GETIJ ANALYSE EN VOORSPELLINGEN M.B.V. ARIMA-MODELLEN

- tekst -

S.A.P.J. Hendriksen

september 1981

Vakgroep: Vloestofmechanica
Docent: Prof.dr.ir.J.P.Th. Kalkwijk
Publikatie no: R/1981/161D

GETIJ ANALYSE M.B.V. ARIMA-MODELLEN

S.A.P.J. HENDRIKSEN

SEPTEMBER 1981

INHOUDS OPGAVE.

FIGUREN.

0 INLEIDING.

Doel van dit onderzoek "GETIJ ANALYSE M.B.V. ARIMA-MODELLEN(1)" is, het voorspellen van stormvloed hoogten m.b.v. ARIMA-modellen.
Als onderzoek gebied is het getij bij Vlissingen in 1976 gekozen.
De basis gedachte achter ARIMA-modellen is, dat het mogelijk moet zijn om een gegeven serie waarnemingen te modelleren m.b.v. de statistische eigenschappen van die serie. Daarbij spelen vooral de auto-correlatie, de partiele auto-correlatie en de verdelingsdichtheid functie een belangrijke rol. In hoofdstuk I is de theorie van ARIMA-modellen kort samen gevat.
In de na volgende hoofdstukken wordt het onderzoek als volgt uitgesplitst:

* algemene tijdreeks analyse
* ARIMA-model complete signaal, plus voorspelling hiermee
* afpellen astronomisch deel m.b.v. kleinste kwadraten methode
* analyse rest signaal
* ARIMA-model rest signaal voor rustige en storm perioden
* simulatie voorspelling gedurende een storm periode

---

(1) Auto Regressive Integrated Moving Average

HOFFDSTUK I INTRODUCTIE ARIMA MODELLEN.

1.1 INTRODUCTIE.

ARIMA modellen worden toegepast op discrete,stationaire waarnemings reeksen. In feite is dit niet geheel juist en zorgt de 'I' (integrated) van ARIMA modellen dat aan deze stationairiteits eis kan worden voldaan.
Allereerst introduceren we hier enkele eenvoudige operatoren:

* terugschuif operator (backshift operator) B:

$$Bz(t)=z(t-1) \quad (B^m)z(t)=z(t-m)$$

* vooruitschuif operator (forwardshift operator) F:

$$Fz(t)=z(t+1) \quad (F^m)z(t)=z(t+m)$$

* terugwaartse differentie operator (backward difference operator) ∇:

$$\nabla z(t)=z(t)-z(t-1) \quad \nabla z(t)=(1-B)z(t)$$

Het basis idee gaat uit van een lineair filter model. Hiermee wordt een serie waarnemingen beschreven, waarvan opeenvolgende waarden de output <z(t)> vormen van een lineair tijdsonafhankelijk systeem waarvan de input <a(t)> gegeven is als een serie onafhankelijke, normaal verdeelde random 'pulsen', met gemiddelde nul en variantie σ_a^2.
Deze z(t)'s kunnen nu worden uitgedrukt in de a(t)'s met behulp van het lineair filter:

$$z(t)=Y_0a(t)+Y_1a(t-1)+Y_2a(t-2)+ \ldots$$

$$z(t)=Y(B)a(t)$$

$\Upsilon(B)$ heet het lineair filter of de overdrachts functie.
Bij alle hierna te beschouwen systemen (modellen cq filters), gaan we er vanuit, dat het gemiddelde (u) van het uitgangssignaal pas na het filter bij dit uitgangssignaal wordt opgeteld (zie figuur 0).

MA of Moving Average modellen: nu is z(t) lineair afhankelijk van een eindig aantal voorafgaande 'pulsen' a(t):

$$z(t)=a(t)-\theta_1 a(t-1)-\theta_2 a(t-2)- \ldots -\theta_q a(t-q)$$

$$z(t)=\theta(B)a(t)$$

$$\theta(B)=\theta_0-\theta_1 B-\theta_2 B - \ldots -\theta_q B^q$$

$\theta(B)$ is de Moving Average operator van de orde q.
Het MA(q) model heeft q+1 onbekende parameters:

$$\theta_1,\theta_2, \ldots ,\theta_q,\sigma^2_a$$

Het is mogelijk uit bovenstaande vergelijking voor z(t) de a(t)'s te elimineren, dan ontstaat de volgende uitdrukking:

$$z(t)=\pi_1 z(t-1)+\pi_2 z(t-2)+ \ldots+a(t)$$

AR of Auto Regressive modellen: voor een puur AR model geldt dat de nieuwe waarde van een proces uitgedrukt kan worden als een eindige lineaire combinatie van voorgaande waarden plus een puls a(t):

$$z(t)=\phi_1 z(t-1)+\phi_2 z(t-2)+ \ldots +\phi_p z(t-p)+a(t)$$

$$\phi(B)z(t)=a(t)$$

$$\phi(B)=1-\phi_1 B-\phi_2 B^2- \ldots -\phi_p B^p.$$

$\phi(B)$ is de Auto Regressive operator van de orde p.
Het AR(p) model heeft p+1 onbekende parameters:

$$\phi_1,\phi_2, \ldots ,\phi_p,\sigma^2_a.$$

Inplaats van $\phi(B)z(t)=a(t)$ kunnen we ook $z(t)=\Psi(B)a(t)$ schrijven met $\Psi(B)=\phi^{-1}(B)$. Voor stationaire processen moeten

$$\phi_1,\phi_2, \dots ,\phi_p$$

zodanig worden gekozen, dat de gewichten $\Psi_1,\Psi_2,\dots$ een convergente serie vormen.

Mixed Auto Regressive Moving Average (ARMA) model:

$$\phi(B)z(t)=\theta(B)a(t)$$

Dit model heeft p+q+1 onbekende parameters

Het ARIMA(p,d,q) model tenslotte ziet er als volgt uit:

$$\phi(B)\nabla^d z(t)=\theta(B)a(t)$$

1.2 ENIGE DEFINITIES UIT DE WAARSCHIJNLIJKHEIDS REKENING.

Definitie stochastisch proces.

Een waarschijnlijkheidsexperiment wordt beschreven door zijn uitkomsten, die tezamen de ruimte S vormen, als mede een klasse van deelverzamelingen van S waarvan de elementen gebeurtenissen worden genoemd en waaraan waarschijnlijkheden zijn toegekend.
Door aan elke uitkomst k volgens een zekere regel een tijdsfunctie $x(t;k)$ toe te kennen wordt een stelsel van functies, een voor elke k, gevormd. Een dergelijk stelsel heet stochastisch proces, en is in feite een functie van twee variabelen t en k. Het domein van k is de ruimte S, het domein van t is de verzameling van reele getallen, die tezamen de gehele tijdas beslaan. De grootheid $x(t;k)$ kan derhalve voorstellen ;

a) voor een specifieke uitkomst k_i zal $x(t;k_i)$ alleen van t afhankelijk zijn; $x(t;k_i)$ is dus een tijdsfunctie. Deze tijdsfunctie wordt een realisering van het

stochastische proces genoemd.

b) voor een zeker tijdstip t_i zal de grootheid $x(t_i;k)$ alleen afhankelijk zijn van de uitkomst k: $x(t_i;k)$ zal een random variabele zijn.

Voorbeeld: stel men is geinteresseerd in de beschrijving van de golfhoogte als functie van de tijd t en de plaats k, welke kan worden voorgesteld als een stochastisch proces $x(t;k)$:

zie figuur 1

* $x(t;k_1)$ en $x(t;k_2)$ etc. zijn voorbeelden van a) (beschrijft in een punt het verloop van de hoogte)

* $x(t_i;k)$ een voorbeeld van b), beschrijft het hele oppervlak op $t=t_i$.

Mathematische verwachting of het gemiddelde van $x(t;k)$:

$$u_x(t)=E\{x(t;k)\}$$

variantie van $x(t;k)$:

$$\delta_x^2(t)=E\{(x(t;k)-u_x(t))^2\}$$

auto covariantie van $x(t;k)$:

$$y(t_1,t_2)=E\{(x(t_1;k)-u_x(t_1))(x(t_2;k)-u_x(t_2))\}$$

Voor stationaire processen geldt:

$$\left.\begin{array}{l} u_x(t)=u_x \\ \delta_x^2(t)=\delta_x^2 \end{array}\right| \quad \text{onafhankelijk van } t$$

$y(t_1,t_2)=y(\tau)$ alleen afhankelijk van $\tau=t_2-t_1$

De autocorrelatie functie geeft aan in hoeverre er een

samenhang bestaat tussen de waarde $x(t)$ en de waarde van x k tijdstappen eerder: $x(t-k)$. Een correlatie coefficient, <r>, 1.0 duidt op volledige gecorreleerdheid, terwijl een coefficient 0.0 op ongecorreleerdheid duidt:

$$r(k)=y(k)/y(0)$$

De autocorrelatie functie heeft de volgende eigenschappen:

- $r(0)=1$
- $r(k)=r(-k)$ acf is een even functie
- $|r(k)|\leq 1.0$

De partiele autocorrelatie functie (pacf) is een maat voor hoeveel correlatie er nog bestaat na een tijdsinterval k (zie ook $1.4.2).

