  en  $\theta$  ingevuld worden.

Een voorbeeld. Stel dat voor de Annan-data als PWM schattingen gevonden worden  $\hat{\mu} = 253$ ,  $\hat{\sigma} = 53$  en  $\hat{\theta} = -0.1$  bij  $n = 16$  (wel erg klein voor asymptotische formules). Stel  $T = 10$  jaar, dus  $p = 0.9$ ; dan vindt men

$$\hat{x}_p = 253 + 53 \cdot \frac{1 - (-\ln 0.9)^{-0.1}}{-0.1} = 386.7$$

$$d_1 = 1; d_2 = 2.5237; d_3 = -156.14$$

$$w_{11} = 1.2915, \dots, w_{33} = 0.6815; \hat{\text{var}}(\hat{x}_p) = 41.7^2$$

$$0.95\text{-betrouwbaarheidsinterval: } 386.7 \pm 1.96 \times 41.7: 305 < x_{0.9} < 468.$$

### 6.1.2 ML-schatting bij de GEV verdeling

Als men in plaats van PWM als schattingsmethode ML kiest, verandert slechts de asymptotische covariantiematrix. Jenkinson (1977) geeft daarvoor Tabel 4.

Tabel 4 Elementen van de asymptotische covariantiematrix van de ML-schatters van de GEV parameters.

$\theta$	$w_{11}$	$w_{12}$	$w_{13}$	$w_{22}$	$w_{23}$	$w_{33}$
-0.6	1.289	1.019	0.233	1.367	-0.250	1.048
-0.4	1.291	0.799	0.264	1.046	-0.089	0.835
-0.2	1.278	0.571	0.273	0.806	0.045	0.645
0	1.249	0.337	0.258	0.653	0.147	0.477
0.2	1.202	0.092	0.220	0.584	0.214	0.330
0.4	1.141	-0.164	0.162	0.595	0.242	0.206
0.6	1.077	-0.433	0.094	0.679	0.225	0.104

Bijbehorende interpolatie-formules van Jenkinson (1977) zijn:

$$w_{11} = 1.2489 - 0.1943 \theta - 0.2239 \theta^2 + 0.0486 \theta^3 + 0.1136 \theta^4$$

$$w_{12} = 0.3366 - 1.1976 \theta - 0.1193 \theta^2 - 0.0347 \theta^3 - 0.0047 \theta^4$$

$$w_{13} = 0.2581 - 0.1359 \theta - 0.2970 \theta^2 + 0.0556 \theta^3 + 0.0947 \theta^4$$

$$w_{22} = 0.6428 - 0.5547 \theta + 1.0634 \theta^2 - 0.0521 \theta^3 - 0.0971 \theta^4$$

$$w_{23} = 0.1470 + 0.4272 \theta - 0.4389 \theta^2 - 0.0868 \theta^3 - 0.0118 \theta^4$$

$$w_{33} = 0.4768 - 0.7866 \theta + 0.2705 \theta^2$$

### 6.1.3 PWM-schatting bij $GEV(\theta = 0)$

Nu geldt  $x_p = \mu + \sigma z$  met  $z = -\ln(-\ln p)$

$$d_1 = \frac{\partial}{\partial \mu} x_p = 1; \quad d_2 = \frac{\partial}{\partial \sigma} x_p = z$$

$$C = \frac{\sigma^2}{n} \cdot \begin{bmatrix} 1.1128 & 0.2287 \\ 0.2287 & 0.8046 \end{bmatrix} \quad \text{volgens Hosking (1986), en}$$

$$\text{var}(\hat{x}_p) = \sum_{i=1}^2 \sum_{j=1}^2 d_i d_j c_{ij} = \frac{\sigma^2}{n} (1.1128 + 0.4574 z + 0.8046 z^2)$$

### 6.1.4 ML-schatting bij $GEV(\theta = 0)$

Nu geldt

$$C = \frac{6\sigma^2}{n\pi^2} \begin{bmatrix} \frac{\pi^2}{6} + (1-\gamma)^2 & 1-\gamma \\ 1-\gamma & 1 \end{bmatrix} = \frac{\sigma^2}{n} \begin{bmatrix} 1.1087 & 0.2570 \\ 0.2570 & 0.6079 \end{bmatrix}$$

derhalve

$$\text{var}(\hat{x}_p) \approx \frac{\sigma^2}{n} (1.1087 + 0.5140 z + 0.6079 z^2)$$

