

# Tijdreeksanalyse voor (eco)hydrologen

## Achtergronddocumentatie en cursushandleiding





14 september 2005

# Tijdreeksanalyse voor (eco)hydrologen

Achtergronddocumentatie en cursushandleiding

© 2005 Kiwa N.V.  
Alle rechten voorbehouden.  
Niets uit deze uitgave mag  
worden vervoelvoudigd,  
opgeslagen in een  
geautomatiseerd  
gegevensbestand, of  
openbaar gemaakt, in enige  
vorm of op enige wijze,  
hetzij elektronisch,  
mechanisch, door  
fotokopieën, opnamen, of  
enig andere manier, zonder  
voorafgaande schriftelijke  
toestemming van de  
uitgever.

Kiwa N.V.  
Water Research  
Groningenhaven 7  
Postbus 1072  
3430 BB Nieuwegein

Telefoon 030 60 69 511  
Fax 030 60 61 165  
Internet [www.kiwa.nl](http://www.kiwa.nl)



# Colofon

**Titel**

Tijdreeksanalyse voor (eco)hydrologen,  
Achtergronddocumentatie en cursushandleiding

**Projectnummer**

306.449.400

**Projectmanager**

Jan Willem Kooiman

**Kwaliteitsborger**

Kees Maas

**Auteurs**

Jos von Asmuth  
Martin Knotters  
Kees Maas



Deze handleiding is niet openbaar en slechts verstrekt aan de deelnemers van de cursus.  
Eventuele verspreiding daarbuiten vindt alleen plaats door Kiwa en Alterra zelf.

# Voorwoord

Dit document bevat de achtergronddocumentatie en handleiding behorende bij de cursus 'tijdreeksanalyse voor (eco)hydrologen'. De inhoudelijke uitwerking van deze cursus is verzorgd door Kiwa en Alterra, in een eendrachtig samenwerkingsverband. Een wat ons betreft bijzonder aspect van de cursus is dat de omlijning van de inhoud ervan voor een groot deel bepaald is door de deelnemers zelf, te weten een consortium van waterschappen.

Ook de praktische organisatie van de cursus heeft in handen gelegen van de deelnemers, onder aanvoering van de waterschappen Velt en Vecht (in het bijzonder Thomas de Meij, van wie het initiatief is uitgegaan en die de 'administratieve lasten' dapper heeft gedragen) en Groot-Salland (Francis de Graaf, die de cursus onderdak heeft geboden en de rol van gastvrouw heeft willen vervullen). Onze bijzondere dank gaat dan ook uit naar hen, maar natuurlijk ook naar de andere deelnemers die inbreng hebben geleverd en daarmee deze cursus daadwerkelijk mogelijk hebben gemaakt. We hopen dat al deze inbreng en betrokkenheid hebben geleid tot een cursus die goed past bij de vragen en kennisbehoeften van iedereen die erbij aanwezig zal zijn.

Jos von Asmuth  
Martin Knotters  
Kees Maas





# Inhoud

|          |                                                                       |           |
|----------|-----------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Inleiding</b>                                                      | <b>5</b>  |
| <b>2</b> | <b>Tijdreeksanalyse vanuit statistisch perspectief</b>                | <b>7</b>  |
| 2.1      | Inleiding                                                             | 7         |
| 2.2      | Over tijdreeksen, tijdreeksanalyse en haar toepassingen               | 7         |
| 2.3      | Enkele basisbegrippen                                                 | 8         |
| 2.3.1    | Reeksen, processen en modellen, continu en discreet                   | 9         |
| 2.3.2    | Statistische begrippen                                                | 11        |
| 2.4      | Eenvoudige, lineaire regressie                                        | 13        |
| 2.5      | Over geheugen en tijdreeksen                                          | 14        |
| 2.6      | Lineaire, stationaire processen                                       | 14        |
| 2.6.1    | Autoregressieve processen (AR)                                        | 14        |
| 2.6.2    | Moving-Averageprocessen (MA)                                          | 16        |
| 2.6.3    | Autoregressive-Moving Average processen (ARMA)                        | 17        |
| 2.7      | Niet-stationaire lineaire processen                                   | 18        |
| 2.7.1    | Verschilwaarden                                                       | 18        |
| 2.7.2    | Autoregressive Integrated Moving Average processen (ARIMA)            | 18        |
| 2.7.3    | Seizoen-nietstationaire processen                                     | 18        |
| 2.7.4    | Seasonal Autoregressive Integrated Moving Average processen (SARIMA)  | 19        |
| 2.8      | Samenhang tussen reeksen                                              | 19        |
| 2.8.1    | Inleiding                                                             | 19        |
| 2.8.2    | Cross-covariantie, cross-correlatie                                   | 20        |
| 2.8.3    | Transfer-ruismodellen                                                 | 20        |
| <b>3</b> | <b>Tijdreeksanalyse vanuit fysisch-hydrologisch perspectief</b>       | <b>23</b> |
| 3.1      | Over statistiek en fysica                                             | 23        |
| 3.2      | De transferfunctie van een Box-Jenkinsmodel                           | 24        |
| 3.3      | Van Box-Jenkins- naar PIRFICT-model                                   | 25        |
| 3.4      | De responsfunctie van verschillende typen systemen                    | 28        |
| 3.4.1    | Een bodemkolom: 1-dimensionale stroming                               | 28        |
| 3.4.2    | Een poldersysteem: 2- dimensionale stroming                           | 29        |
| 3.4.3    | De onverzadigde zone: Convectie en Dispersie                          | 30        |
| 3.4.4    | Gekoppelde systemen: een convolutie van responsfuncties               | 31        |
| 3.4.5    | De Pearson III functie: een hybride oplossing                         | 33        |
| 3.5      | Wanneer is een grondwatersysteem niet-linear?                         | 34        |
| 3.6      | Tijdreeksanalyse met drempel-nietlineariteit                          | 35        |
| <b>4</b> | <b>Tijdreeksanalyse met meerdere verklarende reeksen</b>              | <b>37</b> |
| 4.1      | Verschillende typen invloedsfactoren                                  | 37        |
| 4.2      | De responsfunctie en interpretatie van verschillende invloedsfactoren | 37        |
| 4.2.1    | Het type PREC oftewel neerslag(overschot)                             | 37        |

|          |                                                                           |           |
|----------|---------------------------------------------------------------------------|-----------|
| 4.2.2    | Het type EVAP oftewel verdamping                                          | 40        |
| 4.2.3    | Het type WELL oftewel onttrekkingen                                       | 41        |
| 4.2.4    | Het type River oftewel oppervlaktewaterpeil (polder-, rivier- of zeepeil) | 44        |
| 4.2.5    | De typen Impulse en BOXJEN                                                | 46        |
| 4.3      | De effecten van hydrologische ingrepen: interventieanalyse                | 47        |
| 4.4      | Trendanalyses                                                             | 48        |
| 4.5      | Het ruismodel                                                             | 49        |
| 4.5.1    | De residuen van het transfermodel                                         | 49        |
| 4.5.2    | Wat is autocorrelatie?                                                    | 49        |
| 4.5.3    | Belang van het ruismodel                                                  | 49        |
| <b>5</b> | <b>Het proces van tijdreeksmodellering</b>                                | <b>51</b> |
| 5.1      | Inleiding                                                                 | 51        |
| 5.2      | In den beginne: de gegevens                                               | 51        |
| 5.2.1    | Opschonen van de dataset                                                  | 51        |
| 5.2.2    | Aanlooperperioden en beginvoorwaarden                                     | 52        |
| 5.2.3    | Over het interpoleren en 'sampelen' van de data                           | 53        |
| 5.3      | Modelidentificatie, calibratie en verificatie                             | 53        |
| 5.4      | Automatische criteria voor modelselectie                                  | 54        |
| 5.5      | Beoordelen van modellen vanuit statistisch perspectief                    | 55        |
| 5.5.1    | Verificatie of validatie?                                                 | 55        |
| 5.6      | Beoordelen van modellen vanuit fysisch-hydrologisch perspectief           | 56        |
| 5.6.1    | De drainagebasis                                                          | 57        |
| 5.6.2    | De verdampingsfactor                                                      | 58        |
| 5.6.3    | Ruimtelijke samenhang van de resultaten                                   | 58        |
| 5.7      | Het simuleren van grondwaterstanden                                       | 60        |
| 5.7.1    | Toepassingen                                                              | 60        |
| 5.7.2    | Stochastisch versus deterministisch simuleren                             | 60        |
| <b>6</b> | <b>Monitoring en tijdreeksmodellering</b>                                 | <b>63</b> |
| 6.1      | Inleiding                                                                 | 63        |
| 6.2      | Monitoring in ruimte, tijd en ruimte en tijd                              | 63        |
| 6.2.1    | Waarom monitoren?                                                         | 63        |
| 6.2.2    | Het ontwerpen van een monitoringstrategie                                 | 64        |
| 6.2.3    | Strategieën voor monitoring in ruimte                                     | 65        |
| 6.2.4    | Strategieën voor monitoring in tijd                                       | 65        |
| 6.2.5    | Strategieën voor monitoring in ruimte-tijd                                | 66        |
| 6.3      | Monitoring voor tijdreeksmodellen                                         | 66        |
| 6.3.1    | Het schatten van modelparameters                                          | 66        |
| 6.3.2    | Waarnemingsfrequentie voor het schatten van trends                        | 69        |
| 6.3.3    | Interventiemodellen                                                       | 72        |
|          | <b>Bibliografie</b>                                                       | <b>73</b> |



## Dag 1 - De basis van tijdreeksmodellering

| Tijdstip      | Programma                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | Docent | Handl. |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|--------|
| 9.15 - 9.30   | Ontvangst met koffie                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |        |        |
| 9.30 - 10.00  | <b>Inleiding</b> <ul style="list-style-type: none"> <li>- Introductie, kennismaking</li> <li>- Doelstellingen, overzicht programma</li> </ul>                                                                                                                                                                                                                                                                                                                                                                | JvA    |        |
| 10.00 - 10.45 | <b>Theorieblok: statistische tijdreeksanalyse</b> <ul style="list-style-type: none"> <li>- Tijdreeksanalyse vanuit statistisch perspectief</li> <li>- Verschil met andere methoden (m.n. multiple regressie en fysische modellen als SWAP)</li> <li>- Wat kun je met tijdreeksanalyse?<br/>(klimaatonafhankelijke grondwaterkarakteristieken, invloed van externe factoren)</li> </ul>                                                                                                                       | MK     | H-2    |
| 10.45 - 11.00 | Koffiepauze                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |        |        |
| 11.00 - 11.30 | <b>Practicumblok: kennismaking met Menyanthes</b> <ul style="list-style-type: none"> <li>- Wat is Menyanthes?</li> <li>- Menyanthes als 'datamining'-tool</li> <li>- input en output van het model</li> <li>- beoordelen van de reeksen voordat je gaat modelleren</li> <li>- handmatig fitten tijdreeksmodel</li> </ul>                                                                                                                                                                                     | JvA    | P-1    |
| 11.00 - 11.30 | <b>Theorieblok: hydrologische tijdreeksanalyse</b> <ul style="list-style-type: none"> <li>- Tijdreeksanalyse vanuit hydrologisch perspectief</li> <li>- Continu versus discreet</li> <li>- Omgaan met data</li> <li>- Aanlooperperioden en beginvoorwaarden</li> </ul>                                                                                                                                                                                                                                       | JvA    | H-3    |
| 11.30 - 12.15 | <b>Practicumblok: Box-Jenkins versus Pirfict</b> <ul style="list-style-type: none"> <li>- Modellering van een makkelijke reeks</li> <li>- modellering van een reeks met Box-Jenkins en Pirfict</li> <li>- resampling van de input</li> <li>- verschillende frequenties (bijvoorbeeld van een dagreeks een tweede reeks met kwartaalcijfers maken en dan verschillen in model beoordelen)</li> <li>- lezen van een tijdreeks voor modellering; reactiesnelheden systeem, effect van peilbeheersing</li> </ul> | JvA    | P-2    |
| 12.15 - 13.00 | Lunch                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |        |        |
| 13.00 - 13.30 | <b>Theorieblok: Statistische Beoordeling van modellen</b> <ul style="list-style-type: none"> <li>- Beoordeling van modellen: wanneer is een model goed (vuistregels voor grootte verklaarde</li> </ul>                                                                                                                                                                                                                                                                                                       | MK     | H-5.5  |

|               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |     |       |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-------|
|               | variantie, Akaike's information criterion, beoordeling lineariteit en stationariteit)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |     |       |
| 13.30 – 14.00 | <b>Theorieblok vervolg: Hydrologische beoordeling van modellen</b> <ul style="list-style-type: none"> <li>- Hydrologische interpretatie en beoordeling van modellen</li> <li>- Ruimtelijk patroon</li> </ul>                                                                                                                                                                                                                                                                                                                                                            | JvA | H-5.6 |
| 14.00 – 14.45 | <b>Practicumblok 3: Oprekken van een vierjarige meetreeks naar een dertigjarige tijdreeks</b> <ul style="list-style-type: none"> <li>- principes en praktijk van klimaatonafhankelijke reeksen</li> <li>- beoordelen van een reeks voordat je gaat modelleren; is dit wel geschikt om te modelleren?</li> </ul>                                                                                                                                                                                                                                                         | JvA | P-3   |
| 14.45 – 15.15 | <b>Vervolg theorieblok: Monitoring in ruimte en tijd</b><br>Docenten: Martin Knotters <ul style="list-style-type: none"> <li>- eisen aan de data</li> <li>- hoe lang moet je gemeten hebben (vuistregels)</li> <li>- frequentie van de waarnemingen, kwartaalcijfers tegenover divers (mag ik drie jaar kwartaalcijfers gebruiken voor een tijdreeksmodel)</li> <li>- ongelijke frequentie van meetreeksen, hoe ga je daar mee om?</li> <li>- resamplen van de input; waarom, wat doet het?</li> <li>- welke fouten kom je tegen in de opzet van meetnetten?</li> </ul> | MK  | H-6   |
| 15.15 – 15.30 | Theepauze                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |     |       |
| 15.30 – 16.30 | <b>Practicumblok: Modelleren effect van wateraanvoer</b><br>Docent: Martin Knotters<br>Case Study 3: <ul style="list-style-type: none"> <li>- interventieanalyse 'Physical basis...'</li> </ul>                                                                                                                                                                                                                                                                                                                                                                         | MK  | P-4   |