De pacf is als volgt gedefinieerd:
Stel de autocorrelatie coeffiecienten $r(k)$ $(k=1,2,3,\ldots)$ zijn bekend, dan is het volgende stelsel vergelijkingen op te stellen:

$$\begin{vmatrix} 1 & r(1) & \cdots & r(k-1) \\ r(1) & 1 & \cdots & r(k-2) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ r(k-1) & r(k-2) & \cdots & 1 \end{vmatrix} \begin{vmatrix} \phi(k1) \\ \phi(k2) \\ \cdot \\ \cdot \\ \cdot \\ \phi(kk) \end{vmatrix} = \begin{vmatrix} r(1) \\ r(2) \\ \cdot \\ \cdot \\ \cdot \\ r(k) \end{vmatrix}$$

De rij $\phi(11),\phi(22),\phi(33),\ldots$ noemt men de partiele autocorrelatie coefficienten. Deze zijn met matrix rekening eenvoudig uit deze vergelijkingen op te lossen.

N.B. Als we te maken hebben met een AR(p) proces, en $k=p$, dan zijn $\phi(p1),\phi(p2),\ldots,\phi(pp)$ juist de model parameters $\phi_1,\phi_2,\ldots,\phi_p$.

1.3.1. MOVING AVERAGE PROCES.

Stel $\{a(t)\}$ is een puur random proces met gemiddelde nul en variantie σ_a^2 . $\{x(t)\}$ is nu een MA(q) proces als geldt:

$$x(t)=\theta_0 a(t)-\theta_1 a(t-1)-\theta_2 a(t-2)-\ldots-\theta_q a(t-q)$$

Hierin zijn de $\{\theta_i\}$ constant, en is $\theta_0=1$ (meestal).
Voor x(t) zijn de volgende eigenschappen af te leiden:

$$u_x=E\{x(t)\}=0$$

$$var\{x(t)\}=\sigma_a^2 \sum \theta^2$$

$$Cov(x(t),x(t+k))=y(k)= \begin{cases} 0 & k>p \\ \sigma_a^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k} & k=0,1,2,\ldots q \\ y(-k) & k<0 \end{cases}$$

$$\text{autocorrelatie functie } r(k)= \begin{cases} 0 & k>q \\ \sum_{i=0}^{q-k} \theta_i \theta_{i+k} \Big/ \sum_{0} \theta_i^2 & k=0,1,2,\ldots,q \\ r(-k) & k<0 \end{cases}$$

Voor een MA(1) proces geeft dit dus voor de acf:

MA(1): $x(t)=a(t)-\theta a(t-1)$

$$r(k)= \begin{cases} 1 & k=0 \\ \theta/(1+\theta^2) & k=1 \\ 0 & |k|>1 \end{cases}$$

Aan de $\{\theta_i\}$ van het MA proces moeten eisen gesteld worden om de invertibiliteit van het model te verzekeren. Deze eis komt er op neer dat de z(t)'s in de a(t)'s uit te drukken zijn, en dat ook het omgekeerde mogelijk is. Deze eis komt

overeen met het begrip 'inverse' uit de functie theorie, waarbij ge-eist wordt dat iedere x een y hoort en omgekeerd bij iedere y een x.

zie figuur 2

------ : y=g(x) en x=f(y) invertibel
...... : y=h(x), maar x is niet in y uit te drukken, niet invertibel.

Deze invertibiliteits eis is ook uit te leggen aan de hand van weer twee MA(1) processen.
Stel we hebben de processen a) $x(t)=a(t)-\theta a(t-1)$ en b) $x(t)=a(t)-(1/\theta)a(t-1)$. De beide processen hebben dezelfde auto correlatie functie:

$$r(k)=\theta/(1+\theta^2) =(1/\theta)/(1+1/\theta^2)$$

Drukken we nu a(t) uit in termen van x(t),x(t-1),... dan vinden we:

a) $a(t)=x(t)- \theta x(t-1)+ \theta^2 x(t-2)- \ldots$

b) $a(t)=x(t)-(1/\theta)x(t-1)+(1/\theta)^2 x(t-2)- \ldots$

Het zal duidelijk zijn dat voor $|\theta|<1$ serie a) convergeert. Dus een schattings procedure die werkt met het schatten van de rest serie a(t) zal automatisch naar model a) leiden. De invertibiliteits conditie geeft de garantie dat er een uniek MA proces is bij een gegeven acf.

Een MA proces ziet er in het algemeen aldus uit:

$$x(t)=(\theta_0+\theta_1 B+\theta_2 B^2+ \ldots +\theta_q B^q)a(t)=\theta(B)a(t)$$

Dit proces is invertibel als de wortels van de vergelijking

$$\theta(B)=\theta_0+\theta_1 B+\theta_2 B^2+ \ldots +\theta_q B =0$$

allen buiten de eenheids cirkel liggen.

overeen met het begrip 'inverse' uit de functietheorie, waarbij geëist wordt dat ieder x een y voortbrengt en omgekeerd bij iedere y een x.

zie figuur 2

------ : y=f(x) en x=f'(y) invertibel

...... : y=h(x), maar x is niet in y uit te drukken, niet invertibel

Deze invertibiliteitseis is ook uit te leggen aan de hand van het twee MA(1) processen.
Stel we hebben de processen a) $x(t)=e(t)-\theta e(t-1)$ en b) $x(t)=e(t)-(1/\theta)e(t-1)$. Deze beide processen hebben dezelfde autocorrelatiefunctie:

$$r(1)=\theta/(1+\theta^2)=(1/\theta)/(1+1/\theta^2)$$

Drukken we nu de $e(t)$ uit in termen van $x(t),x(t-1),\ldots$ dan vinden we:

$$a)\quad e(t)=x(t)+\theta x(t-1)+\theta^2x(t-2)+\ldots$$

$$b)\quad e(t)=x(t)+(1/\theta)x(t-1)+(1/\theta)^2x(t-2)+\ldots$$

Het zal duidelijk zijn dat voor $|\theta|<1$ serie a) convergeert. Bij een schattingsprocedure die werkt met het schatten van de rest serie $e(t)$ zal automatisch naar model a) leiden. De invertibiliteitsconditie wordt dus gegarandeerd dat er een uniek MA proces bij een gegeven acf.

Een MA-proces ziet er in het algemeen als volgt uit:

$$x(t)=(\theta_0+\theta_1B+\theta_2B^2+\ldots+\theta_qB^q)e(t)=\theta(B)e(t)$$

Dit proces is invertibel als de wortels van de vergelijking

$$\theta(B)=\theta_0+\theta_1B+\theta_2B^2+\ldots+\theta_qB^q=0$$

allen buiten de eenheidscirkel liggen.

1.3.2. AUTO REGRESSIVE PROCES.

Stel $a(t)$ is een puur random proces met gemiddelde nul en variantie σ_a^2. $\{x(t)\}$ is nu een AR(p) process als geldt:

$$x(t)=\phi_1 x(t-1)+\phi_2 x(t-2)+ \ldots +\phi_p x(t-p)+a(t)$$

Dit komt sterk overeen met een 'multiple regression model', maar $x(t)$ is nier ge-'regressed' naar een onafhankelijk variabele, maar naar waarden uit het verleden van $x(t)$, vandaar ook de naam 'auto' regressive.
We beginnen met het beschouwen van een AR(1) proces, ook wel een MARKOV proces genoemd:

$$x(t)=\phi x(t-1)+a(t) \quad AR(1)$$

Dit is als volgt om te zetten in een MA(..) proces:

$$x(t)=\phi(\phi x(t-2)+a(t-1))+a(t)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$=a(t)+\phi a(t-1)+\phi^2 a(t-2)+ \ldots$$