### 6.2 POT data

Een kwantiel ofwel T-jaar-herhalingswaarde ( $x_p$ ) kan in het geval van peaks over threshold reeksen, beschreven met behulp van de GP verdeling, geschat worden met behulp van de inverse verdelingsfunctie:

$$x_p = \begin{cases} \nu + \sigma \ln(\lambda T) & ; \theta=0 \end{cases} \quad (6.3a)$$

$$x_p = \begin{cases} \nu + \sigma [1 - (\lambda T)^{-\theta}] / \theta; & \theta \neq 0 \end{cases} \quad (6.3b)$$

, waarin  $\lambda$  het verwacht aantal pieken per jaar is.

Voor de variantie van de geschatte kwantielen geldt bij benadering (Van Montfort en Witter, 1986):

$$\text{var}(\hat{x}_p) = \begin{cases} \frac{\sigma^2}{\lambda M} [1 + \ln^2(\lambda T)] & ; \theta=0 \end{cases} \quad (6.4a)$$

$$\text{var}(\hat{x}_p) = \begin{cases} f_1 \text{var}(\hat{\lambda}) + f_2 \text{var}(\hat{\sigma}) + f_3 \text{var}(\hat{\theta}) + f_4 \text{cov}(\hat{\sigma}, \hat{\theta}); & \theta \neq 0 \end{cases} \quad (6.4b)$$

waarin

$M$  : reekslengte in jaren (=  $n/\hat{\lambda}$ , met  $\lambda$  de Poisson parameter)

$f_1$ :  $(\sigma/\lambda)^2 (\lambda T)^{-2\theta}$

$f_2$ :  $\{1 - (\lambda T)^{-\theta}\}^2 / \theta^2$

$f_3$ :  $\frac{1}{2} \sigma^2 \ln^4(\lambda T)$

$f_4$ :  $\frac{1}{2} \sigma \theta \ln^4(\lambda T) - \frac{1}{2} \sigma \ln^3(\lambda T)$ .

In Vergelijking 6.4b geldt  $\text{var}(\hat{\lambda}) = \lambda/M$ ; de termen  $\text{var}(\hat{\sigma})$ ,  $\text{var}(\hat{\theta})$  en

$\text{cov}(\hat{\sigma}, \hat{\theta})$  volgen uit de tweede afgeleiden van de log-likelihood

(Hoofdstuk 5.4). Deze formules volgen uit een Taylor-ontwikkeling van Vergelijking 6.3 en toepassen van de benaderingsformules in Kendall en Stuart (1977) voor de variantie van functies van kansvariabelen. Merk op dat indien wordt uitgegaan van een vast aantal waarnemingen (Methode 2) in Vergelijking 6.4b geldt  $f_1 \text{var}(\hat{\lambda}) = 0$ .

## 7 CONCLUSIES EN AANBEVELINGEN

De GEV en in enigszins mindere mate de GP verdeling zijn goed gedocumenteerde kandidaatverdelingen voor respectievelijk maxima en groottes van overschrijdingen van een drempelwaarde. Bij de GP verdeling lijken bij lage drempelwaarden wel enige problemen te verwachten met betrekking tot de veronderstelde Poisson verdeling van het jaarlijks aantal overschrijdingen. Lage drempelwaarden kunnen daar ook leiden tot een slechte aanpassing in de rechterstaart van de verdeling. Links censureren lijkt dan een logische oplossing (Van Montfort en Witter, 1986).