## Dag 2 - Verdieping en praktische problemen

| Tijdstip      | Programma                                                                                                                                                                                                                                                                                                                                                                          | Docent | Handl. |
|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|--------|
| 9.15 - 9.30   | Ontvangst met koffie                                                                                                                                                                                                                                                                                                                                                               |        |        |
| 9.30 - 10.15  | <b>Theorieblok: verdieping achtergronden</b> <ul style="list-style-type: none"> <li>- niet-lineariteit (wanneer nodig, wanneer overbodig)</li> <li>- overzicht van beschikbare tijdreeksprogramma's en toepassingen, dwarsverbanden en ontwikkelingen (Menyanthes, Vidente, model van TNO)</li> <li>- Kalman-filter (wat is dat nou weer?)</li> </ul>                              | MK     | H-3.5  |
| 10.15 - 10.45 | <b>Theorieblok: Meerdere verklarende reeksen</b> <ul style="list-style-type: none"> <li>- principes van modellering van externe invloeden (winning, rivier, wateraanvoer)</li> <li>- uit elkaar rafelen van een reeks in componenten</li> <li>- grootte van de invloed van een invloedsfactor vaststellen</li> </ul>                                                               | JvA    | H-4    |
| 10.45 - 11.15 | <b>Practicumblok: verdieping van de modellering met Menyanthes</b> <ul style="list-style-type: none"> <li>- analyses met meerdere verklarende factoren</li> <li>- terugkomen op kwaliteitsbeoordeling model: verklarende variantie; heb je eigenlijk wel alle invloedsfactoren gemeten die relevant zijn?</li> </ul>                                                               | JvA    | P-5    |
| 11.15 - 11.30 | Koffiepauze                                                                                                                                                                                                                                                                                                                                                                        |        |        |
| 11.30 - 12.15 | <b>Practicumblok: niet-lineariteit in Menyanthes</b> <ul style="list-style-type: none"> <li>- niet-lineariteit (wanneer nodig, wanneer overbodig)</li> </ul>                                                                                                                                                                                                                       | MK     | P-6    |
| 12.15 - 13.00 | Lunch                                                                                                                                                                                                                                                                                                                                                                              |        |        |
| 13.00 - 13.30 | <b>Presentatie Verschillende Cases:</b> <ul style="list-style-type: none"> <li>- aansprekend voorbeeld van modellering van een externe bron met voorbeelden van praktische problemen, oplossingen en uitkomsten</li> <li>- koppeling maken met de ecohydrologie; wat kan tijdreeksanalyse door de hydroloog betekenen voor ecohydrologen?; impuls-responsonderzoek KIWA</li> </ul> | JvA    |        |
| 13.30 - 14.00 | <b>Practicumblok: verdieping modellering, tips en tricks.</b> <ul style="list-style-type: none"> <li>- nadruk op kalibratieproblemen en diagnostiek van je model.</li> <li>- modellering van diverse reeksen met valkuilen</li> <li>- cursisten door oefening bewust maken van het</li> </ul>                                                                                      | JvA    |        |

|               |                                                           |    |     |
|---------------|-----------------------------------------------------------|----|-----|
|               | toepassingsbereik en de valkuilen van<br>tijdreeksanalyse |    |     |
| 15.15 – 15.30 | Theepauze                                                 |    |     |
| 15.30 – 16.00 | <b>Practicumblok – Do-it-yourself in Excel</b>            | MK | P-7 |
| 16.00 – 16.30 | Afsluiting cursus en evaluatie.                           | MK |     |

## Deelnemerslijst Tijdreeksanalyse voor (eco)hydrologen

Cursus blok 1, 4 en 6 October

| Blok | #  | Naam                   | Organisatie                   | Adres        |         |            |
|------|----|------------------------|-------------------------------|--------------|---------|------------|
| 1    | 1  | Francis de Graaf       | waterschap Groot Salland      | Postbus 60   | 8000 AB | Zwolle     |
| 1    | 2  | Guus van den Berg      | waterschap Groot Salland      | Postbus 60   | 8000 AB | Zwolle     |
| 1    | 3  | Tanja Bogdanova        | waterschap Groot Salland      | Postbus 60   | 8000 AB | Zwolle     |
| 1    | 4  | Tom Grobbee            | waterschap Velt en Vecht      | Postbus 330  | 7740 AH | Coevorden  |
| 1    | 5  | Pieter Filius          | waterschap Velt en Vecht      | Postbus 330  | 7740 AH | Coevorden  |
| 1    | 6  | Jacques Esenkbrink     | waterschap Reest en Wieden    | Postbus 120  | 7940 AC | Meppel     |
| 1    | 7  | Almer Bolman           | waterschap Vallei en Eem      | Postbus 330  | 3830 AJ | Leusden    |
| 1    | 8  | Matthijs van den Brink | waterschap Vallei en Eem      | Postbus 330  | 3830 AJ | Leusden    |
| 1    | 9  | Dolf van de Voort      | waterschap Brabantse Delta    | Postbus 5520 | 4801 DZ | Breda      |
| 1    | 10 | Gert van den Houten    | waterschap Rijn en IJssel     | Postbus 148  | 7000 AC | Doetinchem |
| 1    | 11 | Bert Docter            | waterschap Rijn en IJssel     | Postbus 148  | 7000 AC | Doetinchem |
| 1    | 12 | Pierre Kamps           | waterleidingbedrijf Amsterdam | Postbus 8169 | 1005 AD | Amsterdam  |
| 1    | 13 | Thomas de Meij         | waterschap Velt en Vecht      | Postbus 330  | 7740 AH | Coevorden  |
| 1    | 14 |                        |                               |              |         |            |
| 1    | 15 |                        |                               |              |         |            |

Cursus blok 2, 8 en 10 November

| Blok | #  | Naam                    | Organisatie                    | Adres         |         |            |
|------|----|-------------------------|--------------------------------|---------------|---------|------------|
| 2    | 16 | Margo Meeuwissen        | waterschap Groot Salland       | Postbus 60    | 8000 AB | Zwolle     |
| 2    | 17 | Jos Alblas              | waterschap Groot Salland       | Postbus 60    | 8000 AB | Zwolle     |
| 2    | 18 | Hilde van Mensvoort     | waterschap Groot Salland       | Postbus 60    | 8000 AB | Zwolle     |
| 2    | 19 | Bert Bijkerk            | waterschap Groot Salland       | Postbus 60    | 8000 AB | Zwolle     |
| 2    | 20 | Jan Willem van den Barg | waterschap Velt en Vecht       | Postbus 330   | 7740 AH | Coevorden  |
| 2    | 21 | Ellen Bollen-Weide      | waterschap Rijn en IJssel      | Postbus 148   | 7000 AC | Doetinchem |
| 2    | 22 | Roel van der Veen       | waterschap Rijn en IJssel      | Postbus 148   | 7000 AC | Doetinchem |
| 2    | 23 | Sjaak Vorstermans       | Staatsbosbeheer                | Postbus 330   | 5000 AH | Tilburg    |
| 2    | 24 | Eise Harkema            | Staatsbosbeheer                | Postbus 1300  | 3970 BH | Driebergen |
| 2    | 25 | Cor Beets               | Staatsbosbeheer                | Postbus 58147 | 1040 HD | Amsterdam  |
| 2    | 26 | Henk Hut                | Staatsbosbeheer                | Trompsingel 1 | 9724 CX | Groningen  |
| 2    | 27 | Philip Nienhuis         | waterleidingbedrijf Amsterdam  | Postbus 8169  | 1005 AD | Amsterdam  |
| 2    | 28 | Zwannie Visser          | waterschap Reest en Wieden     | Postbus 120   | 7940 AC | Meppel     |
| 2    | 29 | Henk van Hardevelt      | hoogheemraadschap van Rijnland | Postbus 156   | 2300 AD | Leiden     |
| 2    | 30 | Ans Elffrink            | waterschap Veluwe              | Postbus 4142  | 7320 AC | Apeldoorn  |



# 1 Inleiding

Hydrologen en ecohydrologen hebben een dynamisch vak. Zij hebben te maken met fenomenen die veranderlijk zijn in de tijd: de grondwaterstand, het weer, de waterkwaliteit, de samenstelling van de vegetatie, noem maar op. Over al deze fenomenen wordt informatie verzameld, wat niet zelden leidt tot grote databestanden met daarin lange tijdreeksen van gegevens. Voordat deze 'ruwe' gegevens gebruikt kunnen worden bij het maken van keuzes en het nemen van beslissingen in het waterbeheer, moeten zij eerst worden geanalyseerd. Voor (eco)hydrologen is de schone taak weggelegd om informatie te verzamelen en te verwerken tot informatie die benut kan worden bij het waterbeheer en -beleid. Bij deze verwerking kan een (eco)hydroloog kiezen uit een groot aantal methoden, die vaak geheel uiteenlopende uitgangspunten en achtergronden hebben. Het is vaak niet eenvoudig hieruit een keuze te maken. 'Moet ik een empirisch of een fysisch-mechanistisch model gebruiken?' en 'Moet ik iets doen met statistiek en onzekerheid of niet?' zijn vragen die een (eco)hydroloog, de verschillende methoden overziend, zich zal stellen.

Gelukkig liggen de alternatieve methoden waaruit een (eco)hydroloog moet kiezen steeds minder ver uit elkaar. De laatste jaren zijn tussen de uitersten van het methodenspectrum namelijk tussenvormen ontstaan: methoden die zowel van fysische als van statistische kennis gebruik maken. Er zijn relatief eenvoudige fysisch verklaarbare tijdreeksmodellen ontwikkeld waarmee schattingen en voorspellingen kunnen worden verricht met gekwantificeerde nauwkeurigheid. Een voorbeeld van een methodiek waarin statistische en fysische kennis is geïntegreerd is het pakket Menyanthes voor de analyse van grondwaterstandstijdreeksen.

Het doel van deze cursus en cursushandleiding is om (eco)hydrologen zodanig wegwijs te maken in de wereld van de tijdreeksmodellering, dat zij zelfstandig tijdreeksen kunnen modelleren, modellen kunnen beoordelen en resultaten kunnen interpreteren. Nu is de wereld van tijdreeksmodellering erg groot; enige afbakening is dan ook gewenst. Om te beginnen beperken wij ons tot tijdreeksen van grondwaterstanden, een enkel uitstapje misschien daargelaten. Voorts zoeken wij het niet in de puur empirisch-statistische, noch in de strikt fysisch-mechanistische benadering. Wij kiezen voor modellen die fysisch gebaseerd zijn en een statistische beschrijving van nauwkeurigheid geven. De voorbeelden en practica zijn gebaseerd op het pakket Menyanthes, dat dit soort modellen bevat.

De opbouw van deze handleiding is als volgt. In hoofdstuk 2 introduceren wij een aantal basisbegrippen uit de tijdreeksanalyse, en behandelen wij tijdreeksanalyse vanuit een statistisch perspectief. De belangrijkste empirisch-statistische tijdreeksmodellen komen hierbij aan de orde. Vervolgens beschouwen wij in hoofdstuk 3 tijdreeksmodellering van grondwaterstandsreeksen vanuit de hydrologie. In hoofdstuk 4 gaan wij in op de modellering van grondwaterstandsreeksen die door meerdere factoren



worden beïnvloed, zoals neerslag, verdamping, rivierwaterstanden, onttrekkingen en andere ingrepen van mensen in de hydrologie. Hoofdstuk 5 beschrijft het proces van tijdreeksmodellering: hoe kom je van verzamelde data tot een tijdreeksmodel en hoe beoordeel je de kwaliteit van deze modellen? In hoofdstuk 6 staan wij stil bij het monitoringproces: het proces van verzamelen van tijdreeksen tot en met het verwerken, het modelleren en de gevolgtrekking.

## 2 Tijdreeksanalyse vanuit statistisch perspectief

### 2.1 Inleiding

In dit hoofdstuk worden een aantal methoden beschreven waarmee tijdreeksen kunnen worden geanalyseerd en veel voorkomende vragen kunnen worden beantwoord. Dit hoofdstuk is sterk op het gebruik van methoden in de praktijk gericht. Meer theoretische achtergronden kunnen worden gevonden in bijv. (Box & Jenkins, 1970; Chatfield, 1989; Hipel & McLeod, 1994). Wij beperken ons hier tot methoden voor toepassing in het 'tijddomein'. Methoden voor toepassing in het frequentiedomein, zoals Fourieranalyse, blijven buiten beschouwing (zie Priestley, 1981). De opbouw van het hoofdstuk is als volgt. Paragraaf 2.2 geeft een introductie in toepassingen van tijdreeksanalyse, met voorbeelden van tijdreeksen die in onze leefomgeving worden verzameld. In paragraaf 2.3 worden enkele basisbegrippen uit de theorie van de tijdreeksanalyse verklaard. Vervolgens beginnen we in 2.4 met lineaire regressie, en in 2.5 komt het geheugen van systemen aan de orde. Daarna gaan we in paragraaf 2.6 in op lineaire, stationaire processen en in paragraaf 2.7 op niet-stationaire, lineaire processen. In paragraaf 2.8 komen als laatste relaties tussen reeksen aan de orde.

### 2.2 Over tijdreeksen, tijdreeksanalyse en haar toepassingen

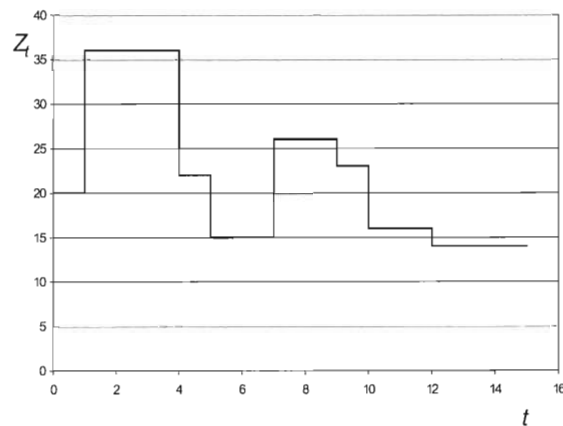
In de natuur varieert er van alles in de tijd: de grondwaterstand op een locatie beweegt zich voortdurend in op- of neerwaartse richting, rivieren en beken voeren dan weer veel, dan weer weinig water af, enzovoort. Om uiteenlopende redenen worden er waarnemingen aan deze processen op achtereenvolgende tijdstippen verzameld en opgeslagen in zogenaamde *tijdreeksen*: waarnemingen in chronologische volgorde. Tijdreeksanalyse omvat een aantal methoden om uit tijdreeksen inzicht te verkrijgen in processen en allerlei vragen te beantwoorden. Het dynamische 'gedrag', de temporele variatie, kan worden beschreven met tijdreeksmodellen, die kunnen worden gebruikt om doelparameters te schatten. Dit zijn bijvoorbeeld de verwachte waarden op een zeker tijdstip, bijvoorbeeld de verwachte grondwaterstand aan het begin van het groeiseizoen, of de waarschijnlijkheid waarmee kritische niveaus worden overschreden op een kritisch tijdstip of gedurende een kritische periode. Een aantal recente toepassingen van tijdreeksmodellering in het landelijk gebied zijn:

- **Toepassing:** Het schatten van grondwaterstandskarakteristieken, zoals gemiddeld hoogste en laagste grondwaterstand (GHG en GLG), duurlijnen en regimecurves;  
**Methoden:** Transfer-ruismodellering (TFN), stochastische simulatie (Bierkens & Bron, 2000; Kotters, 2001);
- **Toepassing:** Verdrogingsstudies;