Uit $x(t)=\phi x(t-1)+a(t)$ volgt ook $(1-\phi B)x(t)=a(t)$ dus

$$x(t)=a(t)/(1-\phi B)=(1+\phi B+\phi^2 B^2+ \ldots)a(t)$$

$$=a(t)+\phi a(t-1)+\phi^2 a(t-2)+\ldots$$

Verder is voor $x(t)$ af te leiden:

$$Var\{x(t)\}=\sigma_a^2(1+\phi^2+\phi^4+ \ldots)$$

De variantie is eindig, aangenomen dat $|\phi|<1$, en hieruit volgt dan weer dat:

$$\sigma_x^2=\sigma_a^2/(1-\phi^2)$$

De auto covariantie functie:

$$y(k)=E\{x(t).x(t-k)\}=\sigma_a^2 \sum^{k} \phi^{i} . \phi^{i+k} \qquad k>0$$

gaat voor $|\phi|<1$ over in:

$$\gamma(k)=\phi^{k}\sigma^2_a/(1-\phi^2)=\phi^{k}\sigma^2_x$$

De autocorrelatie functie wordt dan:

$$r(k)=\phi \quad (k=0,\pm1,\pm2, \ldots)$$

Hier onder volgen drie figuren van de autocorrelatie en de partiele autocorrelatie functie van twee AR(1) processen en een AR(2) proces. Voor ϕ is respectievelijk 0.8, -0.8 en (0.8 -0.4):

zie figuur 3

Een AR proces ziet er in het algemeen aldus uit:

$$(1-\phi_1 B-\phi_2 B^2- \ldots -\phi_p B^p)x(t)=a(t)$$

$$\text{of } x(t)=a(t)/(1-\phi_1 B- \ldots -\phi_p B^p)=f(B)a(t)$$

$$\text{met } f(B)=(1+s_1 B+s_2 B^2+ \ldots)$$

De variantie van een AR proces is eindig, aangenomen dat $\sum(s^2_i)$ convergeert.
Het is gebleken dat als $\sum |s_i|$ convergeert dit al een voldoende voorwaarde is voor stationairiteit.

Als wordt aangenomen dat het proces stationair is, dan kunnen de Yule-Walker vergelijkingen worden toegepast om de ϕ_i te bereken:

$$r(k)=\phi_1 r(k-1)+ \ldots +\phi_p r(k-p)$$

Zij vormen een verzameling differentie vergelijkingen met als algemene oplossing:

$$r(k)=A_1 \pi_1^{|k|} + \ldots + A_p \pi_p^{|k|}$$

Als stationairiteits eis geldt weer dat alle wortels van de vergelijking:

$$\phi(B)=1-\phi_1 B- \ldots-\phi_p B^p=0$$

buiten de eenheids cirkel moeten liggen.

Voorbeeld: AR(2) proces, π_1 en π_2 zijn de wortels van

$$y^2-\phi_1 y-\phi_2=0$$

hieruit volgt dan dat $|\pi_i|<1$ als:

$$|(\phi_1 \pm \sqrt{\phi_1^2+4\phi_2})/2|<1$$

Het stationairiteits gebied is de driehoek:

$$\phi_1+\phi_2<1$$

$$\phi_1-\phi_2>-1$$

$$\phi_2>1$$

De wortels zijn reeel als:

$$\phi_1^2+4\phi_2>0$$

de acf neemt dan exponentieel met k af. En de wortels zijn complex als:

$$\phi_1^2+4\phi_2<0$$

de acf is dan een gedempte sinus golf.

1.4.1. SCHATTEN VAN DE PARAMETERS VAN EEN AR PROCES.

Het schatten van de parameters van een AR(p) proces:

$$x(t)=\phi_1 x(t-1)+ \ldots +\phi_p x(t-p)+a(t)$$

met N waarnemingen x(1),x(2),...,x(N) gegeven, kan geschieden door de parameters ϕ_1, ... ,ϕ_p te bepalen m.b.v. de kleinste kwadraten methode. Door de volgende vergelijking te minimaliseren:

$$S=\sum \{x(t)-\phi_1 x(t-1)- \ldots -\phi_p x(t-p)\}^2$$

In het eerste orde geval vinden we:

$$\phi_1=\{\sum x(t).x(t+1)\}/\{\sum x(t)^2\}$$

Hieruit volgt dan

$$\phi_1 = y(1)/y(0) = r(1)$$

Het is interessant om nog op te merken dat deze schatter precies dezelfde is is, als de schatter die te voorschijn zou komen als we de auto regressive vergelijking:

$$x(t) = \phi_1 x(t-1) + a(t)$$

zouden behandelen als een gewone regressie met x(t) als onafhankelijk variabele.

Voor een tweede orde AR proces, kunnen soortgelijke benaderingen gedaan worden:

$$\phi_2 = r(1)(1-r(2))/(1-r(1)^2)$$

$$\phi_2 = (r(2)-r(1)^2)/(1-r(1)^2)$$

Een tweede methode voor het bepalen van de parameters, is het stelsel vergelijkingen oplossen dat ontstaat na substitutie van de geschatte autocorrelatie coefficient in de eerste p Yule-Walker vergelijkingen:

$$\text{algemeen: } r(k) = \phi_1 r(k-1) + \ldots + \phi_p r(k-p)$$

Uitschrijven voor r(1) t/m r(p) geeft p vergelijkingen met p onbekenden:

$$r(1) = \phi_1 \; 1 \; + \phi_2 r(2) + \ldots + \phi_p r(p-1)$$

$$r(2) = \phi_1 r(1) + \phi_2 \; 1 \; + \ldots + \phi_p r(p-2)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$r(p) = \phi_1 r(p-1) + \ldots\ldots\ldots + \phi_p \; 1 \; .$$

1.4.2. HET BEPALEN VAN DE ORDE VAN EEN AR PROCES.

Een belangrijk hulp middel bij het bepalen van de orde van een AR proces is de partiele autocorrelatie functie.
Stel, we 'fitten' een model van de orde m. De laatste coefficient ϕ_m noemen we hier π_m en deze is een maat voor 'de rest correlatie' na een tijds interval m. Een 'rest correlatie' waarmee een model van de orde (m-1) geen rekening zou houden. π_m wordt de m' partiele auto correlatie coeffiecient genoemd. Maken we nu een grafiek waarin we π_m uitzetten tegen m, dan hebben we het partieel autocorrellogram. Waarden die significant verschillen van nul geven de orde van het AR proces aan.

1.4.3. SCHATTEN VAN PARAMETERS VAN EEN MA PROCES.

Voor het schatten van de parameters van een MA proces zijn,in tegen stelling tot een AR proces, geen expliciete schatters te formuleren.
Inplaats daarvan moet een numerieke iteratie uitgevoerd worden (Newton-Rapson procedure). Laten we om te beginnen een MA(1) proces beschouwen:

$$x(t)=a(t)-\theta_1 a(t-1)$$

Kies nu een geschikte startwaarde voor θ_1, bijvoorbeeld zoals volgt uit de vergelijking:

$$\theta_1=r(1)*(1-\theta_1^2)$$

Nu kan de overeenkomstige som van de kwadraten van de 'rest' berekend worden:

$$a(t)=x(t)-\theta_1 a(t-1)$$

met a(0)=0 vinden we:

$$a(1)=x(1)$$

1.4.4. [illegible]

Een belangrijk hulpmiddel bij het bepalen van de orde van een AR proces is de partiële autocorrelatie functie.
Stel we 'fitten' een model van de orde m. De laatste coefficient noemen we hier π_m en deze is een maat voor 'de 'best correlatie' na een tijdsinterval m. Een 'best correlatie' waarmee een model van de orde (m-1) geen rekening houdt. π_m wordt de 'partiële autocorrelatie coefficient' genoemd. Maken we nu een grafiek waarin we π_m uitzetten tegen m, dan hebben we het partieel autocorrelogram. Waarden die significant verschillen van nul geven de orde van het AR proces aan.

1.4.5. SCHATTEN VAN PARAMETERS VAN EEN MA PROCES.

Voor het schatten van de parameters van een MA proces zijn, in tegenstelling tot een AR proces, geen expliciete schatters te formuleren.
In plaats daarvan moeten numerieke methoden (iteratie) worden (Newton-Raphson procedure). Laten we om te beginnen een MA(1) proces beschouwen:

$$x(t)=a(t)-\theta_1 a(t-1)$$

Kiezen we een geschikte startwaarde voor θ_1, dan kan de volgende vergelijking:

$$r_1=-\theta_1/(1+\theta_1^2)$$

En kan de over ... van de kwadraten van de 'rest' berekend worden:

$$a(t)=x(t)+\theta_1 a(t-1)$$

met a(0)=0 vinden we:

$$a(1)=x(1)$$

$$a(n)=x(n)-\theta_1 a(n-1)$$

en dus kan $\sum a(t)^2$ berekend worden.
Kies vervolgens een nieuwe waarde voor θ_1 en bereken opnieuw $\sum a(t)^2$. Deze sommen van kwadraten kunnen worden uitgezet tegen θ_1. Die waarde van θ_1 die de kleiste som van kwadraten levert, is de beste schatting.
Deze methode kan nog verbeterd worden door de waarde van a(0) te berekenen door terugwaartse voorspelling. Het is echter gebleken dat dit alleen nuttig is in het geval dat $\theta_1 = \pm 1$.