De PWM en ML schattingsmethoden voor de GEV en GP verdeling zijn adequaat, en goed gedocumenteerd; bij de PWM schattingsmethode zijn nog wel ontwikkelingen te verwachten met betrekking tot optimale schatters voor de kansgewogen momenten  $M_{p,r,s}$  (Vergelijking 5.2). Toetsen op al dan niet nul zijn van de vormparameter  $\theta$  van de GEV of GP verdeling, is op eenvoudige wijze mogelijk op basis van zowel de ML als de PWM schatter van  $\theta$ . Kwantielschattingen en hun standaard afwijking zijn ook te geven. Nog onvoldoende inzicht bestaat in de grootte van die standaard afwijking voor enige relevante hydrologische variabelen. Ook bestaat nog weinig inzicht in optimale methoden van combineren van meetreeksen voor het schatten van de T-jaar-herhalingswaarden bij grote herhalingsstijd T.

De conclusie lijkt gerechtvaardigd dat de methodiek wel voldoende ontwikkeld is; nu is naar analogie van het Flood Studies Report (NERC, 1975) behoefte aan richtlijnen voor frequentie analyse en aan resultaten van frequentie analyse van hydrologische variabelen.

Vaak blijft de relatie tussen veiligheid en het statistische begrip herhalingstijd onderbelicht. De bovengenoemde richtlijnen zullen dan ook, al dan niet onderbouwd door inzicht in veiligheidsaspecten, informatie moeten geven over te kiezen herhalingstijden bij specifieke ontwerp vragen.

#### DANKBETUIGING

Dank is verschuldigd aan het KNMI voor het verlenen van toestemming voor reproductie van de figuren 2, 3, 4 en 5.

#### LITERATUUR

- BUISHAND, T.A., 1977. Stochastic modelling of daily rainfall sequences. Mededelingen Landbouwhogeschool 77-3. Veenman & Zonen, Wageningen.
- BUISHAND, T.A., 1983. Uitzonderlijk hoge neerslaghoeveelheden en de theorie van de extreme waarden. Cultuurtechnisch Tijdschrift, 23: 9-20.
- BUISHAND, T.A., 1985. Het sneeuwdek in Nederland. KNMI, De Bilt.
- BUISHAND, T.A., 1986. Extreme-value analysis of climatological data. Third International Conference on Statistical Climatology, 23-27 June 1986. Vienna, 145-158.
- BUISHAND, T.A., 1987. De invloed van regionale verschillen in het voorkomen van zware neerslaggebeurtenissen op de berekende gemiddelde overstortingsfrequenties. H<sub>2</sub>O, 20: 486-489.
- BUISHAND, T.A., 1988. Statistics of extremes in climatology. Aangeboden aan Statistica Neerlandica.
- BUISHAND, T.A., en C.A. VELDS, 1980. Neerslag en verdamping. KNMI, De Bilt.
- DAVID, H.A., 1981. Order statistics. John Wiley, New York.
- DGW, 1986. Onderzoek basispeilen. Tussentijds verslag, juni 1986. Rijkswaterstaat, Dienst Getijdenwateren, Den Haag.
- DWW, 1984. Een hernieuwd onderzoek naar de maatgevende afvoer van de Rijn te Lobith. Dienst Weg- en Waterbouwkunde van Rijkswaterstaat, Den Haag.
- GOMES, M.I., 1981. An i-dimensional limiting distribution function of largest values and its relevance to the statistical theory of