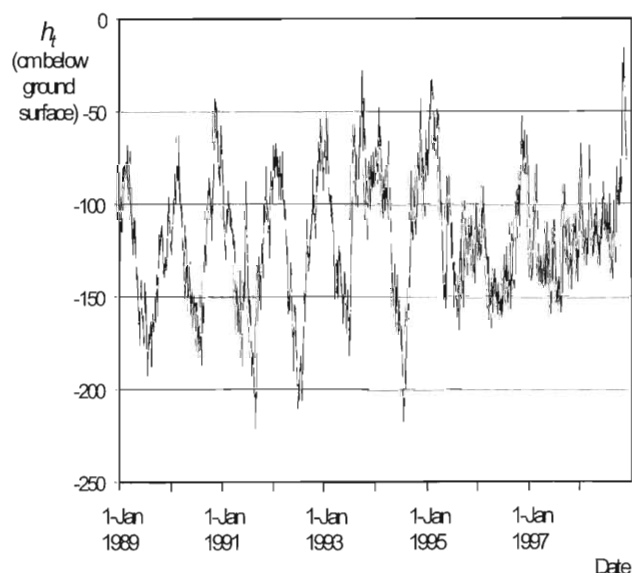
- Methoden:** TFN-modellering, trendanalyse (o.a. (Rolf, 1989; Maas & Von Asmuth, 2003);
- **Toepassing:** Afvoervoorspellingen;  
*Methoden:* deseasonalization, interventieanalyse, TFN-modellering;
  - **Toepassing:** Aantonen van effecten van een ingreep op de grondwaterstanden in de omgeving;  
*Methoden:* Kalman-filtering, TFN-modellering, interventieanalyse (o.a. (Gehrels, 1999; Cirkel & Von Asmuth, 2004);
  - **Toepassing:** Voorspelling effect hydrologische ingrepen op grondwaterstandsfluctuatie  
*Methoden:* TFN-modellering, fysische interpretatie, stochastische simulatie (Knotters, 2001);
  - **Toepassing:** Analyse van de geohydrologische eigenschappen van de ondergrond (Maas & Von Asmuth, 2004).
  - **Toepassing:** Analyse van pompproef resultaten (Maas, 2005).
  - **Toepassing:** Effect van een vernattingsmaatregel op grondwaterstanden en/of nitraatconcentraties;
  - **Toepassing:** Analyse van waarnemingen aan het gedrag van dieren;
  - **Toepassing:** Analyse van gemiddelde jaarlijkse zeestanden (Van Geer *et al.*, 1999);

### 2.3 Enkele basisbegrippen

In deze paragraaf zullen een aantal basisbegrippen uit de tijdreeksanalyse worden toegelicht, die van belang zijn voor een goede toepassing van de methoden die elders in dit hoofdstuk worden besproken. In paragraaf 2.3.1 gaan wij in op tijdreeksen, processen en tijdreeksmodellen in het algemeen. In paragraaf 2.3.2 bespreken wij een aantal statistische begrippen die specifiek betrekking hebben op de tijdreeksmodellen die in deze syllabus worden behandeld.



figuur 1: Continue tijdreeks, discrete variabele

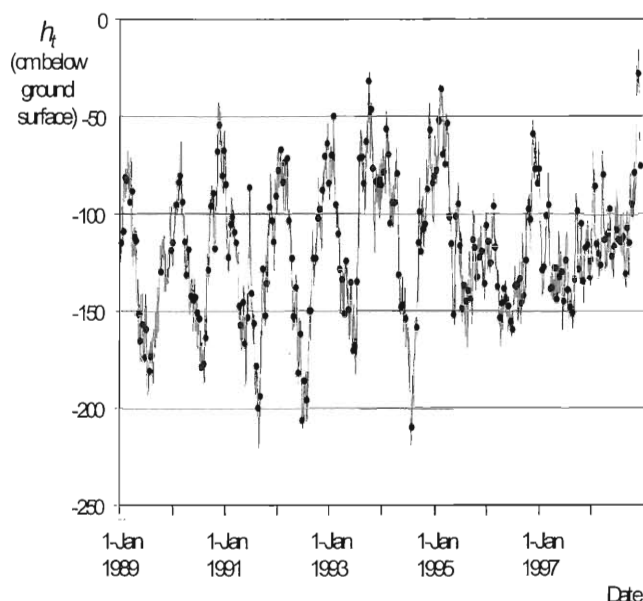


figuur 2: Continue tijdreeks, continue variabele

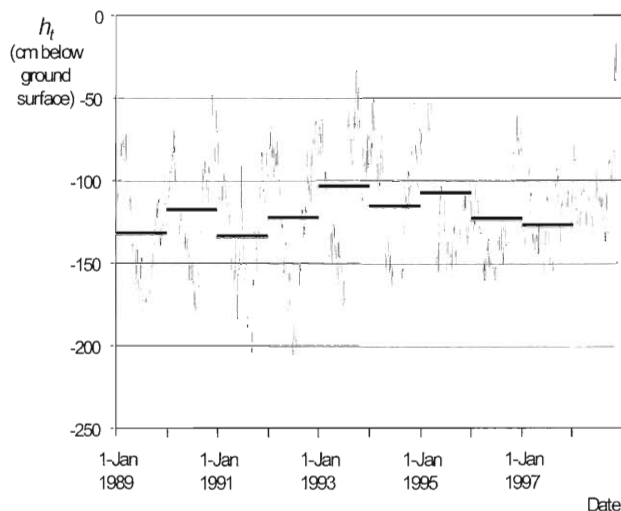
### 2.3.1 Reeksen, processen en modellen, continu en discreet

Een *tijdreeks* is een set waarnemingen in chronologische volgorde. Bij een *continue tijdreeks* is er op ieder moment in de tijd een waarneming. De variabele zelf kan dan overigens discreet of continue zijn (zie resp. figuur 1 en 2).

Bij een *discrete tijdreeks* zijn de waarnemingen verzameld op discrete tijdstippen, gescheiden door intervallen. De lengte van deze intervallen kan ongelijkmatig zijn (*unevenly spaced time series*) of gelijkmatig of equidistant (*equally spaced discrete time series*): uur-, dag-, maand-, jaar-, enz.-waarnemingen. Dit kunnen uur-, dag-, maand-, enz.-gemiddelden of -sommen zijn, maar ook waarnemingen aan een continu proces. Let bij berekeningen op schrikkeljaren en het variërende aantal dagen in een maand! Figuur 3 en 4 geven voorbeelden van een discrete tijdreeks.



figuur 3: Continue variabele, waargenomen op discrete tijdstippen (halfmaandelijks waargenomen grondwaterstanden).



figuur 4: Continue variabele, discrete tijdreeks: jaargemiddelde grondwaterstanden.

Vaak verdient een maand- of jaargemiddelde de voorkeur boven een maand- of jaarsom, om te voorkomen dat een deel van de variatie verklaard wordt uit de lengte van de maanden of jaren. Het voordeel van equidistante tijdreeksen is dat met relatief eenvoudige mathematische methoden analyses kunnen worden uitgevoerd. De methoden die in dit hoofdstuk worden behandeld zijn hier voorbeelden van. Recent zijn methoden ontwikkeld om onregelmatig waargenomen tijdreeksen van de grondwaterstand te modelleren. Deze methoden zijn gebaseerd op dagsommen van het potentiële neerslagoverschot (neerslag minus potentiële verdamping), waarmee de grondwaterstand samenhangt (Bierkens et al., 1999b; Von Asmuth et al., 2001; Von Asmuth et al., 2002). Dagsommen van de neerslag en de verdamping worden door het Koninklijk Nederlands Meteorologisch Instituut (KNMI) verstrekt voor een groot aantal locaties in Nederland.

Het woord *tijd*-reeksanalyse impliceert niet dat de methoden alleen toepasbaar zijn op waarnemingen op achtereenvolgende tijdstippen; het interval mag ook ruimtelijk zijn. Voorbeelden zijn waarnemingen op regelmatige afstand in de loop van een meanderende rivier of beek en waarnemingen langs routes of transectwaarnemingen. Als men uitsluitend is geïnteresseerd in de positie langs de rivier, beek of route dan kunnen methoden uit de tijdreeksanalyse met succes worden toegepast. Bij dit soort reeksen kan ook van causaliteit sprake zijn, evenals bij tijdreeksen. Een *proces* is iets dat zich in de tijd ontwikkelt, afspeelt, voortplant, gedraagt. Een *stochastisch proces* kunnen wij niet precies beschrijven, wij zijn tenminste over een deel van het proces onzeker. De toekomstige toestand kunnen we niet exact voorspellen en de historie kunnen wij niet exact reconstrueren. Een voorbeeld is de buitentemperatuur. Weermannen of -vrouwen voorspellen aan het eind van het journaal hoe warm het morgen zal zijn. Zij zullen nooit zeggen: 'morgen wordt het 23 graden', maar eerder: 'morgen kan de temperatuur oplopen tot zo'n 23 graden'. Sinds enige tijd presenteren zij grafieken waarbij naast een voorspelling ook een band wordt aangegeven van de 'onzekerheid'. Opmerkelijk is dat zij dan zeggen 'Zoals u ziet is de onzekerheid groot', en niet 'Wij zijn erg onzeker', terwijl ze dat natuurlijk wel zijn.

Een *deterministisch proces* kunnen wij, in tegenstelling tot een stochastisch proces, precies beschrijven. De toekomstige toestand kunnen wij exact uitrekenen en de historie kunnen wij exact reconstrueren. Een model dat een deterministisch proces beschrijft heeft één uitkomst. Een model dat een *stochastisch proces* beschrijft heeft oneindig veel mogelijke uitkomsten of *realisaties*. De zojuist genoemde band rond de temperatuurvoorspellingen in het journaal geeft het traject aan waarbinnen deze realisaties liggen. Je kunt deze realisaties zien als de uitkomsten van een kansexperiment, zoals een loterij. Elke realisatie is een mogelijkheid van het verloop van bijvoorbeeld temperatuur in de tijd.

De tijdreeksmodellen die in deze cursus aan de orde komen beschrijven stochastische processen. Wij beperken ons in dit hoofdstuk tot de lineaire modellen voor equidistante, discrete tijdreeksen, zoals die bijvoorbeeld worden beschreven door (Box & Jenkins, 1970; Hipel & McLeod, 1994).

(Niet)lineariteit en niet-lineaire tijdreeksanalyse, en wat dat precies inhoudt, komt ter sprake in paragraaf 3.5. Wij benadrukken dat bij tijdreeksmodellering de waargenomen reeks zelf wordt beschouwd als een realisatie van een stochastisch proces. Dit proces, dat afhangt van het waarnemingsinterval, moet niet worden verward met het onderliggende fysische proces dat naast meetfouten de bron is van variatie in de waargenomen reeks. De veronderstelling van een stochastisch proces impliceert niet dat de natuur stochastisch is. In principe zijn stochastische processen gebaseerd op data, en bestaan ze alleen in onze verbeelding om statistische verwerking en gevolgtrekking mogelijk te maken. Een belangrijke veronderstelling die vaak bij tijdreeksmodellering wordt gedaan om een mathematisch eenvoudig model te bereiken, is die van *stationariteit*. Een proces is stationair als de statistische eigenschappen niet veranderen in de tijd. Je kunt je afvragen of de processen in het landelijk gebied van nature stationair zijn. Als je bijvoorbeeld een reeks halfmaandelijke neerslag-overschotten (figuur 3) alsmaar verlengt kom je op een gegeven moment in een ander klimaat terecht. Als je een reeks steeds maar inkort houd je op den duur een stukje 'helling' over tussen bijvoorbeeld de winter en de zomer. Ook de waarnemingsfrequentie hangt samen met de veronderstelling van stationariteit: in de halfmaandelijke cijfers zit een seizoeneffect dat in jaarcijfers niet zichtbaar is. Samenvattend kunnen bij stationariteit de volgende opmerkingen worden gemaakt:

1. Stationariteit is geen eigenschap van de natuur zelf, maar van het stochastische proces dat door de onderzoeker is bedacht om de natuur met behulp van een beperkt aantal waarnemingen te kunnen beschrijven;
2. Stationariteit veronderstel je om mathematische eenvoud te bereiken;
3. Gegeven een bepaalde lengte van de tijdreeks en een bepaalde waarnemingsfrequentie kun je iets veronderstellen over stationariteit.

*Strong/strict stationarity* wil zeggen dat alle statistische eigenschappen onafhankelijk zijn van de tijd, dus niet veranderen bij verschuiving in de tijd. Vaak is het voldoende om een afgezwakte veronderstelling te doen: *weak stationarity of order k* wil zeggen dat de 'statistische momenten tot orde  $k$ ' alleen afhangen van de tijdverschillen en niet van tijdstippen. *Second order stationarity* wil zeggen dat het stochastische proces kan worden beschreven met het gemiddelde, de variantie en de autocorrelatiefunctie. Deze begrippen worden in paragraaf 2.3.2 uitgelegd. *Second order stationarity* wordt ook wel *covariance stationarity* genoemd.

### 2.3.2 Statistische begrippen

We zullen nu ingaan op tweede-orde stationaire stochastische processen. Stel dat we een equidistante tijdreeks hebben van  $n$  waarnemingen,  $z_1, z_2, z_3, \dots, z_n$ . We kunnen het proces niet exact beschrijven, dus  $\{z_i\}$  zou in onze ogen een realisatie van een stochastisch proces  $\{Z_i\}$  moeten zijn (deterministische en stochastische grootheden onderscheiden we met resp.

een kleine en een hoofdletter). Het gemiddelde is gedefinieerd als de verwachtingswaarde van  $Z_t$ :

$$\mu = E[Z_t] \quad (1)$$

en kan uit een waargenomen tijdreeks geschat worden met het rekenkundige gemiddelde  $\bar{z}$ :

$$\hat{\mu} = \bar{z} = \frac{1}{n} \sum_{t=1}^n z_t \quad (2)$$

Met een dakje geven we aan dat  $\mu$  wordt geschat. ( $\hat{\mu}$  nadert tot  $\mu$  als  $n$  tot oneindig nadert).