Ook voor MA processen van hogere orde is dit de aangewezen methode. Er moeten alleen meer start waarden gekozen worden. Nieuwe keuzen van start waarden moeten wel slim gedaan worden om in de richting van de optimale parameters te komen. Box en Jenkins(1970, Section 7.2) hebben hiervoor een procedure ontwikkeld.

1.4.4. HET BEPALEN VAN DE ORDE VAN EEN MA PROCES.

De orde van een MA proces is meestal vrij eenvoudig te bepalen met behulp van de geschatte autocorrelatie functie. Dit komt door de eenvoudige vorm van de theoretische autocorrelatie functie van een MA proces:

zie figuur 4

Als de θ_i van een MA proces bekend zijn, dan kunnen hiermee de theoretische autocorrelaties van het proces berekend worden. Zij het een zuiver MA proces was.
Uit de theoretische autocorellogrammen van figuur 4 blijkt

dat voor een MA(1) proces er een autocorrelatie coefficient ≠0 is, en dat voor een MA(2) proces er twee autocorrelaties ≠0 zijn. Dit gaat zo door voor hogere orde MA processen.

1.4.5. SCHATTEN VAN DE PARAMETERS VAN EEN ARMA PROCES.

De schattings problemen van de ARMA parameters zijn soort gelijk als de schattings problemen van een MA proces.
Ook hier zullen startwaarden gekozen moeten worden, en de som van de kwadraten van de rest berekend moeten worden.
Door deze procedure te herhalen moeten weer de optimale schatters gevonden worden: <marquart algoritm>.
Voor de keuze van het aantal AR en/of MA termen geven Box en Jenkins de in figuur 5 en 6 opgenomen informatie als leidraad.

1.4.6. SCHATTEN VAN DE PARAMETERS VAN EEN ARIMA PROCES.

In de praktijk zijn de meeste waarnemings series niet stationair en modellen die hievoor beschreven zijn, zijn dan niet direct toepasbaar. Dan kunnen we de waarnemings serie eerst een of meer malen differentieren om zo stationairiteit te bereiken.
Vervolgens kunnen dan de procedures toegepast worden die beschreven zijn voor AR, MA en/of ARMA processen.

1.4.7. SEIZOEN MODELLEN.

In de praktijk vertonen waarnemings series een zekere

waarnemingen herhaalt.

Beschouwen we nu bijv een waarnemings serie met maandelijkse gegevens (s=12), dan kan volgens BOX en JENKINS het volgende model opgesteld worden:

$$\phi_p(B)\Phi_P(B^{12})w(t) = \theta_q(B)\Theta_Q(B^{12})a(t)$$

met B^{12} : $B^{12}w(t)=w(t-12)$

Hiermee is weer een stationair model mee gedefinieerd, aangenomen dat de wortels van:

$$\phi_p(B)\Phi_P(B^{12})=0$$

buiten de eenheids cirkel liggen.

Mocht de waarnemings serie niet stationair zijn, dan kan de volgende manier van differentieren helpen bij het stationair maken van de waarnemings serie:

$$w(t)= \nabla^d \nabla^D x(t)$$

met $\nabla^{12} x(t)=x(t)-x(t-12)$

en $\nabla\nabla^{12} x(t)=\nabla^{12} x(t)-\nabla^{12} x(t-1)=x(t)-x(t-1)-x(t-12)+x(t-13)$

HOOFDSTUK II DE WAARNEMINGS SERIE

2.1. ALGEMEEN.

In dit hoofdstuk geven we aan hoe we de getij gegevens van Vlissingen geanalyseerd hebben.
Van het getij zijn uur waarenemingen gegeven, te beginnen op 1 januari 1976 om 1.00u a.m. en eidigend op 31 december 1976 om 24.00u p.m., in totaal dus 8784 uur waarnemingen.
Allereerst hebben we de waarnemingen grafisch weergegeven (figuur 7) Vervolgens zijn maximum,minimum en gemiddelde bepaald, daarna trachten we het getij te analyseren met behulp van het autocorrelogram. We onderzoeken het signaal op trend en proberen rustige perioden en storm perioden aan te geven.
Tot slot analyseren we het signaal uitgaande van een aantal bekende astronomische componenten, en onderzoeken of de eigenschappen van de restserie, die ontstaat door het astronomisch getij van het meet signaal af te trekken.

2.2.1. EERSTE VERKENNING.

Zoals al gezegd hebben we eerst het maximum, minimum en het gemiddelde berekend:

het maximum is	389.0cm	
		397.47 cm
het gemiddelde is	-8.52cm	
het minimum is	-266.0cm	255.53 cm

Uit de grafiek van het signaal blijkt dat er vaker (meer) verhoging optreedt dan verlaging door invloeden buiten de

astronomische invloeden. Bovendien geven die invloeden een grotere verhoging dan verlaging van de waterstand, zoals ook blijkt uit het maximum en het minimum.
Dit betekent dat het gemiddelde dat we hier gevonden hebben infeite hoger ligt dan de evenwichts stand waarom het astronomisch getij zich beweegt.
Daarom is uitgegaan van een analyse van het getij bij Vlissingen door Rijks Waterstaat (Hatyan 1974), waar als middenstand aangegeven wordt -8.86 cm (t.o.v. NAP).
Hierna wordt verder steeds gerekend met het gegeven signaal gecorrigeerd voor de middenstand:

$$y(t)=y(t)+8.86$$

2.2.2. ONDERZOEK M.B.V. HET AUTOCORRELLOGRAM.

Het autocorrellogram of de autocorrelatie functie is een grafische weergave van de autocorrelatie coefficienten r(k) voor verschillende tijds intervallen k.
Over het interpreteren van het autocorrellogram kunnen de volgende algemene opmerkingen gemaakt worden.

a) voor een serie random waarnemingen kan voor voldoend grote N aangetoond worden, dat r(k) bij benadering van de orde N(0,1/N) is, d.w.z. dat 19 van de 20 r(k) waarden tussen $\pm\sqrt{N}$ zullen liggen.

b) als het autocorrellogram van een waarnemings serie snel naar nul gaat, dan hebben we met een AR proces te doen.

c) als de waarnemings serie een steeds wisselend teken heeft, dan heeft het autocorrellogram dat ook.

d)als de waarnemings serie niet stationair is, dan nadert r(k) slechts zeer langzaam naar nul.

e) als de waarnemings serie een sinus vormig verloop heeft (dus periodiek is), dan vertoont r(k) een zelfde sinus vormig verloop. Het periodieke effect is uit de waarnemings serie weg te werken door het juiste periodieke gemiddelde van iedere waarneming af te trekken.

f) uitzonderlijke waarnemingen: een of meer hiervan kunnen een belangrijke invloed hebben op het autocorrellogram, het is dan ook verstandig ze te corrigeren.

g) tot slot valt nog op te merken dat het interpreteren van een auto correllogram aanzienlijk wat ervaring vraagt.

In figuur 8 is het autocorrellogram van de meet serie getekend. Hieruit zou een periodiciteit van 12 cq 25 uur kunnen volgen. Hetgeen bij nadere beschouwing van de meet serie niet blijkt op te gaan.

2.2.3. TREND ONDERZOEK.

Doormiddel van een tekentoets hebben we onderzocht of er een trend aanwezig is in de waarnemings serie. Hierbij wordt onderzocht hoe vaak een volgende waarneming groter is dan zijn voorganger. Als dit aantal in de buurt van 0.5*N ligt dan is er op grond van deze test geen reden om aan te nemen dat er een trend aanwezig is in de waarnemings serie.
Volgens de theorie is het 90% betrouwbaarheids interval voor N=8784:

tussen 4338 en 4444

We vonden echter dat 4209 maal de tweede waarde groter is dan de eerste. Dit zou beteken dat het gemiddelde van de waarnemings serie afneemt naar het einde toe. Of wel dat het gemiddelde over het begin van de meet serie hoger is dan over de hele serie. Dit is te verklaren uit het stormachtige

weer aan het begin van het jaar en daardoor een gemiddeld hoger tij. Als we hiervan uitgaan, kunnen we toch aannemen dat er geen trend is.