- extremes. C. Tallie et al. (eds.), *Statistical distributions in scientific work*, Vol. 6, 389-410. D. Reidel, Dordrecht.
- GREENWOOD, J.A., J.M. LANDWEHR, N.C. MATALAS, and J.R. WALLIS, 1979. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15: 1049-1054.
- HAAN, C.T., 1977. *Statistical Methods in Hydrology*. The Iowa State University Press/Ames.
- HOSKING, J.R.M., 1985. A correction for the bias of maximum-likelihood estimators of Gumbel parameters. *Journal of Hydrology*, 78: 393-396.
- HOSKING, J.R.M., 1986. The theory of probability weighted moments. IBM T.J. Watson Research Center, New York.
- HOSKING, J.R.M., J.R. WALLIS, and E.F. WOOD, 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27: 251-261.
- HOSKING, J.R.M. and J.R. Wallis, 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29: 339-349.
- JENKINSON, A.F., 1977. The analysis of meteorological and other geophysical extremes. Meteorological Office Met.0. 13 Branch memorandum No. 58, Bracknell.
- KENDALL, M.G. and A. STUART, 1977. *The advanced theory of statistics*, vol. 1. Griffin & Co., London.
- KNMI, 1984. *Klimatologische gegevens van Nederlandse stations*. No. 11. Frequentietabellen en reeksen van dagen volgens bepaalde criteria. KNMI, De Bilt.
- KOOT, A.C.J., 1977. *Inzameling en transport van rioolwater*. Waltman, Delft.
- LEADBETTER, M.R., G. LINDGREN, and H. ROOTZEN, 1983. *Extremes and related properties of random sequences and series*. Springer Verlag, New York.
- LEVERT, C., 1959. Een nomogram voor betrouwbaarheids grenzen van kwantilen en overschrijdingskansen bij Normaal verdeelde grootheden. *Statistica Neerlandica*, 13: 3-14.
- NERC, 1975. *The Flood Studies Report*. Natural Environments Research Council, London.
- PICKANDS, J., 1975. *Statistical inference using extreme order sta-*

- tistics. *Annals of Statistics*, 3: 119-131.
- SINCLAIR, C.D. and M.I. AHMAD, 1988. Location invariant plotting positions for PWM estimation of the parameters of the GEV distribution. *Journal of Hydrology*, 99: 271-279.
- SMITH, R.L., 1986. Extreme value theory based on the  $r$ -largest annual events. *Journal of Hydrology*, 86: 27-43.
- VAN MONTFORT, M.A.J., 1969. Inleiding over de verdeling van extremen. *Statistica Neerlandica*, 23: 97-111.
- VAN MONTFORT, M.A.J., 1982. Modellen voor maxima en minima, schattingen en betrouwbaarheidsintervallen, keuze tussen modellen. Vakgroep Wiskunde Nota 82-2, LUW, Wageningen.
- VAN MONTFORT, M.A.J., and J.V. WITTER, 1985. Testing exponentiality against generalized Pareto distribution. *Journal of Hydrology*, 78: 305-315.
- VAN MONTFORT, M.A.J., and J.V. WITTER, 1986. The generalized Pareto distribution applied to rainfall depths. *Hydrological Sciences Journal*, 31: 151-162.
- WIERINGA, J. en J.P. RIJKOORT, 1983. Windklimaat van Nederland. KNMI, De Bilt.
- WITTER, J.V., 1984. Heterogeneity of Dutch rainfall. Proefschrift Landbouwhogeschool, Wageningen.





## Aantekeningen

## Aantekeningen

## Aantekeningen



RAPPORTEN EN NOTA'S  
COMMISSIE VOOR HYDROLOGISCH ONDERZOEK TNO

- No. 1. Tweede rapport en aanbevelingen van de Contactgroep  
Archivering en Automatische Verwerking van hydrologische  
gegevens TNO.  
Januari 1977
- No. 2. Verslag en aanbevelingen van de ad hoc Groep  
Grondwatermodellen en Computerprogrammatuur TNO.  
Juli 1978
- No. 3. De droogte in 1976.  
Een samenvatting en overzicht van de over de droogte van 1976  
verschenen literatuur - (P.K.M. v.d. Heijde).  
Augustus 1978
- No. 4. Nederlandse activiteiten in internationaal hydrologisch  
verband.  
Lezingenserie, gehouden op 25 april 1978 te Delft, aangevuld  
met (schematische) overzichten van internationale organisaties  
en een overzicht van hun vertegenwoordigers in Nederland.  
Augustus 1978
- No. 5. Waterkwaliteit in grondwaterstromingsstelsels.  
Een verslag van de workshop op 1 en 2 april 1980 te Wageningen  
- (J.C. Hooghart), aangevuld met discussiebijdragen en een  
inventarisatie van onderzoek in Nederland.  
Augustus 1980
- No. 6. Derde rapport en aanbevelingen van de Contactgroep Archivering  
en Automatische verwerking van hydrologische gegevens TNO.  
Februari 1981
- No. 7. Overzicht van de wensen van hydrologen en waterbeheerders ten  
aanzien van het operationele regenwaarnemingsnet van het KNMI  
- J.C. Hooghart.  
Oktober 1981