Rond het gemiddelde  $\mu$  fluctueert het stochastische proces met een bepaalde spreiding of variatie. Een maat voor deze spreiding is de variantie. De variantie van het stochastische proces  $\{Z_t\}$  is gedefinieerd als de verwachtingswaarde van de kwadratische afwijkingen t.o.v. het gemiddelde:

$$\sigma_z^2 = E[(Z_t - \mu)^2] \quad (3)$$

$\sigma_z^2$  kan worden geschat met:

$$\hat{\sigma}_z^2 = \frac{1}{n-1} \sum_{t=1}^n (z_t - \bar{z})^2 \quad (4)$$

De waarde van  $z$  op een bepaald tijdstip  $t$  kan samenhangen van de waarde van  $z$  op een ander tijdstip,  $t+k$ . De mate waarin sprake is van samenhang tussen waarden die met een zogenaamde *lag*  $k$  van elkaar zijn gescheiden, kan worden beschreven met de autocovariantie. 'Auto-' wil zeggen dat wij naar de samenhang kijken binnen één proces. De autocovariantie bij een lag  $k$  is gedefinieerd als:

$$C_k = E[(Z_t - \mu)(Z_{t+k} - \mu)] \quad (5)$$

$$C_0 = \sigma_z^2.$$

Bij lag 0 is de autocovariantie dus gelijk aan de variantie. De autocorrelatiefunctie (ACF) bij lag  $k$  is een geschaalde vorm van de autocovariantie:

$$\rho_k = \frac{C_k}{C_0} \quad (6)$$

De *sample* of geschatte autocovariantiefunctie bij lag  $k$  kan uit een reeks worden berekend met:

$$\hat{C}_k = \frac{1}{n_k} \sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z}) \quad (7)$$

waarin  $n_k$  het aantal gesommeerde termen is, met een maximum van  $n - k$ ; termen waar een waarde van  $z_t$  of  $z_{t+k}$  ontbreekt zijn niet meegeteld.

Naarmate de waarden van  $z_t$  minder verschillen van die van  $z_{t+k}$  zal de autocovariantiefunctie een hogere waarde krijgen.

De sample ACF bij lag  $k$  wordt geschat met:

$$\hat{\rho}_k = \left(1 - \frac{k}{n}\right) \frac{\hat{C}_k}{\hat{C}_0} \quad (8)$$

## 2.4 Eenvoudige, lineaire regressie

Tijdreeksmodellen kunnen in principe worden opgevat als regressiemodellen. Een eenvoudig lineair tijdreeksmodel heeft dan de vorm:

$$Z_t = b_0 + b_1 X_t + \varepsilon_t \quad (9)$$

Hierin kan  $X_t$  bijvoorbeeld gelijk zijn aan  $Z$  op het vorige tijdstip,  $Z_{t-1}$ :

$$Z_t = b_0 + b_1 Z_{t-1} + \varepsilon_t \quad (10)$$

Een dergelijk model noemen we ook wel een dynamisch regressiemodel. Het model in (10) is een univariaat model, omdat er geen andere verklarende variabelen dan  $Z$  in voorkomen. Het kan zijn dat men de temporele variatie van de ene variabele wil verklaren uit dat van een andere variabele. Een voorbeeld is het verklaren van de fluctuatie van de grondwaterstand uit de fluctuatie van het neerslagoverschot. Een eenvoudig model zou dan vergelijking (9) kunnen zijn, waarin  $Z_t$  de grondwaterstand is en  $X_t$  het neerslagoverschot tussen bijvoorbeeld tijdstip  $t-1$  en  $t$ . Een belangrijke veronderstelling bij regressiemodellering is echter dat de residuen  $\varepsilon$  onderling ongecorreleerd zijn en niet samenhangen met het niveau van de responsvariabele  $Z$ . Met een eenvoudig lineair model als in vergelijking (9) is vaak niet aan deze veronderstellingen te voldoen; de 'ruis'  $\varepsilon$  zal vaak gekleurd zijn, d.w.z. een correlatiestructuur vertonen. Om dit te vermijden moeten termen aan het regressiemodel worden toegevoegd die terugslaan op vorige tijdstippen, zoals bijvoorbeeld  $b_1 Z_{t-1}$  in vergelijking (10). Bij tijdreeksmodellering zoeken we naar modellen waarbij de ruisterm  $\varepsilon$  wit is, dat wil zeggen ongecorreleerd en gelijk verdeeld (overal evenveel variatie). Een dynamisch regressiemodel voor de relatie tussen neerslagoverschot en grondwaterstand de bij een halfmaandelijke frequentie is waargenomen zou de volgende structuur kunnen hebben:

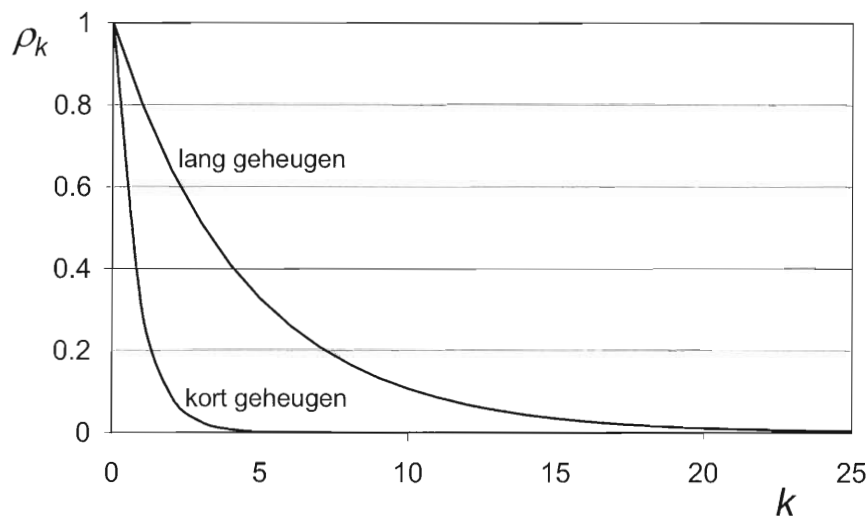


$$Z_t = b_0 + b_1 Z_{t-1} + b_2 X_t + \varepsilon_t \quad (11)$$

Dit dynamische regressiemodel is gelijk aan het ARX(1,0)-model (ARX= autoregressief met een exogene variabele).

## 2.5 Over geheugen en tijdreeksen

'Geheugen' is een woord dat vaak valt bij tijdreeksanalyse. Het kan als volgt worden uitgelegd: het geheugen is de afstand in de tijd waarover de toestand van een variabele nog afhangt van eerdere gebeurtenissen. Een lang geheugen komt tot uiting in hoge waarden voor de coëfficiënt  $b_1$  in vergelijking (10) en (11). Een autocorrelatiefunctie beschrijft de samenhang met de waarden op vorige tijdstippen, en geeft dus inzicht in het geheugen. Hoe sneller de ACF naar 0 gaat, hoe korter het geheugen. Figuur 5 illustreert dit.



figuur 5 ACF voor een tijdreeks met een lang en een kort geheugen

## 2.6 Lineaire, stationaire processen

### 2.6.1 Autoregressieve processen (AR)

Bij bepaalde processen is het denkbaar dat de toestand op een bepaald tijdstip gecorreleerd is met de toestand op het vorige tijdstip. Dit soort reeksen kunnen we beschrijven met autoregressieve (AR-)processen. Een autoregressief proces van orde 1, een AR(1)-proces of Markovproces is gegeven door:

$$Z_t - \mu = \phi_1 (Z_{t-1} - \mu) + \varepsilon_t \quad (12)$$

met  $\mu$  het gemiddelde niveau,  $\phi_1$  de AR-parameter,  $\varepsilon_t$  de ruisterm met gemiddelde 0 en variantie  $\sigma_\varepsilon^2$ .  $\varepsilon_t$  is verondersteld gelijk en onafhankelijk verdeeld te zijn (IID), dus:

$$E[\varepsilon_t \varepsilon_{t-k}] = \begin{cases} \sigma_\varepsilon^2 & \text{als } k = 0 \\ 0 & \text{als } k \neq 0 \end{cases} \quad (13)$$

voor alle  $t$ . Een autoregressief proces van orde  $p$ ,  $AR(p)$  is gegeven door:

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \phi_2(Z_{t-2} - \mu) + \dots + \phi_p(Z_{t-p} - \mu) + \varepsilon_t \quad (14)$$

De AR-parameters moeten in een bepaald bereik liggen om aan de eis van stationariteit te voldoen. Voor een  $AR(1)$ -proces geldt:  $|\phi_1| < 1$ . Dit is als volgt na te gaan. Het  $AR(1)$ -proces in vergelijking (12) kan worden geschreven als:

$$Z_t - \mu = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \dots \quad (15)$$

Ga nu na wat het effect is van de waarde van  $\phi_1$  op de samenhang tussen  $Z_t - \mu$  en  $\varepsilon_{t-k}$  met  $k = 1, 2, 3, \dots$ . Een belangrijk hulpmiddel bij het identificeren van een  $AR(p)$ -proces is de autocorrelatiefunctie (ACF). We zullen nu de theoretische ACF van een  $AR(p)$ -proces afleiden. Eerst vermenigvuldigen we de termen van het  $AR(p)$ -proces in vergelijking (14) met  $(Z_{t-k} - \mu)$ , zodat we krijgen:

$$\begin{aligned} (Z_{t-k} - \mu)(Z_t - \mu) = & \phi_1(Z_{t-k} - \mu)(Z_{t-1} - \mu) + \phi_2(Z_{t-k} - \mu)(Z_{t-2} - \mu) + \dots \\ & + \phi_p(Z_{t-k} - \mu)(Z_{t-p} - \mu) + (Z_{t-k} - \mu)\varepsilon_t \end{aligned} \quad (16)$$

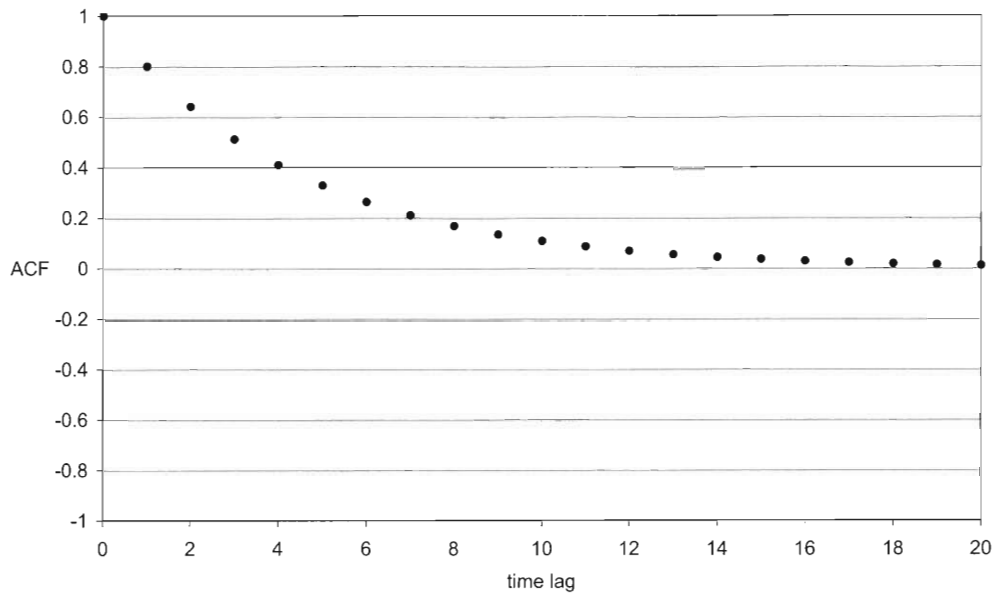
Nu nemen we de verwachtingen van vergelijking (16). We krijgen dan:

$$C_k = \phi_1 C_{k-1} + \phi_2 C_{k-2} + \dots + \phi_p C_{k-p} \quad (17)$$

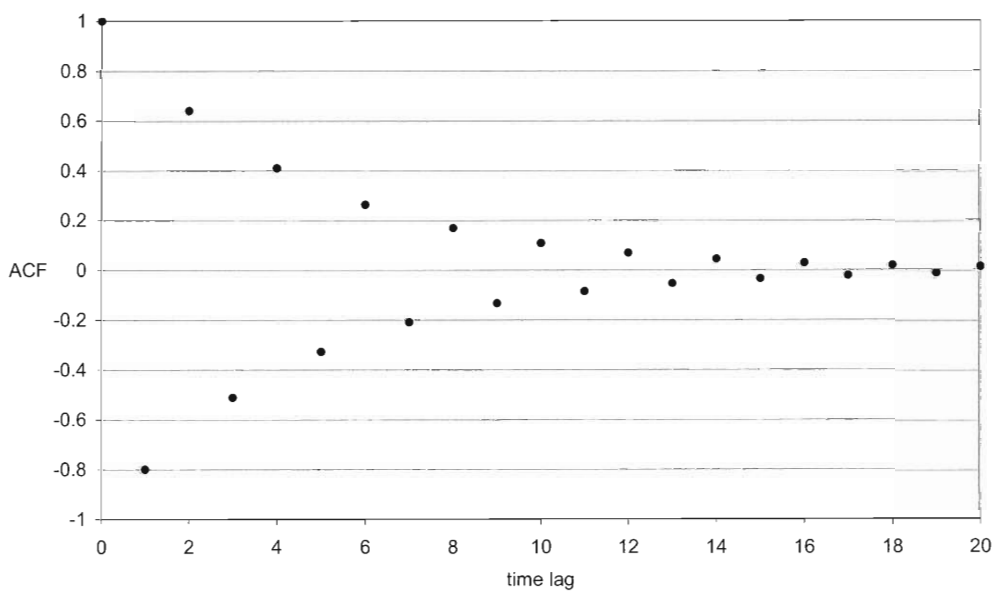
met  $k > 0$ .  $E[(Z_{t-k} - \mu)\varepsilon_t]$  is gelijk aan nul voor  $k > 0$ , omdat  $Z_{t-k}$  alleen maar afhangt van het foutenproces tot en met  $t - k$  en  $\varepsilon_t$  hiermee ongecorrleerd is. De theoretische ACF kan nu worden verkregen door vergelijking (17) te delen door  $C(0)$ :

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (18)$$

met  $k > 0$ . In figuur 6 en 7 staan de theoretische ACF's weergegeven voor  $AR(1)$ -processen met resp.  $\phi_1 = 0.8$  en  $\phi_1 = -0.8$  (zie vergelijking (1)).