2.2.4. DOUBLE MASS CURVE (DMC).

Door $\geq$ x(t) uit te zetten tegen $\geq$ (x(t+k) is het mogelijk een indruk te krijgen van het aantal buiten gewone waarnemingen. In dit geval dus de tijdstippen dat er door bijzondere omstandigheden een extra hoge of juist juist een extra lage waterstand optreedt.
De keuze van k heeft invloed op het resultaat, k moet dus verstandig gekozen worden. Om dit duidelijk te maken zijn eerst wat voorbeelden gemaakt.
Uit figuur 9.Ic blijkt dat voor een ongestoord sinus vormig signaal (met waarden ook <0) de DMC een rechte lijn op levert als k=6, precies de periode van het signaal.
In de figuren 9.II en 9.III zijn nogmaals de DMC-en van twee signalen met een verstoring getekend, nu is er echter voor gezorgd dat de signalen geen waarden <0 aannemen door er een constante (> dan het minimum) bij op te tellen.
Ook nu weer blijkt dat het teken van de DMC-en alleen zinvol is als k=6 (de periode van de signalen).
Figuur 9.Ic geeft een keurige rechte lijn voor het ongestoorde signaal. In de figuren 9.IIc en 9.IIIc is duidelijk de invloed van de verstoringen op het signaal te zien.
Omdat voor het gegeven getij signaal geen exacte periode is aan te geven, is het niet mogelijk om m.b.v de DMC de storm perioden en de rustige perioden aan te wijzen.
In figuur 10 zijn, een gedeelte van, de DMC-en van het meetsignaal gegeven voor k=12 respectievelijk k=25.

2.2.5. HET VOORTSCHRIJDEND GEMIDDELDE (VRTG).

Wederom hebben we twee voorbeelden gegenereerd om de werking van het VRTG uit te leggen.
Het uitgangs punt is dat steeds van k opeenvolgende waarnemingen het gemiddelde wordt berekend, en wel volgens onderstaande formule:

$$VRTG\ (k)=\{0.5x(i)+x(i+1)+\ \ldots\ +x(i+k)+0.5x(i+k+1)\}/k$$

Uit de figuren 11.I en 11.II blijkt nu dat het voor het VRTG niet noodzakelijk is of het signaal positieve en/of negatieve waarden kent, door een constante bij het signaal op te tellen komt het alleen op een ander niveau te liggen.
Wel blijkt ook nu weer de keuze van k van belang te zijn.
Zolang k niet gelijk is aan T (de periode) blijkt dat het VRTG dezelfde frequentie vertoont als het signaal zelf.
Als k=T dan is het VRTG constant zolang er geen verstoringen optreden. Verstoringen zijn dus aan te wijzen door het afwijken van het constante niveau.
Maar weer is deze methode niet geschikt voor ons signaal omdat
Zoals ook blijkt uit figuur 12, het VRTG van het meetsignaal voor k=12 respectievelijk k=25.

2.2.7. CONCLUSIE.

Op de hiervoor aangegeven methoden is het niet mogelijk om aan te geven wat rustige perioden en wat storm perioden zijn. In het hierna volgende zullen we eerst de harmonische componenten van het meetsignaal aftrekken en vervolgens het rest signaal opnieuw beschouwen.
In een poging nu al iets te kunnen zeggen over het meetsignaal is een histogram (figuur 13) gemaakt. Maar

aangezien dit histogram min of meer een uniforme verdeling oplevert, is nog steeds geen eenduidige keuze te maken van welke periode rustig is en welke stormachtig. In $4.2.2 komen we hierop terug.

HOOFDSTUK III ARIMA MODEL VOOR HET MEET SIGNAAL.

3.1. WAARNEMINGS SERIE.

De waarnemings serie waarvoor we een ARIMA model gaan ontwikkelen, bestaat uit 156 waarnemeingen. Deze aan een sluitende serie waarnemingen is gekozen in een rustige periode, d.w.z. weinig verstoringen van het harmonische getij.
Hoe deze rustige periode is bepaald is aangegeven in 4.2.6, de periode loopt van uur 3001 tot en uur 3156.

zie figuur 14

3.2. ARIMA MODEL SINUS.

In figuur 15 is een sinus getekend met periode T=12. In figuur 16 is dezelfde sinus getekend met daarop een kleine verstoring. Voor dit signaal gaan we een ARIMA model (figuur 17) opstellen met behulp van het programma "BOX 0" uit de APL bibliotheek:

)lib 42 BOX1

Allereerst zijn de autocorrellogrammen van de 1 en 2 maal gedifferentieerde serie berekend en getekend.
Vervolgens ontwikkelen we het volgende model $(1\ 0\ 0)\times(0\ 1\ 0)_{12}$:

$$(1+\phi_1)\ z(t)=a(t)$$

$$z(t)+\phi_1 z(t-1)=a(t)$$

$$\{z(t)-z(t-12)\}+\phi_1\{z(t-1)-z(t-13)\}=a(t)$$

Door een maal te differentieren met periode 12, wordt de sinus van het signaal gehaald en wordt alleen de ruis nog gemodelleerd:

rest variantie =0.02
'significance' =0.42

In figuur 17.V zijn voorspellingen met dit model gedaan. Dit model voorspelt, zoals te verwachten, keurig de sinus.

In figuur 17.VI is gekeken naar een $(0\ 0\ 0)x(1\ 0\ 0)_{12}$ model:

$$z(t)=\phi_1 z(t-12)+a(t)$$

Gezien onze kennis van het signaal, zouden wij direct $\phi_1=1$ kiezen, het programma vindt echter: 0.8

restvariantie =7.28
'significance'=0.0

Vergelijken we de rest variantie en de 'significance' van deze twee modellen, dan komt het eerste model als beste uit de bus. Hetgeen wij al direct konden inzien.
Hoewel niet geheel juist zullen we de meet serie in ieder geval een maal differentieren met periode 12.

3.3. HET ARIMA MODEL.

Het ARIMA model is opgesteld m.b.v. een interactief APL programma. Hierbij is het alleen nodig de waarnemings reeks in te vullen en vervolgens een aantal vragen te beantwoorden.
Door dit een aantal keren te doen, volgens de voorgeschreven procedures, kan uit eindelijk tot een ARIMA model gekomen worden. Met dit model kunnen dan voorspellingen gedaan worden. Voor de hier gebruikte waarnemings serie vonden we uiteindelijk dat een $(1\ 0\ 0)x(1\ 1\ 0)_{12}$ model het best voldoet.

Het hierna volgende is een copy van het programma waarmee het bovenstaande ARIMA model is gevonden.

zie figuur 17B (I t/m V)

3.4. CONCLUSIE.

Uit de gevonden voorspellingen met hun 90% betrouwbaarheids interval, blikt dat dit model niet als geschikt beschouwd kan worden om rustig weer getij te voorspellen, laat staan de waaterstand gedurende een storm.
Al na een uur voorspellen is het betrouwbaarheids interval zo ruim (110.9-52.6=58.3) dat niets gezegd kan worden over het eventueel optreden van HHW (Hoog Hoog Water). Het zal duidelijkzijn dat dit model voornamelijk het periodieke deel van het signaal beschrijft en niet de verstoringen daarop (de betrouwbaarheids intervallen zijn een veelvoud van de verstoringen. Voor deze rustig weer periode liggen de verstoringen in de orde grootte:

$$|verstoring| < 15.0$$

HOOFDSTUK IV HARMONISCH DEEL BEREKEN

4.1 TOEPASSEN KLEINSTE KWADRATEN METHODE (KKM) OP HET GETIJ SIGNAAL.

Aangezien we hier op betrekkelijk eenvoudige manier de harmonische komponenten, die in het gemeten signaal voorkomen, willen bepalen, maken we gebruik van de KKM. Deze geeft voor routine analyses gewoonlijk zeer bevredigende resultaten.
De reden voor de keuze van slechts een eenvoudige methode is, dat we het getij signaal willen modelleren met ARIMA modellen. En zodra we een meer ingewikkelde methode kiezen, dan zouden we beter die ingewikkelde methode kunnen voortzetten voor het modelleren van hele signaal.

4.2. HET PRINCIPE VAN DE KKM.

Stel de waarnemings reeks $y(i)$ is gegeven en deze reeks willen we 'zo goed' mogelijk beschrijven met de gegeven functie $Y(i)=c*i+b*i^2$. De KKM geeft aan hoe b en c te bereken zijn zodanig dat $\sum \{y(i)-Y(i)\}^2$ zo klein mogelijk is.
Deze eis is als volgt om te zetten in eisen voor b en c:

1) minimaliseer $\sum \{y(i)-Y(i)\}^2=\sum \{y(i)-c*i-b*i^2\}^2$

2) de bovenstaande ongelijkheid is minimaal, als de afgeleiden naar b en c van deze ongelijkheid nul zijn:

$$\frac{dS}{db}\sum\{\ \ldots\ \}^2=0$$

$$\frac{dS}{dc}\sum\{\ \ldots\ \}^2=0$$

Dit levert twee vergelijkingen met twee onbekenden, waaruit de twee onbekenden zijn op te lossen, en waarmee een optimale keuze volgens de KKM gedaan is.