- No. 8. \*) Verklarende Hydrologische Woordenlijst van de Gespreksgroep  
Hydrologische Terminologie.  
8a. I Water in de onverzadigde zone  
    II Water in de verzadigde zone  
        Januari 1982  
8b. III Atmosferisch water  
        Juni 1983  
8c. IV Oppervlaktewater  
        Maart 1985
- No. 9. Waterkwaliteit en waterkwantiteit in het IJsselmeergebied.  
Een verslag van de 2e CHO-studiebijeenkomst op 2 en 3 november  
1981, De Eemhof, Zuidelijk Flevoland - (J.C. Hooghart),  
aangevuld met discussiebijdragen.  
Februari 1982.
- No. 10. Rapport en aanbevelingen van de Contactgroep  
Grondwatermodellen, CHO-TNO  
April 1982.
- No. 11. Inventarisatie Grondwaterkwaliteitsmodellen.  
L.J.M. Boumans  
Oktober 1982
- No. 12. Grondwaterkwaliteit in relatie met onderzoek en beleid.  
Een verslag van de 3e CHO-studiebijeenkomst op 15 maart 1983  
te Wageningen - (J.C. Hooghart), aangevuld met  
discussiebijdragen.  
Juni 1983
- No. 13. Vergelijking van modellen voor het onverzadigd  
grondwatersysteem en de verdamping.  
Een verslag van de 4e CHO-studiebijeenkomst op 24 oktober 1984  
in samenwerking met de Studiegroep Hupselse Beek - (J.C.  
Hooghart).  
Maart/april 1985

\*) Verouderd, zie de nieuwe uitgave no. 16.

- No. 14. Meten, meetnetten en optimale meetnetontwerpen ten dienste van het waterbeheer.  
Een verslag van:  
- Voorjaarsbijeenkomst van de KIVI Sectie Waterbeheer:  
"Meten voor waterbeheer", mei 1984.  
- Colloquium van de Studiegroep Statistiek in de hydrologie  
CHO-TNO: "Meetnetontwerp en optimalisatie", november 1984  
Februari 1986
- No. 15. Het hydrologisch systeem in het grensgebied Luik-Maasbracht.  
Le système hydrologique dans la région frontalière  
Liège-Maasbracht.  
Een verslag van de 5e CHO-studiebijeenkomst op 13 december  
1985 in samenwerking met de Nationale IHP-comité's van België  
en Nederland en de Contactgroep Hydrologie van het Nationaal  
Fonds voor Wetenschappelijk Onderzoek uit België - (J.C.  
Hooghart).  
April 1986
- No. 16. Verklarende Hydrologische Woordenlijst van de Gespreksgroep  
Hydrologische Terminologie, waarin opgenomen de hoofdstukken:  
I Algemene termen  
II Atmosferisch Water  
III Water in de onverzadigde zone  
IV Water in de verzadigde zone  
V Oppervlaktewater  
Oktober 1986, hernieuwde uitgave
- No. 17. Duurzaamheid rioolleidingen; een literatuurstudie naar  
aantastingsmechanismen.  
R.B. Polder  
Februari 1987
- No. 18. Ruimtelijke variabiliteit van bodem en water.  
Een verslag van de 6e CHO-studiebijeenkomst op 22 oktober 1986  
- (J.C. Hooghart)  
Februari 1987



- No. 19. Van Penman naar Makkink; een nieuwe berekeningswijze voor de klimatologische verdampingsgetallen.  
Eindrapport van de project- en begeleidingsgroep  
Verdampingsberekeningen - (J.C. Hooghart en W.N. Lablans)  
(in bewerking)
- No. 20. Tijdreeksen in bodem en water.  
Bijdragen aan de lezingendag op 26 januari 1989 van de NRLO werkgroep "Ruimtelijke variabiliteit in Bodem en Water" en de Studiegroep Statistiek in de Hydrologie van de CHO-TNO.  
December 1988

Voor bestellingen en informatie:

CHO-TNO

Postbus 297

2501 BD 's-Gravenhage