figuur 6 Theoretische ACF voor een AR(1)-proces met  $\phi_1 = 0.8$



figuur 7 Theoretische ACF voor een AR(1)-proces met  $\phi_1 = -0.8$

### 2.6.2 Moving-Averageprocessen (MA)

Bij een Moving-Average- of MA-proces hangt de toestand op een bepaald tijdstip af van een 'random shock' op dat tijdstip en een random shock die op een of meer vorige tijdstippen optrad. Een eerste orde Moving-Averageproces, MA(1), is gegeven door:

$$Z_t - \mu = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (19)$$

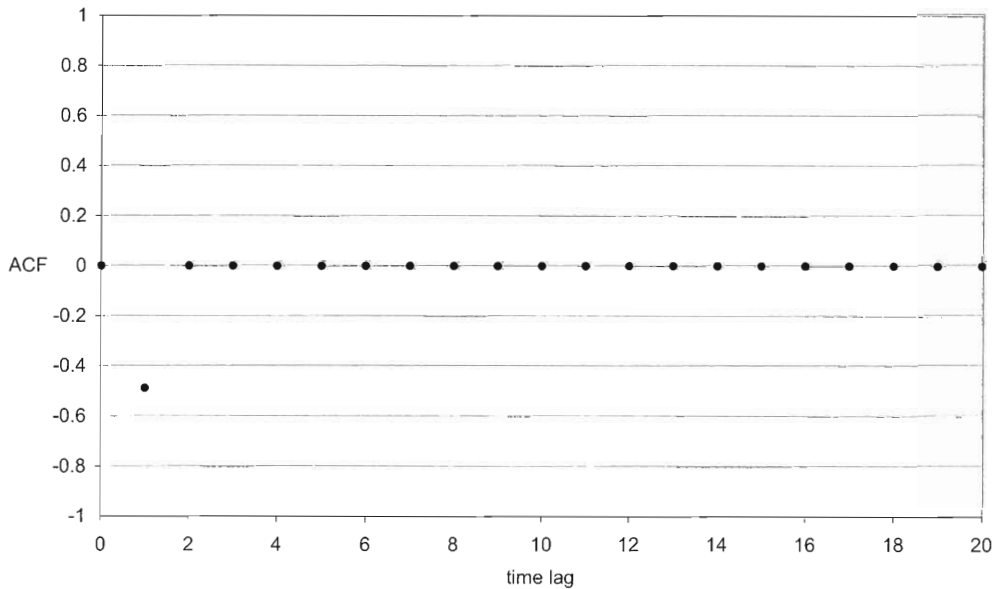
Hierin zijn  $\varepsilon_t$  en  $\varepsilon_{t-1}$  random shocks, deel uitmakend van een witte-ruisproces met gemiddelde 0 en eindige en constante variantie. Een MA( $q$ )-proces is gegeven door:

$$Z_t - \mu = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (20)$$

Omdat  $\{\varepsilon_t\}$  een stationair proces is en  $Z_t$  een lineaire combinatie van de  $\varepsilon_t$ 's is gelden er geen stationariteitseisen voor de MA-parameters. Het is echter verstandig om bepaalde restricties aan de waarden van de MA-parameters te verbinden. Om achter het nut hiervan te komen schrijven we het MA(1)-model in vergelijking (16) als:

$$Z_t - \mu = \varepsilon_t - \theta_1(Z_{t-1} - \mu) - \theta_1^2(Z_{t-2} - \mu) - \theta_1^3(Z_{t-3} - \mu) - \dots \quad (21)$$

Nu gaan we na wat het effect van de waarde van  $\theta_1$  is op de samenhang tussen  $Z_t$  en  $Z_{t-k}$ ,  $k = 1, 2, 3, \dots$ . Het blijkt dat  $|\theta_1| < 1$  moet zijn, willen gebeurtenissen uit het verre verleden niet meer invloed hebben dan recente gebeurtenissen. We noemen dit de voorwaarde van invertibiliteit. De theoretische ACF voor een MA(1)-proces met  $\theta_1 = 0.8$  is weergegeven in figuur 8.



figuur 8 Theoretische ACF voor een MA(1)-proces met  $\theta_1 = 0.8$

### 2.6.3 Autoregressive-Moving Average processen (ARMA)

Een tijdreeks kan zowel kenmerken van een autoregressief proces als van een moving-averageproces bevatten. Een autoregressive-moving average of ARMA( $p, q$ )-proces is gegeven door:

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \phi_2(Z_{t-2} - \mu) + \dots + \phi_p(Z_{t-p} - \mu) + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (22)$$

## 2.7 Niet-stationaire lineaire processen

### 2.7.1 Verschilwaarden

Door verschilwaarden te berekenen kan een bepaalde trend uit de reeks worden gefilterd (differentiëren):

$$\begin{aligned}\nabla Z_t &= (Z_t - \mu) - (Z_{t-1} - \mu) \\ \nabla^2 Z_t &= \nabla Z_t - \nabla Z_{t-1}\end{aligned}\tag{23}$$

enzovoort, net zolang tot er een reeks van verschilwaarden ontstaat waarvan het niveau niet meer van de tijd afhangt.

### 2.7.2 Autoregressive Integrated Moving Average processen (ARIMA)

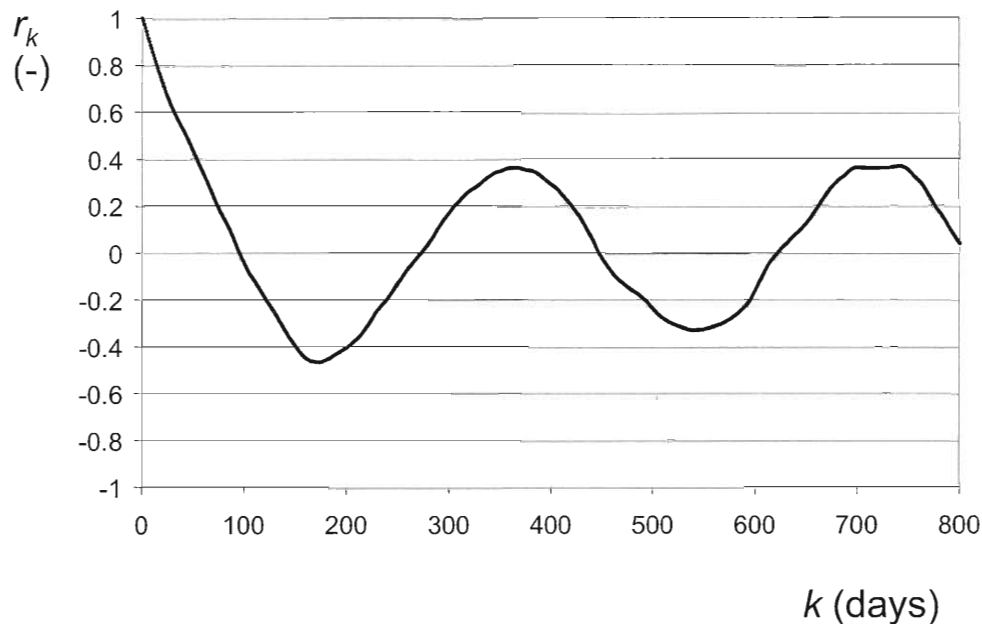
Dit is in feite een ARMA model voor stationaire verschilwaarden:

$$\begin{aligned}\nabla^d Z_t &= \phi_1 \nabla^d Z_{t-1} + \phi_2 \nabla^d Z_{t-2} + \dots + \phi_p \nabla^d Z_{t-p} \\ &\quad + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}\end{aligned}\tag{24}$$

Dit noemen we een ARIMA( $p,d,q$ )-model.

### 2.7.3 Seizoen-nietstationaire processen

Een vorm van nietstationariteit die we in de natuur vaak tegenkomen is seizoenaliteit. Figuur 8 laat de geschatte ACF zien voor een tijdreeks van dagelijks waargenomen grondwaterstanden. De geschatte ACF geeft duidelijk het jaarlijkse seizoenspatroon weer. Naast seizoensvariatie in het gemiddelde, kan de variantie zelf variëren met de seizoenen. Ondiepe grondwaterspiegels in het natte seizoen kunnen bijvoorbeeld meer variëren dan diepe grondwaterspiegels in het droge seizoen, als gevolg van de afgenomen bergingscapaciteit in de onverzadigde zone gedurende het natte seizoen. Als de variantie niet constant is in de tijd, i.e. heteroscedasticiteit, dan moet de variantie constant worden gemaakt met een deseizoenalisatieprocedure of met een Box-Coxtransformatie van de tijdreeks (Hipel en McLeod, 1994).



figuur 9 Sample ACF van een dagelijks waargenomen reeks grondwaterstanden

#### 2.7.4 Seasonal Autoregressive Integrated Moving Average processen (SARIMA)

Het ARIMA-model in vergelijking (24) is uit te breiden tot een SARIMA- $(p,d,q) \times (P,D,Q)$ -model, waarbij  $(P,D,Q)$  de seizoenscomponent beschrijft. Bij het seasonal autoregressive moving average proces zijn de verschilwaarden voor de zogenaamde seizoensafstand berekend, teneinde een seizoentrend weg te filteren. Bij maandcijfers is de seizoensafstand bijvoorbeeld 12. SARIMA-modellen zijn vrij snel erg ingewikkeld. Daarom is het raadzaam om met een eenvoudige voorbeeldreeks te controleren of het model beschrijft wat je bedoelt.

## 2.8 Samenhang tussen reeksen

### 2.8.1 Inleiding

Tot nu toe hebben we naar het dynamische gedrag van één variabele gekeken. Het is echter denkbaar dat twee of meer processen met elkaar samenhangen. Zo hangt bijvoorbeeld de grondwaterstand op tijdstip  $t$  niet alleen af van de grondwaterstanden op voorgaande tijdstippen, maar ook van het neerslagoverschot op voorgaande tijdstippen. Is er een neerslagoverschot opgebouwd dan stijgt de grondwaterstand, is er een neerslagtekort ontstaan dan daalt de grondwaterstand. Duidelijk is dat het omgekeerde niet geldt: het neerslagoverschot reageert niet op veranderingen in de grondwaterstand. De vorm van samenhang zoals tussen neerslagoverschot en grondwaterstand wordt in de tijdreeksliteratuur wel eens causaliteit genoemd. In elk geval is het heel plausibel dat er een samenhang is tussen neerslagoverschot en grondwaterstand.

### 2.8.2 Cross-covariantie, cross-correlatie

Stel dat er twee tijdreeksen zijn waargenomen op  $N$  equidistante tijdstippen,  $x_t, t = 1 \dots N$  en  $z_t, t = 1 \dots N$ . We willen weten of er samenhang is tussen de stochastische processen  $\{X_t\}$  en  $\{Z_t\}$ . Een methode is regressie-analyse, waarbij  $x_t, t = 1 \dots N$  wordt uitgezet tegen  $z_t, t = 1 \dots N$  of andersom. Het nadeel is dat er geen inzicht ontstaat in het verband tussen bijvoorbeeld  $x_t$  en  $z_{t+k}$ . De cross-covariantiefunctie en de cross-correlatiefunctie geven dit verband wel. De cross-covariantiefunctie kan worden geschat met:

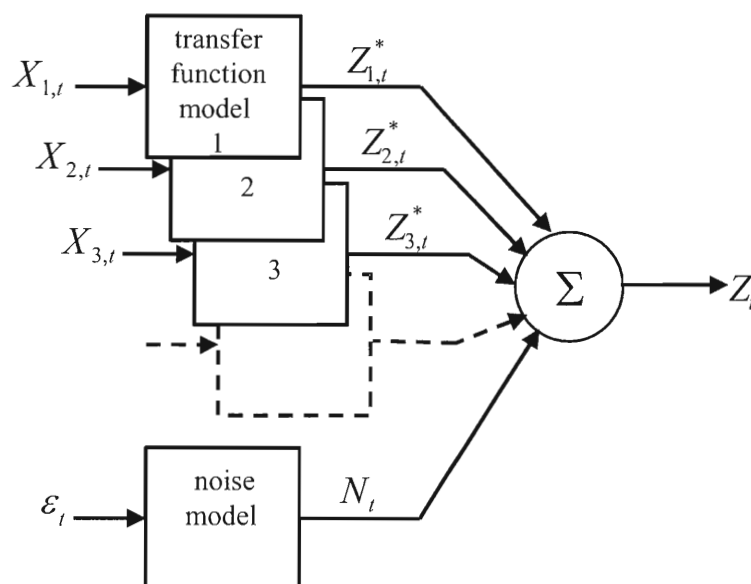
$$\hat{C}_{xz}(k) = \frac{1}{n_k} \sum_{t=1}^{n-k} (x_t - \bar{x})(z_{t+k} - \bar{z}) \quad (25)$$

voor positieve lags  $k$ . Hierin is  $n_k$  het aantal termen waarover is gesommeerd; termen die ontbrekende waarnemingen bevatten vallen af. De cross-correlatiefunctie wordt in combinatie met vergelijking (25) geschat met:

$$\hat{\rho}_{xz}(k) = \left(1 - \frac{k}{n}\right) \frac{\hat{C}_{xz}(k)}{\sqrt{\hat{C}_x(0)\hat{C}_z(0)}} \quad (26)$$

### 2.8.3 Transfer-ruismodellen

Een klasse tijdreeksmodellen voor de lineaire dynamische samenhang tussen één of meer invoerreeksen en een uitvoerreeks is het transferfunctiemodel met additionele ruis (TFN) van Box en Jenkins (1976). Voor toepassingen op reeksen uit het natuurlijke milieu verwijzen wij naar Hipel en McLeod (1994). Figuur 10 geeft de algemene vorm van het TFN-model schematisch weer.



figuur 10 Schematische weergave van de structuur van een TFN-model

Als we één invoerreeks  $\{X_t\}$  beschouwen, dan is het TFN-model gedefinieerd als:

$$Z_t = Z_t^* + N_t \quad (27)$$

waarin:

$$Z_t^* = \sum_{i=1}^r \delta_i Z_{t-1}^* + \omega_0 X_{t-b} - \sum_{j=1}^s \omega_j X_{t-j-b} \quad (28)$$

de transfercomponent is, en:

$$N_t - \mu = \sum_{i=1}^p \phi_i (N_{t-1} - \mu) + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (29)$$

de ruiscomponent is. Het subscript  $b$  is een vertraging, die gelijk is aan het aantal tijdstappen waarna een reactie op een verandering in de invoer wordt waargenomen in de uitvoer. De uitbreiding naar meer invoerreeksen is rechttoe, rechtaan. De transfercomponent in vergelijking (28) kan worden geschreven als:

$$Z_t^* = \nu_0 X_t + \nu_1 X_{t-1} + \nu_2 X_{t-2} + \dots \quad (30)$$

De gewichten  $\nu_0, \nu_1, \nu_2, \dots$  vormen samen de impuls-responsfunctie  $\nu(k)$ . De theoretische impuls-responsfunctie geeft dezelfde autoregressieve en moving-averagekarakteristieken weer als de theoretische crosscorrelatiefunctie, zie paragraaf 2.3.

Box en Jenkins (1976) geven een procedure voor het identificeren van de orde van TFN-modellen. Deze procedure is samengevat in de volgende stappen:

1. Een passend univariaat tijdreeksmodel wordt gekalibreerd op de invoerreeks  $\{x_t\}$ . De witte-ruisreeks van residuen die hierna resten wordt de *prewhitened* invoerreeks  $\{\alpha_t\}$  genoemd;
2. De uitvoerreeks  $\{z_t\}$  wordt gefilterd met het univariate tijdreeksmodel voor de invoerreeks uit de vorige stap. Dit levert een reeks  $\{\beta_t\}$  op;
3. De residuele cross-correlatiefunctie (residuele CCF)  $\hat{\rho}_{\alpha\beta}(k)$  wordt berekend voor de reeksen  $\{\alpha_t\}$  en  $\{\beta_t\}$ :

$$\hat{\rho}_{\alpha\beta}(k) = \left(1 - \frac{k}{n}\right) \frac{\hat{C}_{\alpha\beta}(k)}{\sqrt{\hat{C}_{\alpha}(0)\hat{C}_{\beta}(0)}} \quad (31)$$



4. Op basis van de residuele CCF die is gegeven in vergelijking (31) worden de parameters geïdentificeerd die zijn vereist in de transferfunctie  $\nu(k)$  in vergelijking (25). Box en Jenkins (1976) lieten zien dat de theoretische CCF tussen  $\alpha_i$  en  $\beta_i$  zich direct proportioneel verhoudt tot  $\nu(k)$ .
5. Vervolgens wordt een ruismodel geïdentificeerd voor de reeks:

$$\hat{n}_i = (z_i - \bar{z}) - \sum_{k=1}^K \hat{\nu}(k)(x_i - \bar{x}) \quad (32)$$

waarin  $K$  de *lag* is waarop de impuls-responsfunctie  $\hat{\nu}(k)$  is uitgedempt, en waarbij gebruik wordt gemaakt van de geschatte ACF voor  $\hat{n}_i$ , die gegeven is in vergelijking (8).