4.3. OPLOSSINGS METHODE.

Voor het harmonische getij geldt dat het een sommatie is van 'sinus'-en met verschillende amplituden, frequenties en fase veschuivingen:

$$x(t)=\sum_{i=1}^{n} A(i) * \text{sinus}(T(i) * t + F(i))$$

Hierin is:

- .x(t) het harmonische getij op tijdstip t
- .n het aantal harmonische componenten
- .A(i) de amplitude van de i' harmonische component
- .T(i) de hoeksnelheid " " " " in gr/uur
- .F(i) de fase veschuiving " " " in gr/uur

De uitdrukking voor x(t) is ook anders te schrijven m.b.v:

$$\sin(T+F)=\cos(F).\sin(T)+\sin(F).\cos(T)$$

dit geeft voor x(t):

$$x(t)=\sum \{p(i).\sin(T(i) * t)+q(i).\cos(T(i) * t)\}$$

met $p(i)=A(i).\cos(F(i))$
en $q(i)=A(i).\sin(F(i))$

Nu gaat het er om, om p(i) en q(i) zodanig te kiezen, bij

gegeven hoeksnelheden, dat het signaal zo goed mogelijk beschreven wordt.
Met de KKM gaat dit als volgt:

$$S=\sum \{y(t)-x(t)\}^2 \text{ ----> minimaliseer}$$

$$S=\sum \{y(t)-[p(i).\sin(T(i)*t)+q(i).\cos(T(i)*t)]\}^2$$

S is minimaal als de afgeleiden van S naar p(i) en q(i) nul zijn:

$$\frac{dS}{dp(i)}=-2\sum \{y(t)-[\ ...\ ]\} * \sin(T(i)*t)=0$$

uit werken voor i=1 geeft:

$$\frac{dS}{dp(1)}=\sum \{y(t)-[\ ...\ ]\} * \sin(T(1)*t)=0$$

reorganisatie van de termen geeft:

$$\sum y(t).\sin(t(1)*t)=\sum \sin(T(1)*t) * \{[\ ...\ ]\} =$$

$$p(1).\sin(T(1)*t).\sin(T(1)*t)+p(2).\sin(T(2)*t).\sin(T(1)*t)+ ...$$

$$+p(n).\sin(T(n)*t).\sin(T(1)*t)+ ...$$

$$+q(1).\cos(T(1)*t).\sin(T(1)*t)+ ...$$

$$+q(n).\cos(T(n)*t).\sin(T(1)*t)$$

We voeren nu de volgende notatie in:

y(t).sin(T(1)*t)= ysi1	$\sum$ ysi1 = Sysi1
sin(T(1)*t).sin(T(1)*t)=si1kw	$\sum$ si1kw = Ssi1kw
sin(T(2)*t).sin(T(1)*t)=si2si1	$\sum$ si1si2= Ssi2si1
cos(T(n)*t).sin(T(1)*t)=consi1	$\sum$ consi1= Sconsi1

etc.

Dit invullen geeft:

```
Sysi1=p(1)Ssi1kw +p(2)Ssi2si1+ ... +p(n)Ssinsi1+

     +q(1)Sco1si1+    ...          +q(n)Sconsi1
```

Door nu naar alle p(i) en q(i) te differentieren ontstaan 2n vergelijkingen met 2n onbekenden:

```
p(1),p(2), ... ,p(n),
q(1),   ...    ,q(n)
```

die hieruit op te lossen zijn.

Het stelsel ziet er als volgt uit:

```
Sysi1=p(1)Ssi1kw +p(2)Ssi2si1+ ... +p(n)Ssinsi1+ ...+q(n)Sconsi1
Sysi2=p(1)Ssi1si2+p(2)Ssi2kw + ...                  +q(n)Sconsi2
  .                                                      .
  .                                                      .
  .                                                      .
Syco1=p(1)Ssi1co1+                ...               +q(n)Sconco1
  .                                                      .
  .                                                      .
  .                                                      .
Sycon=p(1)Ssi1con+                ...                +q(n)conkw
```

Met behulp van de interactieve computer taal APL is dit stelsel eenvoudig op te lossen:

```
|p(1)|   | Sysi1 |       | Ssi1kw   Ssi2si1    ...        Sconsi1 |
|p(2)|   | Sysi2 |       | Ssi1si2  Ssi2kw     ...        Sconsi2 |
| .  |   |   .   |       |   .                                 .  |
| .  | = |   .   | ±:%   |   .                                 .  |
| .  |   |   .   |       |   .                                 .  |
|q(n)|   | Sycon |       | Ssi1con             ...        Sconkw  |
```

```
PQ=SYSC [:] |MAT|
```

Er kan nog opgemerkt worden dat |MAT| symmetrisch is.

4.4. RESULTATEN.

Rijks Waterstaat heeft het getij bij Vlissingen (in 1974) geanalyseerd met het programma HATYAN.
Hieruit zijn de 16 frequenties met de grootste amplituden gekozen, en bovendien zijn er 3 extra frequenties uit gekozen om een zekere spreiding in het frequentie domein te hebben. Hier onder volgen in tabel de gekozen frequenties, met de bijbehorende amplitude volgens de KKM en volgens HATYAN.

naam	hoeksnelheid (gr/uur)	amplitude KKM (cm)	amplitude HATYAN (cm)
SA	0.04106	10.98	6.41
O1	13.94303	9.62	11.04
K1	15.04106	6.07	6.61
NLK2	27.88607	3.12	8.14
NU2	27.96821	12.18	13.80
N2	28.43972	28.47	27.76
	28.51258	9.31	9.25
M2	28.98410	174.91	173.20
LABDA2	29.45560	5.41	6.05
2MN2	29.52847	12.23	10.55
S2	30.00000	46.79	48.23
K2	30.08213	11.62	13.65
M4	57.96821	13.05	12.38
MS4	58.98410	8.15	8.37
M6	86.95231	9.11	8.11
3MS6	116.95230	4.51	4.75
4MS10	145.93640	1.79	1.57
5MS12	174.92050	1.16	0.87

Trekken we deze harmonischen af van het gemeten signaal, dan houden we een restsignaal over. In dit rest signaal blijken nog veel te duidelijk harmonischen componenten te zitten. Vandaar dat besloten is om 19 nieuwe frequenties extra uit te selecteren, om zo te proberen het merendeel van de harmonischen uit het rest signaal te verwijderen.
Dus dezelfde programma's worden gebruikt met echter 19

nieuwe frequenties en het hiervoor gevonden rest signaal als invoer.
Dit leidt tot de volgende resultaten:

zie figuur 18

In de tabel in figuur 19 is voor 8 frequenties een vegelijkend onderzoek gedaan naar de resultaten van de KKM, door deze te vergelijken met de waarden die volgens de 'Getij tafels voor Nederland 1976' zouden moeten gelden.
De benodigde astronomische argumenten (V +u) en de correctie factor voor de amplitude (f) zijn gehaald uit het dictaat b75 (getij analyse).

zie figuur 19 en 20

De verschillen tussen deze twee tabellen geven geen aanleiding om de resultaten van de KKM te herzien.

4.5. TWEE MAAL KKM, TOEGESTAAN?

Zoals uit het voorgaande blijkt zijn er eerst 19 harmonischen m.b.v de KKM tegelijk bepaald en van het meet signaal afgetrokken. Noemen we het signaal wat overbleef het verschil signaal, dan hebben we dezelfde procedure herhaald met het verschil signaal en 19 nieuwe harmonischen.
De redenen dat we dit deden, en niet ineens met 38 harmonischen werkten, zijn:

a) het programma zou aangepast moeten worden

b) de benodigde rekentijd zou vier maal zo veel worden

c) de gemaakte fout vermoedelijk klein is.

Dit laatste zullen we trachten aan te tonen aan de hand van een paar reken voorbeelden.
Hierna volgen 4 tabellen (zie figuur 21). Deze tabellen bevatten 4 verschillende benaderingen m.b.v. de KKM:

a) beginnend met de frequentie met de grootste amplitude, de fase veschuiving en de amplitude hiervan berekenen, en het gevonden resultaat van het gegeven signaal aftrekken. Dit vervolgens herhalen voor de frequentie met de een na grootste amplitude. Enzovoort..

b) hetzelfde als onder a). Maar nu echter beginnend met de frequentie met de kleinste amplitude.

c) nu worden de respectievelijke amplituden en fase verschuivingen steeds berekend uit het complete meet signaal.

d) in deze tabel staan de resultaten van het toepassen van de KKM voor de 11 frequenties tegelijk.

De verschillen tussen de gevonden resultaten in de vier tabellen zijn dermate klein, dat op grond hiervan vermoed kan worden dat de KKM een sinus waarvan de frequentie gegeven is, hoe dan ook op de best mogelijke manier uit het

signaal. En dat deze best mogelijke manier zeer eenduidig is met betrekking tot de amplitude en de fase verschuiving behorend bij de gegeven frequentie.
Op grond hievan is aangenomen dat door de hier gebezigde werkwijze geen grote fouten gemaakt zijn.

4.6. HET LAAG DOOR LAAT FILTER.

Door van het gegeven signaal de 38 hiervoor bepaalde harmonischen af te trekken, ontstaat het rest signaal.
In dit rest signaal blijken nog hoog frequente componenten te zitten. Aangezien we, bij het verdere onderzoek, daar last van zullen hebben, en deze hoog frequente componenten slechts een zeer kleine amplitude hebben, is besloten deze componenten van het signaal te filteren doormiddel van een laag door laat filter.

Van een laag door laat filter wordt geeist dat de versterking over het frequentie gebied $0<v<v_0$ constant is, terwijl deze voor hogere frequenties $v>v_0$ gelijk aan nul moet zijn. Voor de overbrengings verhouding van een laag door laat filter geldt:

$$|H(v)|=K \qquad 0<|v|<v.$$
$$=0 \qquad |v|>v.$$
$$\arg H(v)=-2\pi v t_0$$

Hieruit volgt voor H(v):

$$H(v)=K\ e^{-jsnvt_0} \qquad 0<|v|<v_0$$
$$=0 \qquad |v|>v_0$$

De impuls responsie h(t) van het laag door laat filter volgt door terug transformatie van H(v):

$$h(t)=F^{-1}\{H(v)\}=\,^{v}_{v_0}{}^{02}\ K\ e^{-j2nvt_0}e^{-j2nvt}dv=K\ \sin(2\pi v_0(t-t_0)/\pi(t-t_0)$$

zie figuur 22

We kiezen dat we lage frequenties met een periode kleiner dan T=5 uur met K=1 uit het rest signaal willen filteren, dan is v_0 in dit geval:

$$v_0=1/T_0=1/5=0.200$$

Het gefilterde signaal op $t=t_0$ kan nu berekend worden door:

$$filter(t_0)= \int h(t_0-\tau).r(t_0)d\tau \qquad r(t)=\text{rest signaal}$$

Aan gezien integreren tussen plus en min oneindig niet mogelijk is, zullen we ons moeten beperken. Deze keuze is gemaakt door te onderzoeken na hoeveel tijdstappen de waarde van h(t) klein is, terwijl er bovendien voor gezorgd moet worden dat:

$$\sum_{i=-t}^{-t} h(i)=1$$

Uit onderzoek is gebleken dat een 33 punts filter hier in dit geval redelijk aan voldoet:

zie figuur 23

Met dit filter is het rest signaal vermenigvuldigd, en zo zijn de hoge frequenties van het rest signaal gefilterd.

4.7. VERGELIJKING MEET EN FILTER SIGNAAL.

Uit het in figuur 28 opgenomen histogram blijkt de invloed van alle voorgaande bewerkingen op het meet signaal.
Het histogram van het meet signaal lijkt min of meer op een uniforme verdeling, terwijl het histogram van het gefilterde signaal meer voldoet aan een normale verdeling. Dit laatste betekent dat we er inderdaad redelijk in geslaagd zijn om het deterministische deel van het signaal te halen.
De figuren 24, 25 en26 geven het meet, bereken,verschil en

filter signaal weer. Hierin is te zien dat een belangrijk gedeelte van de harmonische componenten van het meet signaal gehaald is.
Uit figuur 27 blijkt dat nogmaals uit de autocorrellogrammen van van het 'verschil signaal' en het 'filter signaal'.

HOOFDSTUK V ARIMA MODELLEN.

5.1. KEUZE RUSTIGE EN STORM PERIODE.

De hierna te ontwikkelen modellen moeten gebruikt gaan worden om hoge waterstanden te voorspellen. Hierbij heeft zo'n voorspelling alleen zin als er met voldoende nauwkeurigheid enige uren vooruit voorspeld kan worden, in ieder geval zo'n 3 a 4 uur.
Er wordt voor twee verschillende meteorologische omstandigheden een ARIMA model ontwikkeld:

a) een periode van rustig en stabiel weer

b) een storm periode met zeer hoge waterstanden

Uit het voorgaande histogram voor het gefilterde rest signaal blijkt dat 75% van de waarden tussen +15 en -15 cm ligt. Een langere aan een gesloten periode (meer dan 12 uur) met waarden tussen +15 en -15 cm definieren we hier als een rustige periode.
Verder blijkt dat 97% tussen +60 en -60 cm ligt. Aangezien we geinteresseerd zijn in zeer hoge waterstanden kijken we voor storm perioden naar die perioden waarin waarden groter dan +60 cm voorkomen:

zie figuur 29

zie figuur 30

De modellen worden weer ontwikkeld m.b.v. de interactieve APL programma's.

5.2. RUSTIG WEER MODEL.

Allereerst zijn de autocorrellogrammen voor de respectievelijk 0,1,2 en 3 maal gedifferentieerde serie berekend (zie figuur 31,32 en 33).
In de verschillende figuren is te zien dat er nog steeds iets van periodiciteit in de rustig-weer-serie aanwezig is. Uit alle 4 de figuren zou een periodiciteit van 12 of 13 uur kunnen volgen. Echter het patroon tussen 0 en 12 uur herhaalt zich niet tussen 12 en 24 uur, dus van een eenduidige periodiciteit is geen sprake. Dit maakt een model keuze opgrond van het autocorrellogram zeer moeilijk.
Het APL pakket biedt de mogelijkheid om met behulp van een automatisch programma alle volgende, niet periodieke, modellen te testen:

(1 0 0) (0 0 1) (1 0 1)
(2 0 0) (0 0 2) (2 0 2)
(3 0 0) (0 0 3) (3 0 3)
(1 1 0) (0 1 1) (1 1 1)
(2 1 0) (0 1 2) (2 1 2)
(3 1 0) (0 1 3) (3 1 3)

De keuze welke van de gevonden modellen het 'best' is, wordt gedaan door te kijken:

a) voor welk model de variantie van de rest serie het kleinst is

b) naar de 'significance', dit is een maat voor het al dan niet witte ruis zijn van de rest serie. De waarde "1" betekent "witte ruis", waarden tussen "0.0 en 0.1" beteken "geen witte ruis".

Behalve de bovenstaande 18 modellen, zijn deze modellen ook nog voor 3 verschillende situaties bekeken:

1) 10 waarnemingen gegeven: t=21 t/m t=30

2) 20 waarnemingen gegeven: t=11 t/m t=30

3) 30 waarnemingen gegeven: t=1 t/m t=30

In figuur 34 zijn de belangrijkste resultaten opgenomen. Voor 10 waarnemingen gegeven springen 2 modellen, figuur 35 en 36, in het oog:

- (3 1 0) rest variantie 1.08
  significance 0.35

Hoewel dit model als beste uit de bus kwam, wijken de voorspellingen geheel af van de werkelijke waarde.

-(2 1 0) rest variantie 2.60
  significance 0.71

Dit model blijkt een veel betere voorspelling te geven. Voor de situatie waarin 20 respectievelijk 30 waarnemingen gegeven zijn, komt in beide gevallen ook een (3 1 0) model als beste uit de bus (zie figuur 37 en 38).
Uit figuur 34 blijkt dat het (2 1 0) en het (3 1 0) model steeds de beste twee zijn.
Deze twee modellen (modellen van deze orde) gaan we nu toepassen op het hele signaal van 96 waarnemingen (variantie 20.15)

	(3 1 0)	(2 1 0)
figuur	-39-	-40-
rest variantie	0.79	1.58
significance	0	0
betrouwbaarheid na 3 uur	±6,1cm	±5,4cm (80% interval)
orgineel / computed	goed	ruim voldoende

Omdat de variantie van de rest serie zo klein is (4% van het gegeven signaal), is het feit de rest serie geen witte ruis is, slechts van weinig invloed op de resultaten van het model.
Beide modellen beschrijven de gegeven serie goed (figuur 39.III en 40.V), echter de voorspellingen met beide modellen zijn van weinig nut. De amplitude van het gegeven signaal is 15 cm, terwijl de voorspellingen na drie uur al een spreiding vertonen van ±6,1cm respectievelijk ±5.3cm.

5.3. CONCLUSIE.

Hoewel de gegeven serie een rustige en stabiele periode beslaat, is het op de hier toegepaste manier niet mogelijk om tot een model keuze te komen zodanig dat voorspellingen met grote nauwkeurigheid mogelijk zijn (zie figuur 40.II en 39.II)
Als laatste controle om te zien hoe goed de modellen opzich zijn, zijn de theoretische autocorrellogrammen berekend (figuur 41), volgend uit de AR coefficienten (m.b.v. de Yule-Walker vergelijkingen).
Vergelijking van deze correllogrammen met die van het 1 maal gedifferentieerde gegeven rest signaal (figuur 32B), tonen aan dat het auto correllogram van het (3 1 0) model zeer goed overeenkomt.
Dus het gegeven signaal wordt opzich wel degelijk goed beschreven, alleen voorspellen met het model levert niet veel op.