## 3 Tijdreeksanalyse vanuit fysisch-hydrologisch perspectief

### 3.1 Over statistiek en fysica

Bij het omgaan met (gestoorde) signalen of tijdreeksen zijn er altijd al verschillende scholen geweest. Aan de ene kant van het spectrum staan de 'klassieke' ARIMA tijdreeksmodellen, die in het vorige hoofdstuk aan de orde kwamen. Deze vinden hun oorsprong in de alfa- en gammawetenschappen, en hebben een sterk statistische, beschrijvende invalshoek. Eigenlijk logisch, want in deze hoek van de wetenschap is het functioneren van het systeem dat onderzocht wordt vaak niet of niet op voorhand op een fysisch-deterministische manier te beschrijven. De basis van deze groep van tijdreeksmodellen werd in 1970 gelegd door Box en Jenkins, met de publicatie van het boek 'Time Series Analysis, Forecasting and Control'. Hierin werd de statistische tijdreeksanalysetheorie voor het eerst goed uitgewerkt en uitgebreid behandeld. Vanwege de statistische invalshoek zijn ARIMA modellen zeer generiek van opzet, en toepasbaar op allerlei tijdreeksen, of het nu om grondwaterstanden gaat, het temperatuurverloop van een haardroger, of het grillige verloop van de aandelenkoersen. Juist ook door hun generieke opzet hebben dit soort methoden een grote vlucht genomen, en worden in allerlei disciplines toegepast.

Aan de andere kant van het spectrum staan de methoden die hun oorsprong vinden in de bèta of natuurwetenschappen. In feite gaat de basis daarvan terug tot Einstein (1905), die als eerste de zogenoemde Brownse beweging ofwel de chaotische beweging van stuifmeelkorrels in een glas met water wiskundig wist te beschrijven. Wiener (1949) kan gezien worden als de grondlegger van de filteringsmethoden die hieruit voortvloeiden. Dit soort methoden vonden hun eerste grootschalige toepassing in de navigatie, communicatie en meet en regeltechniek, waar het werken met gestoorde signalen en ruis aan de orde van de dag is. De ontwikkeling van het zogenoemde Kalman filter (Kalman, 1960), vormde bovendien een belangrijke stimulans voor de toepassingen ervan.

In de hydrologie is in eerste instantie vooral veel aandacht geweest voor het gebruik van ARIMA modellen. Een mijlpaal daarbij was het verschijnen van het boek van Hipel en McLeod (1994), dat helemaal gericht was het gebruik van dit soort modellen binnen de hydrologie. Ook in Nederland heeft het gebruik van ARIMA modellen al gauw een grote vlucht genomen, misschien wel vanwege de relatief grote hoeveelheid meetgegevens die hier beschikbaar is. Zo verschenen in 1988 twee artikelen waarin Box-Jenkins tijdreeksmodellen aan een breder publiek werden voorgesteld (Van Geer et al., 1988; Baggelaar, 1988). Ook de verdroging in Nederland is al vroeg in kaart gebracht met behulp van tijdreeksanalyse (Rolf, 1989). De laatste jaren is er echter ook steeds meer aandacht voor de fysisch-hydrologische aspecten

van tijdreeksanalyse. Dit vanuit het inzicht dat het transferegedeelte van een transferruis-model (figuur 10) de fysisch-hydrologische werking van een grondwatersysteem beschrijft, of zou moeten beschrijven.

Deze fysisch-hydrologische kijk op tijdreeksanalyse heeft inmiddels twee belangrijke toepassingen gekregen. Enerzijds door al bij voorbaat (a-priori) gebruik te maken van fysisch-hydrologische wetmatigheden in de kern van het tijdreeksmodel zelf, i.e. de impulsresponsfunctie. Voorbeelden hiervan zullen in dit hoofdstuk de revue passeren. Anderzijds door achteraf (a-posteriori) de tijdreeksanalyse-resultaten te toetsen aan wat men fysisch-hydrologisch gezien zou mogen verwachten. De resultaten kunnen bovendien gebruikt worden om daarmee een schatting te krijgen van de (geo)hydrologische eigenschappen van het onderzochte systeem. Dit komt in paragraaf 5.6 aan de orde.

### 3.2 De transferfunctie van een Box-Jenkinsmodel

De relatie tussen tijdreeksmodellen en de werking van het fysisch-hydrologische systeem dat ze beschrijven is het makkelijkst te begrijpen vanuit het perspectief van de zogenaamde transfer- of responsfunctie. We beginnen dus allereerst met onze aandacht hierop te richten. Om maar direct in de hydrologische wereld van grondwaterstanden en neerslagen terecht te komen, schakelen we over op een iets andere notatie dan tot nu toe. Waar we in het voorgaande hoofdstuk nog gewerkt hebben met algemene symbolen die voor een willekeurige factor gebruikt kunnen worden, schrijven we de basisvergelijkingen van het TFN-model (27-29) voor het analyseren van grondwaterstandsreeksen nu als:

$$h_t = h_t^* + n_t + d \quad (33)$$

$$h_t^* = \Theta_0 p_t + \Theta_1 p_{t-1} + \Theta_2 p_{t-2} + \dots = \sum_{i=0}^{\infty} \Theta_i p_{t-i} \quad (34)$$

$$n_t = \varphi_0 a_t + \varphi_1 a_{t-1} + \varphi_2 a_{t-2} + \dots = \sum_{i=0}^{\infty} \varphi_i a_{t-i} \quad (35)$$

met

$h_t$  = waargenomen grondwaterstand op tijdstip  $t$  (met  $t \in \mathbb{Z}$ , [-]) ten opzichte van NAP [m]

$h_t^*$  = voorspelde grondwaterstand op tijdstip  $t$  op grond van de neerslag, ten opzichte van  $d$  [m]

$n_t$  = de residu-reeks [m]

$d$  = het niveau van  $h_t$  zonder neerslag, of met andere woorden het lokale drainageniveau ten opzichte van NAP [m]

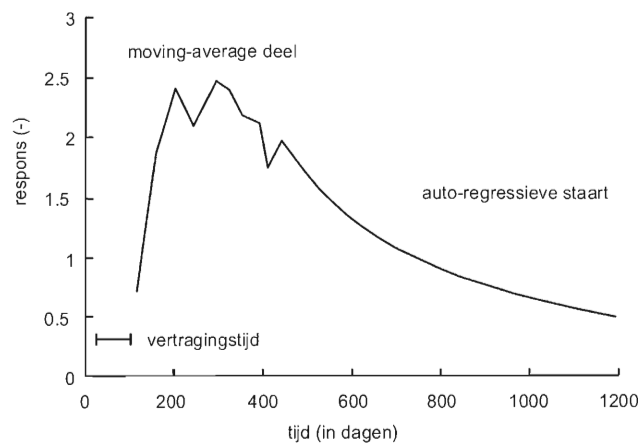
$\Theta_i$  = transferfunctie [-]

$p_t$  = de hoeveelheid neerslag op tijdstip  $t$  [m]

$a_t$  = een discreet witte ruis proces [m] met  $E[a] = 0$  en  $E[a^2] = \sigma_a^2$

$\varphi(t)$  = ruis-transferfunctie [-]

In (34) is het transfer-model uit vergelijking (28) omgeschreven naar een pure MA vorm, zodat de coëfficiënten tezamen de transferfunctie  $\Theta(t)$  vormen. Wanneer je de transferfunctie van een Box-Jenkinsmodel bekijkt (figuur 11) zie je daarin duidelijk de verschillende functies van de AR en MA parameters en vertragingstijd terug. Met het identificeren van de orde van een TFN-model



figuur 11: Voorbeeld van een transferfunctie van een BJ-model met een vertragingstijd, 10 moving-average- en 1 auto-regressieve-parameter.

wordt in feite de vorm gespecificeerd die de transferfunctie aan kan nemen. De moving-average parameters zorgen er voor dat de transferfunctie aan het begin geheel vrij is qua vorm, terwijl een auto-regeressive parameter ervoor zorgt dat de staart van de transferfunctie exponentieel is. De modelleur zoekt met het toevoegen van extra MA-parameters als het ware naar het punt van de transferfunctie van waaraf de staart van de respons benaderd kan worden door een exponentiële functie, wat doorgaans net over de top van de curve is.

Alhoewel het vaak niet expliciet gebeurt, dient de modelleur echter ook de frequentie van de grondwaterstandsreeks  $th$  en die van het neerslagoverschot  $tp$  te specificeren (of wordt hij of zij daartoe gedwongen door de praktijk). Box-Jenkinsmodellen hebben een aantal beperkingen voor wat betreft het omgaan met data met een onregelmatige frequentie, die al in de basis van vergelijkingen (27-29) ligt. Tijd wordt hierin opgevat als een dimensieloze index  $t-i$  met  $i \in \mathbb{Z}$ , waarbij een tijdstap aldus altijd gelijk is aan 1, ongeacht de werkelijke meetfrequentie. Omdat dezelfde index wordt gebruikt voor  $h_t$  en  $p_t$  kunnen bovendien de frequenties van de in- en uitvoerreeksen niet verschillen. De reeks met de laagste frequentie is maatgevend voor de frequentie van het model, en aldus wordt de modelleur gedwongen om de extra waarnemingen die hij of zij heeft van de andere reeksen overboord te zetten. Omdat bovendien het aantal parameters afhangt van de model- en waarnemingsfrequentie, moet de modelleur ook deze twee factoren tegen elkaar afwegen. AR parameters benaderen bovendien bij toenemende frequentie asymptotisch de waarde van 1, wat problemen op kan leveren door een slechte schaling daarvan.

### 3.3 Van Box-Jenkins- naar PIRFICT-model

Net als neerslag, verdamping en grondwaterstandsfluctuaties zijn veel processen die met TFN-modellen worden gemodelleerd in werkelijkheid niet discreet maar continu. Omwille van het leggen van een directe link met de fysica (meer uitleg hierover later), maar ook om de beperkingen van Box-Jenkinsmodellen met betrekking tot de model- en waarnemingsfrequentie te

omzeilen, zullen we hieronder het domein van de discrete Box-Jenkinsmodellen verlaten, en overschakelen naar het continue tijdsdomein en het zogenaamde PIRFICT-model. In continue tijd wordt een sommatie een integraal, zodat we vergelijkingen (33-34) kunnen schrijven als:

$$h(t) = h^*(t) + n(t) + d \quad (36)$$

$$h^*(t) = \int_{-\infty}^t p(\tau) \theta(t - \tau) d\tau \quad (37)$$

Een integraal in de vorm van (37) wordt convolutieintegraal genoemd. Wanneer we vergelijking (35) transformeren naar continue tijd ontstaat een stochastische integraal, waarvan de oplossing onder voorwaarde van een exponentiële IR-functie voor het ruisproces geschreven kan worden als:

$$v(t) = n(t) - e^{-\alpha \Delta t} n(t - \Delta t) \quad (38)$$

waarbij  $v(t)$  de innovatiereeks vormt. Een innovatie is daarbij gedefinieerd als het verschil tussen een waarneming en de zogenaamde 1-stap-vooruit voorspelling van het model. Het hoe en waarom hiervan is eigenlijk een verhaal apart, waarop we hier niet verder in willen gaan, maar volstaan met te verwijzen naar (Von Asmuth en Bierkens, in press.).

Met het enkel transformeren van de vergelijkingen zijn we er natuurlijk nog niet. Ook de beschikbare invoerreeksen moeten getransformeerd worden naar continue reeksen, aangezien de meeste reeksen op de een of andere manier gediscretiseerd zijn om op een digitaal medium opgeslagen te kunnen worden. Veel tijdreeksen, zoals neerslaghoeveelheden, kunnen echter beschouwd worden als een verandering van de primitieve functie van het onderliggende continu proces, ofwel de neerslagintensiteit:

$$P(t_i) - P(t_{i-1}) = \int_{t_{i-1}}^{t_i} p(\tau) d\tau \quad (39)$$

Uit vergelijking (39) valt logischerwijze de neerslagintensiteit  $p(t)$  niet exact op te lossen voor elke  $t$  als  $P_t$  slechts discreet beschikbaar is. We kunnen echter wel een benadering voor  $p(t)$  vinden door aan te nemen dat  $p(t)$  uniform verdeeld is over de periode  $t_{i-1}$  tot  $t_i$ . Vergelijking (39) kan dan geschreven worden als:

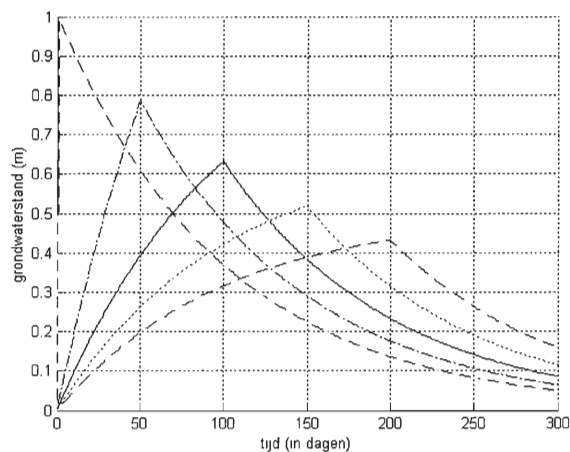
$$p(t) = \frac{P(t_i) - P(t_{i-1})}{t_i - t_{i-1}}, \quad t_{i-1} \leq t < t_i \quad (40)$$