5.4. STORM MODEL.

In figuur 42 is een plot opgenomen van de waarnemings serie. Voor een analyse van de stormvloed nemen we dat deel van de serie met waarden groter dan 25cm (dat zijn er 32 achter elkaar).
Vervolgens hebben we de autocorrellogrammen voor de 0, 1 en 2 maal gedifferentieerde serie berekend (figuur 43).
Hierin is geen periodiciteit meer te herkennen. We kunnen nu weer de in $5.2 aangegeven 18 modellen uit testen.
We proberen deze modellen voor de eerste 10, 12 respectievelijk 18 waarnemingen. In figuur 46 is een overzicht gegeven van de modellen die in alle drie de situaties mogelijk zijn.

Uit de gevonden modellen blijkt dat een (2 1 2) model in

alle drie de gevallen het best voldoet. Dit model gaan we toepassen op alle 32 waarnemingen (zie figuur 47).
Dit levert voor de AR termen:

$$\phi_1 = 1.03 \text{ en } \phi_2 = -0.42$$

en voor de MA termen:

$$\theta_1 = -0.30 \text{ en } \theta_2 = -0.27$$

De variantie van de gegeven serie is 170.45 en de variantie van de nu gevonden rest serie is 8.13; de 'significance' is 0.83, wat betekent dat het rest signaal als witte ruis beschouwd mag worden.

In figuur 47.II zijn 10 voorspellingen gedaan met dit model. Na 3 uur is er al een spreiding van +27.07 cm op de voorspelling (80% betrouwbaarheids interval), terwijl de spreiding van het gegeven signaal +(max-min)/2=+53.0 cm. Dit maakt duidelijk dat de waarde van de voorspelling niet groot is.
Uit figuur 47.III blijkt wel dat het gevonden model de meet serie goed beschrijft.
Tot slot zijn nog de theoretische autocorrellogrammen berekend volgend uit respectievelijk de 2 AR termen en de 2 MA termen (figuur 47.IV). Samen gevoegd leveren deze twee figuren min of meer het autocorrellogram uit figuur 43.IIA.
Hoewel de voorspellingen niet veel hoop geven, gaan we met dit model toch in het volgende hoofdstuk een voorspelling simuleren.

5.5. SLOT CONCLUSIE.

De modellen voor 'rustig weer' en voor 'storm' zijn duidelijk verschillend. Voor het beschrijven van de serie zal dus steeds een keuze gemaakt moeten worden of het 'rustig weer model' of het 'storm model' toegepast moet worden.

[illegible] gelijk aan de waarden [illegible] moet voldoen. Dit [illegible] gedaan [illegible] toegepassen [illegible] (zie figuur 47).
Dit levert voor de ϕ termen:

$$\phi_1=1.03 \text{ en } \phi_2=-0.42$$

En voor de θ termen:

$$\theta_1=-0.30 \text{ en } \theta_2=-0.27$$

De variantie van de gegeven serie is 17.4 en de variantie van de nu gevonden rest serie is 6.13; de 'similiteit' is 0.6[illegible], wat betekent dat het rest signaal als witte ruis beschouwd [illegible] worden.

In figuur 47.II zijn de voorspellingen gemaakt met dit model. Na[illegible] figuur [illegible] een spreiding van +27.0[illegible] de voorspelling (95% betrouwbaarheids interval), terwijl de [illegible] van het gegeven signaal +(max-min)/2=+53.0 cm. Dit maakt dat blijkbaar de waarde van de voorspelling niet groot is.
Uit figuur 47.III blijkt wel dat het gevonden model de reeks goed beschrijft.
Tot slot zijn nog de theoretische autocorrelatiefuncties berekend volgens uit respectievelijk de ϕ termen en de θ termen (figuur 47.IV). [illegible] deze twee figuren [illegible] het autocorrelogram uit figuur 43.III.
Hoewel de voorspellingen niet veel hoop geven, kan we met dit model toch in het vervolg goede voorspellingen [illegible] simuleren.

5.6. SLOT CONCLUSIE.

De modellen voor 'rustig weer' en voor 'storm' zijn duidelijk verschillend. Voor het beschrijven van de serie zal dus steeds een keuze gemaakt moeten worden of het 'rustig weer model' of het 'storm model' toegepast moet worden.

HOOFDSTUK 6 SIMULATIE

6.1. HET MODEL.

Het in het vorige hoofdstuk gevonden model voor de storm periode ziet er als volgt uit:

$$(1-\phi_1 B-\phi_2 B^2)\nabla z(t)=(1-\theta_1 B-\theta_2 B^2)a(t)$$

$$\nabla z(t)-\phi_1 \nabla z(t-1)-\phi_2 \nabla z(t-2)= a(t)-\theta_1 a(t-1)-\theta_2 a(t-2)$$

$$z(t)-z(t-1)-\phi_1 \{z(t-1-z(t-2)\}-\phi_2 \{z(t-2)-z(t-3)\}= a(t)-\theta_1 a(t-1)-\theta_2 a(t-2)$$

$$z(t)=(1+\phi_1)z(t-1)+(\phi_2-\phi_1)z(t-2)-\phi_2 z(t-3)+ +a(t)-\theta_1 a(t-1)-\theta_2 a(t-2)$$

met $\phi_1=1.030$ $\phi_2=-0.42$
$\theta_1=-0.30$ $\theta_2=-0.27$

Dit invullen geeft:

$$z(t)=\underbrace{2.03}_{Y_1}z(t-1)\underbrace{-1.45}_{Y_2}z(t-2)+\underbrace{0.42}_{Y_3}z(t-3)+a(t)+0.30a(t-1)+0.27a(t-2)$$

Hetgeen we hierna gaan toepassen.

6.2. DE SIMULATIE.

De eerste kolom van figuur 48 bevat 27 waarnemingen z(t), t=1,2,...,27 ,die een storm periode weergeven.
Kolom 2 is 2.03*z(t-1), kolom 3 is -1.45*z(t-2) en kolom 4 is 0.42*z(t-3).

Dit zijn de coefficienten zoals hierboven gevonden.
Stel we willen vanaf t=9 gaan voorspellen met het hiervoor gevonden model, dan kunnen we als volgt te werk gaan:

1) we beginnen met a(4) te bereken, onder de aanname dat a(1)=a(2)=a(3)=0 . Voor a(4) vinden we:

$$a(4)=z(4)-Y_1 z(3)-Y_2 z(2)-Y_3 z(1)=14-13.76=0.24$$

2) we bereken vervolgens a(5) t/m a(9):

$$a(5)=z(5)-Y_1 z(4)-Y_2 z(3)-Y_3 z(2)+\theta_1 a(4)=-4.73$$

$$a(6)=z(6)-Y_1 z(5)-Y_2 z(4)-Y_3 z(3)+\theta_1 a(5)+\theta_2 a(4)=-3.1$$

....

$$a(7)=4.03 \quad a(8)=5.69$$

....

$$a(9)=z(9)-Y_1 z(8)-Y_2 z(7)-Y_3 z(6)+\theta_1 a(8)+\theta_2 a(7)=0.56$$

3) in figuur 49 bereken we ieder volgend uur steeds 3 uur vooruit vanaf t=9.

In onderstaande tabel is het resultaat van het voorspellen

samen gevat:

t	3 u. voorsp.	2 u. voorsp.	1 u. voorsp	meet waarde
10	--	--	23.32	22
11	--	24.53	26.82	28
12	21.44	27.65	30.03	36
13	26.48	30.56	43.83	51
14	31.10	50.15	66.85	67
15	53.36	78.81	79.16	77
16	84.47	85.00	79.97	77
17	85.92	78.25	71.33	77
18	75.22	65.49	77.90	88
19	61.86	80.36	103.89	109
20	82.51	118.36	130.85	124
21	126.59	145.90	129.95	123
22	152.25	127.93	111.74	108
23	123.35	68.98	89.97	91

Een vergelijking van de kolommen maakt duidelijk dat 3 uur voorspellen weinig zinvolle waarden geeft. Twee uur voorspellen geeft al verbetering, terwijl een uur voorspellen redelijke waarden oplevert.

6.3. CONCLUSIE.

Het beschrijven van het rest signaal met een ARIMA model gaat goed. Het doen van voorspellingen (meerdere uren vooruit) met zo'n model levert echter geen bruikbare resultaten op.