Op deze manier wordt de benadering van het verloop van de neerslagintensiteit  $p(t)$  beter naarmate de tijdstappen kleiner worden. Als de invoerreeksen omgezet zijn naar continue reeksen kan vergelijking (37) geëvalueerd worden door gebruik te maken van de blokrespons(BR)-functie. Waar in discrete TR-modellen de gewichten van de transferfunctie beschouwd worden als respons op een impuls, is vanuit de continue optiek dit de respons op een invoerreeks in de vorm van een blok. Om verwarring te voorkomen gebruiken we de term

transferfunctie voor het discrete geval en refereren we aan zijn continue equivalent als de BR-functie  $\Theta(t)$ . De term IR-functie kan dan gereserveerd worden voor de respons op een feitelijk instantane impuls. In tegenstelling tot de BR-functie en de transferfunctie van discrete modellen is de IR-functie van een convolutieproces geen functie van  $\Delta t$ , en hangt met andere woorden dus niet af van de modelfrequentie. De IR-functie beschrijft de reactie van de grondwaterspiegel op een plotselinge impuls van neerslag en is onveranderlijk in de tijd en karakteristiek voor een bepaald systeem, zolang de fysica van het systeem niet verandert (Von Asmuth en Maas, 2001). De IR-functie is daarmee verwant aan de zogenaamde 'instantaneous unit hydrograph' die in de oppervlaktewater-hydrologie gebruikt wordt. De BR-functie  $\Theta(t)$  wordt verkregen door de IR-functie te convolueren met een 'blok' neerslagoverschot met eenheidsintensiteit over een periode  $\Delta t$ , wat gelijk staat aan:

$$\Theta(t) = \int_{t-\Delta t}^t \theta(\tau) d\tau \quad (41)$$

In figuur 12 worden een aantal voorbeelden van BR-functies van een enkel lineair reservoir getoond voor verschillende tijdstappen. Vanwege het superpositiebeginsel kan het verloop van de grondwaterspiegel  $h^*(t)$  worden verkregen door de responsfuncties van alle 'blokken' van neerslag bij elkaar op te tellen. Omdat  $\Theta(t)$  een continue functie is, is  $h^*(t)$  zelf ook continu gedefinieerd, en kan er vervolgens voor elke waarneming van  $h(t)$  ook een waarde van de residueeks  $n(t)$  worden verkregen. Wanneer uit deze residueeks met behulp van vergelijking (38) een innovatiereeks gestedilleerd wordt, zijn daarmee alle modelvergelijkingen geëvalueerd en kan het model worden ingebed in een optimalisatiealgoritme (e.g. Marquardt-Levenberg) om de parameters te schatten.



figuur 12: Voorbeelden van blokrespons-functies van een enkel lineair reservoir voor verschillende tijdstappen ( $A=100$ ,  $\alpha=0.01$ ,  $n=1$ ,  $\Delta t = [1 \ 50 \ 100 \ 150 \ 200]$ )

### 3.4 De responsfunctie van verschillende typen systemen

#### 3.4.1 Een bodemkolom: 1-dimensionale stroming

We hebben in het voorgaande in het midden gelaten wat precies de IR-functie  $\theta(t)$  is die we in ons continue TFN-model zullen gaan gebruiken. Dat is daarmee direct een voordeel, omdat we nu vrij zijn om die keuze te maken. Door de formulering van het model kunnen we nu in principe elke continue wiskundige (al dan niet analytisch-hydrologische) functie daarvoor gebruiken. We beginnen de mogelijkheden te verkennen bij de meest eenvoudige optie, het lineaire reservoir (ofwel een emmer met zand van waaruit het water via een gat in de bodem wegstroomt). Martin Knotters en Marc Bierkens (Knotters en Bierkens, 2000)

waren de eersten die een direct verband wisten te leggen tussen de parameters van een tijdreeksmodel en een

eenvoudige schematisatie van een bodemkolom. Zij lieten zien dat een ARX(1,0) model een gediscretiseerde versie is van een enkel lineair reservoir. De werking van een lineair reservoir staat op zijn beurt gelijk aan een simpel fysisch model van een 1-dimensionale grondkolom, waarbij laterale stroming en het gedrag van de onverzadigde zone verwaarloosd worden. De responsfunctie van een dergelijke bodemkolom is:

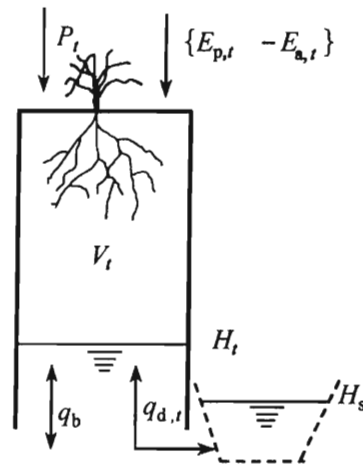
$$\theta(t) = \frac{1}{S} e^{-\frac{t}{cS}} \quad (42)$$

met

$S$  = bergingscoëfficiënt [-]

$c$  = drageweerstand [T]

Vanwege de directe relatie met de fysica kozen Knotters en Bierkens het ARX(1,0) model als meest geschikt ARIMA model voor tijdreeksen van de grondwaterstand. Zodra je immers bij ARIMA modellen kiest voor een hogere orde dan het AR(1) model winnen deze weliswaar aan beschrijvende kracht, maar verliezen hun directe fysische interpretatie. De continue formulering van het PIRFICT model maakt het nu echter ook mogelijk om op



**Figure 1.** Soil column with water balance terms. Input terms include  $P_t$ , potential precipitation excess, i.e., precipitation minus potential evapotranspiration [ $LT^{-1}$ ];  $E_{p,t} - E_{a,t}$ , difference between potential and actual evapotranspiration [ $LT^{-1}$ ]; and  $q_b$ , regional groundwater flux [ $LT^{-1}$ ]. Output terms include  $V_t$ , volumetric water content in the unsaturated zone [ $L$ ]; and  $q_{d,t}$ , drainage flux [ $LT^{-1}$ ].  $H_s$  is the drainage level [ $L$ ];  $H_t$  is table depth [ $L$ ].

figuur 13: Schema van een eenvoudige bodemkolom uit het artikel van Knotters & Bierkens (2000).

zoek gaan te naar functies die meer hydrologische complexiteit kunnen beschrijven, en deze direct in te zetten voor onze analyses.

### 3.4.2 Een poldersysteem: 2-dimensionale stroming

Één stapje complexer dan een lineair reservoir, is een 2 dimensionaal systeem. Het prototype daarvan is een dwarsdoorsnede door een (polder)systeem met evenwijdige, oneindig doorlopende diepe sloten. Als we bovendien veronderstellen dat de doorlatendheid  $K$  van het systeem homogeen is, en dat de opbolling veel kleiner is dan dikte van het pakket, dan geldt voor dit systeem de oplossing van Kraijenhoff van de Leur (Kraijenhof van de Leur, 1958). De responsfunctie van een dergelijk poldersysteem is:

$$\theta_{vz}(t) = \frac{4}{S\pi} \sum_{n=1,3,5,\dots}^{\infty} \frac{1}{n} \sin\left(\frac{n\pi x}{L}\right) e^{\frac{-n^2 t}{j}} \quad (43)$$

met

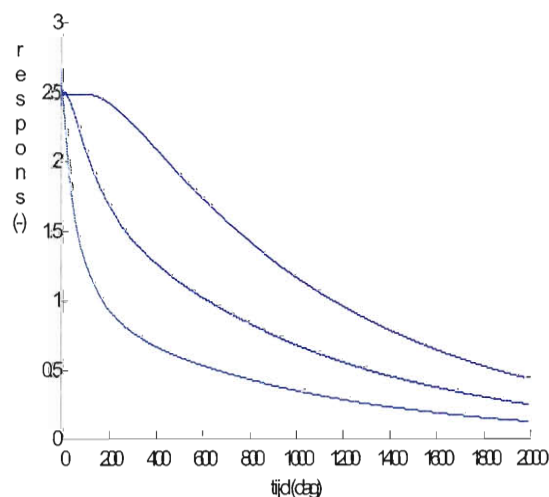
$x$  = afstand tot de dichtstbijzijnde sloot

$L$  = afstand tussen de sloten

$j$  = reservoircoëfficiënt =  $\frac{SL^2}{\pi^2 KH}$

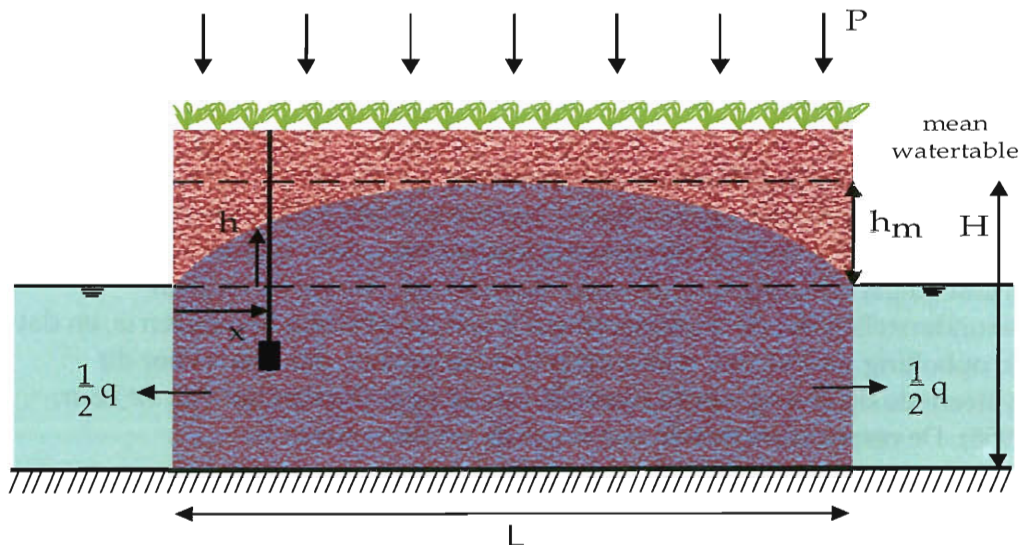
$K$  = doorlatendheid

Wat van belang is, is natuurlijk in hoeverre de neerslagrespons van een dergelijk systeem verschilt van dat van de e-macht van de verticale bodemkolom. Als we een aantal verschillende responsfuncties bekijken valt op dat de neerslag ofwel eerst traag afstroomt, en dan sneller, ofwel eerst sneller en dan trager. De afwijking t.o.v. het exponentiele verloop van het lineaire reservoir wordt het onevenredig afstromend deel genoemd, en is afhankelijk van hoe centraal of juist excentrisch de onderzochte locatie ligt in het systeem. Ook intuïtief is dit wel aan te voelen. Een egaal verspreide hoeveelheid neerslag begint eerst aan de randen af te stromen, daarna pas in het midden.



figuur 14: Voorbeelden van responsfuncties van een 2 dimensionaal poldersysteem volgens Kraijenhoff van de Leur.





figuur 15: Hydrologisch schema van een 2 dimensionaal poldersysteem volgens Kraijenhoff van de Leur.

### 3.4.3 De onverzadigde zone: Convectie en Dispersie

Tot nu toe hebben we de werking van de onverzadigde zone verwaarloosd. Consequentie daarvan is dat zowel bij de e-macht als bij de responsfuncties volgens Kraijenhoff van de Leur de reactie op een neerslagimpuls instantaan is. De neerslag komt in z'n geheel direct bij de verzadigde zone aan, of wel de respons begint niet bij nul, maar direct op z'n hoogste punt, en stroomt daarna af. De meest gedetailleerde dan wel geaccepteerde beschrijving van de stromings- en transportprocessen in de onverzadigde zone wordt op het moment gegeven door de zogenaamde Richards vergelijking. De Richards vergelijking is de basis van onverzadigde zone modellen zoals SWAP (Kroes et al., 2000) en Hydrus (Simunek et al., 1999).

Om in termen van responsfuncties te kunnen blijven moeten we de onverzadigde processen en Richards vergelijking lineariseren. Dit kan door uit te gaan van een 'steady state' situatie, waarbij water door een medium met een uniform vochtgehalte stroomt. De lineaire variant van de Richards vergelijking is de convectie-dispersie vergelijking. Volgens deze vergelijking kan de totale flux geschreven worden als de som van twee termen: een term, die het convectieve of stromingstransport beschrijft, en een tweede term, die de willekeurige menging van de flux beschrijft, als gevolg van diffusie en dispersie van de bewegende vloeistof. In termen van de flux luidt de convectie-dispersie vergelijking (Jury & Roth, 1990):

$$\frac{\partial q}{\partial t} = D \frac{\partial^2 q}{\partial z^2} - v \frac{\partial q}{\partial z} \quad (44)$$

met

$q(z,t)$  = flux op diepte  $z$  t.o.v. maaiveld, en tijdstip  $t$

$D$  = dispersie coefficient

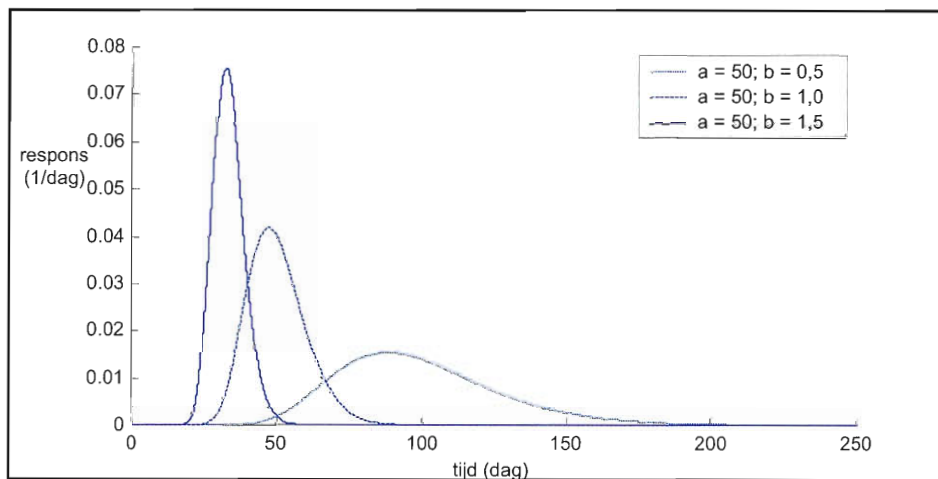
$v$  = effectieve, porie stromingssnelheid, ofwel

$$v = \frac{\text{volume} - \text{flux}}{\text{vochtgehalte}}$$

De oplossing van de convectie-dispersie vergelijking is bekend, en wordt, met een impuls of Dirac delta functie als randvoorwaarde, gegeven door:

$$q(z, t) = \frac{z}{2 \cdot \sqrt{\pi \cdot D \cdot t^3}} e^{-\frac{(z-vt)^2}{4 \cdot D \cdot t}} \quad (45)$$

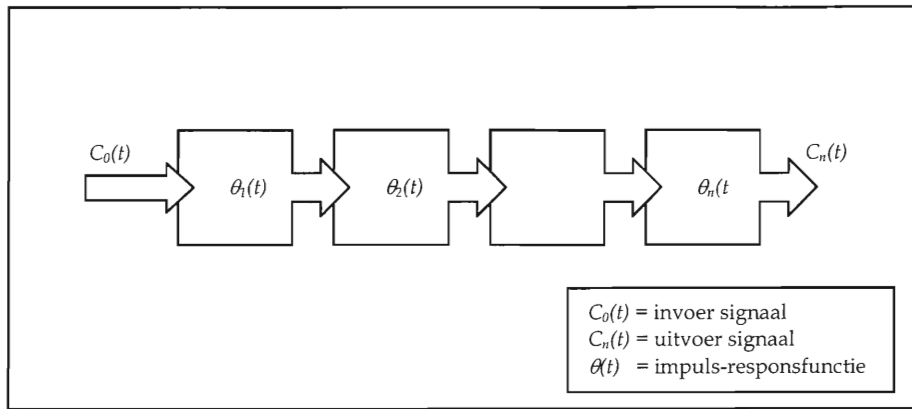
De berekende responsfunctie voor de flux kan gebruikt worden als transferfunctie voor de onverzadigde zone. De functie  $\theta_{bz}(t)$  wordt dan gelijk gesteld met de flux  $q(t)$  op een bepaalde diepte.



figuur 16: Vorm van de IRF voor verschillende waarden van b (oplossing van de CDV)

### 3.4.4 Gekoppelde systemen: een convolutie van responsfuncties

Om aan ons fysisch gebaseerde transfermodel weer een extra stapje complexiteit toe te kunnen voegen zouden we graag de werking van de onverzadigde en de verzadigde zone willen combineren. Deze staan natuurlijk niet los van elkaar, maar zijn gekoppeld. Als we de grondwataanvulling bij de grondwaterspiegel direct zouden kunnen meten, zouden we een apart transfermodel van de onverzadigde zone kunnen maken, maar betrouwbare metingen hiervan zijn nu eenmaal niet makkelijk voorhanden. Wat we echter wel kunnen doen is de werking van de onverzadigde en verzadigde zone combineren, door de responsfuncties daarvan in serie te schakelen.



figuur 17: Een serie systemen, dat een signaal vervormd (Kruithof, 2001)

Beschouw de serie van systemen, dat een invoersignaal omzet in een uitvoersignaal. Toegepast op de stroming van regenwater door de ondergrond kan dit als volgt opgevat worden. De neerslag belandt allereerst op het maaiveld, stroomt daarna door de onverzadigde zone naar de grondwaterspiegel, waar het als grondwateraanvulling de verzadigde zone binnen treedt en een stijging van de grondwaterspiegel veroorzaakt. De grondwateraanvulling stroomt vervolgens via het verzadigde systeem naar de drainagemiddelen, waar het wordt afgevoerd, en uit het systeem verdwijnt. De relatie tussen het invoer- en het uitvoersignaal kan weergegeven worden door een convolutie-integraal. Voor het geval dat het signaal door één systeem getransformeerd wordt, verkrijgen we dezelfde convolutie-integraal als in vergelijking (37):

$$c_{uit}(t) = \int_{-\infty}^{\infty} c_{in}(t - \tau) \cdot \theta(\tau) d\tau \quad (46)$$

Wanneer een signaal vervormd wordt door een serie van systemen, zoals in figuur 17, kan de relatie tussen het invoer- en het uitvoersignaal gegeven worden door een herhaalde convolutie-integraal (Maas, 1994):

$$c_{uit}(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} c_{in}(t - \tau_1) \cdot \theta_1(\tau_2 - \tau_1) \cdot \theta_2(\tau_3 - \tau_2) \cdots \theta_n(\tau_n) d\tau_1 d\tau_2 \cdots d\tau_n \quad (47)$$

Als deze algemene uitdrukking wordt gespecificeerd voor ons probleem met twee systemen, krijgen we:

$$h(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(t - \tau_1) \cdot \theta_{oz}(\tau_2 - \tau_1) \cdot \theta_{vz}(\tau_2) d\tau_1 d\tau_2 \quad (48)$$

waarin

$\theta_{oz}(t)$  = impuls-responsfunctie voor de onverzadigde zone

$\theta_{vz}(t)$  = impuls-responsfunctie voor de verzadigde zone

De samengestelde impuls-responsfunctie van de serie van systemen is dan:

$$\theta(t) = \int_{-\infty}^{\infty} \theta_{oz}(t-\tau) \cdot \theta_{vz}(\tau) \cdot d\tau \quad (49)$$

Als we dit toepassen op een systeem met onverzadigde zone en 2-dimensionale stroming vinden we de volgende uitdrukking voor de impulsresponsfunctie (Kruithof, 2001):

$$\theta(t) = \int_0^{\infty} \frac{a_{oz}}{2\sqrt{\pi(t-\tau)^3}} e^{-\frac{(a_{oz}-b_{vz}(t-\tau))^2}{4(t-\tau)}} \frac{a_{vz}}{\pi} \sum_{n=1,3,5,\dots}^{\infty} \frac{1}{n} \sin(n\pi b_{vz}) e^{-n^2\pi^2 c_{vz}\tau} d\tau \quad (50)$$

Deze integraal is niet in elementaire wiskundige functies uit te drukken. De samengestelde impuls-responsfunctie kan echter wel gevonden worden door de twee afzonderlijke impuls-responsfuncties numeriek te convolveren.

### 3.4.5 De Pearson III functie: een hybride oplossing

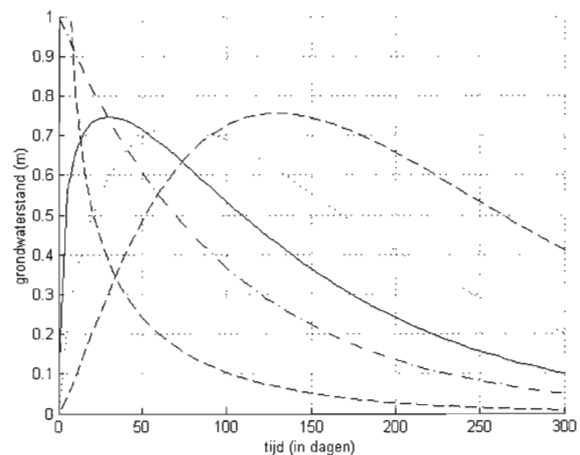
De vorige paragrafen hebben vooral als doel gehad om te verkennen hoe de 'fysisch-hydrologische' respons van een grondwatersysteem op neerslag er zoal uit kan zien, en wat extra fysische complexiteit toevoegt aan het gedrag en de flexibiliteit van de responsfunctie in vergelijking met die van het ARX(1,0) model. De werkelijke natuur, bodemopbouw, processen e.d. in de buitenwereld zijn natuurlijk oneindig veel complexer. Het gaat ons echter niet om het beschrijven van alle complexiteit van het geohydrologische systeem zelf, maar om voldoende flexibiliteit te hebben om met ons transfermodel de relatie tussen neerslagfluctuaties en grondwaterspiegelfluctuaties afdoende te kunnen beschrijven. In die zin kunnen we ook de directe fysische interpretatie van onze responsfuncties loslaten, en overschakelen op functies van algemenere aard met een vergelijkbare flexibiliteit.

Een goed voorbeeld daarvan is de zogenaamde Pearson type III verdelingsfunctie (PIII df). Als we hem toepassen met een extra parameter  $A$  om de oppervlakte te kunnen aanpassen, wordt de PIII df gegeven door:

$$\theta(t) = A \frac{a^n t^{n-1} e^{-at}}{\Gamma(n)} \quad (51)$$

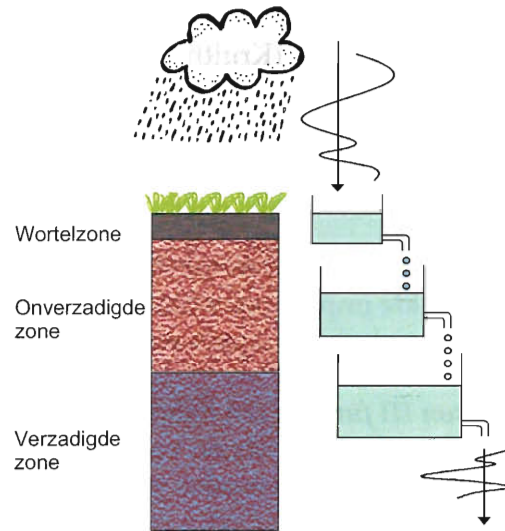
waarbij  $a$ ,  $n$ , en  $A$  de parameters zijn die de vorm van de curve bepalen, en  $\Gamma(n)$  de zogenaamde gamma-

functie. De PIII df kan qua vorm geleidelijk overgaan van een functie die steiler is dan exponentieel, via een exponentiële functie naar een normale verdeling (zie *figuur 18*). In discrete termen omvat de functie daarmee het ARX(1,0) model,



*figuur 18: Voorbeeldkrommen uit de Pearson III df familie ( $A=100$ ,  $\alpha = 0.01$ ,  $n = [0.5 \ 1 \ 1.3 \ 1.7 \ 2.3]$ )* (11)

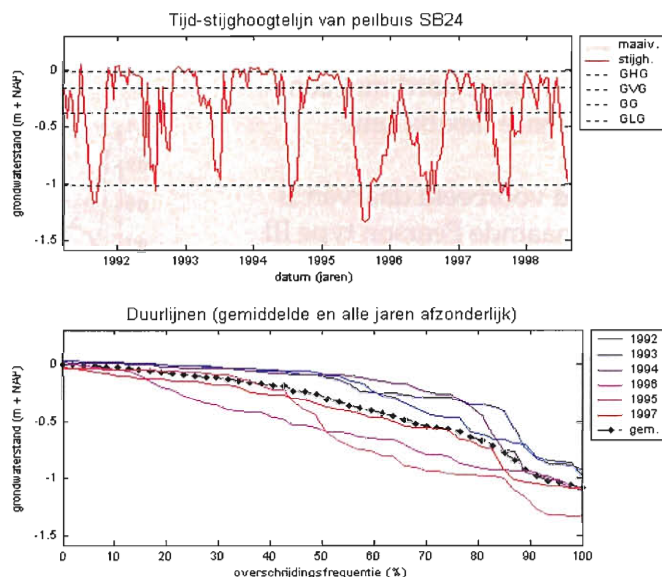
maar bovendien alle ARMAX-modellen waarvan de ordinaten van de transferfunctie een van de krommen uit de PIII familie vormen. De fysische basis van de PIII df ligt in het feit dat het de transferfunctie van een serie van gekoppelde lineaire reservoirs kan beschrijven, waarbij de parameter  $n$  hun aantal weergeeft. Uit een vergelijking van de PIII df met o.a. de convectie-dispersievergelijking (Jury & Roth, 1990) blijkt dat de vorm hiervan nagenoeg identiek is, zodat de PIII df gebruikt kan worden om de gecombineerde respons van de verzadigde en onverzadigde zone te beschrijven. Een voordeel van het gebruik van de Pearson III t.o.v. de fysische oplossing van de respons van een gekoppelde onverzadigde en verzadigde zone zoals in de vorige paragraaf, is enerzijds een stukje rekensnelheid, maar anderzijds ook het vermijden van de schijn van een volledige fysische basis, omdat de fysische betekenis van de parameters door calibratie sowieso vervaagt. De werkelijkheid zal immers nooit gelijk zijn aan de gebruikte hydrologische schematisatie.



figuur 19: De Pearson III formule beschrijft een grondwatersysteem als een serie van lineaire reservoirs.

### 3.5 Wanneer is een grondwatersysteem niet-linear?

Een belangrijke aanname die in traditionele tijdreeksmodellen verborgen zit is dat het grondwatersysteem **lineair** is. Lineariteit wil zeggen dat de grondwaterstand altijd hetzelfde reageert op een eenheids-hoeveelheid neerslag. Het verloop van deze reactie in de tijd wordt **impulsrespons-functie** genoemd.



figuur 20: Bij oppervlakkige afvoer wordt de grondwaterstand 'afgetopt' aan maaiveld.



De lineariteits-aanname gaat eigenlijk tegen de intuïtie in. Je hoeft na een zomerse regenbui maar in de grond te wroeten, of je zult merken dat het water in de bovenste centimeters is blijven hangen. Niet-lineariteit kan, behalve door de **onverzadigde zone**, ook veroorzaakt worden doordat een beek, sloot of ander **drainagemiddel droogvalt**, of doordat anderzijds het **grondwater aan maaiveld** treedt, en al dan niet oppervlakkig wegstroomt. Het optreden van afvoer over maaiveld is iets dat je al vrij gemakkelijk uit de metingen zelf kunt halen (figuur 20). In dat geval zullen de grondwaterstanden en duurlijnen nauwelijks boven maaiveld uitkomen, en er 'afgetopt' uit zien. Bij droogvallende sloten of beekjes zal de grondwaterstand verder weg zakken dan het model aangeeft. Dit type niet-lineariteit wordt **drempel-nietlineariteit** genoemd, omdat het optreedt zodra de grondwaterstand een bepaalde drempelwaarde overschrijdt. Drempel-nietlineariteit uit zich doorgaans doordat het model juist de extreme grondwaterstanden niet goed simuleert.

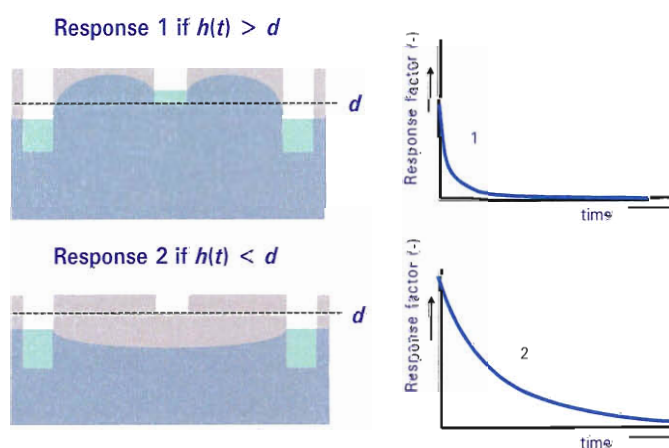
### 3.6 Tijdreeksanalyse met drempel-nietlineariteit

Een oplossing voor het modelleren van grondwatersystemen met drempel-nietlineariteit is o.a. te vinden in (Von Asmuth et al., in prep; Knotters en De Gooijer, 1999; Berendrecht et al., 2004). Alhoewel de methoden op een aantal punten verschillen vertonen is de kern ervan hetzelfde: als de grondwaterstand boven een bepaalde drempel uitkomt treden lokale drainagemiddelen, en daarmee een andere responsfunctie, in werking (figuur 21). De niet-lineariteit zorgt ervoor dat er een aantal belangrijke verschillen zijn ten opzichte van gewone tijdreeksanalyse:

- **Het model werkt (vooral nog) alleen met neerslagoverschot of twee losse neerslag- en verdampingsreeksen**

n. Dit omdat vanwege de wiskundige achtergrond van het model alleen exponentiele responsfuncties mogelijk zijn.

- **Hoog-frequent modelleren is belangrijker** – omdat het tijdstip waarop een bui valt uitmaakt of de neerslag snel of langzaam afgevoerd wordt, is het belangrijker dan bij gewone tijdreeksanalyse om deze gegevens op een hoge frequentie beschikbaar te hebben en mee te nemen. Dit effect wordt echter al voor een belangrijk deel



figuur 21: Als de grondwaterstand boven een bepaalde drempel uitkomt treden lokale drainagemiddelen in werking en is de respons op neerslag veel sneller.

ondervangen door de drempelovergang exact op te lossen in het continue tijdsdomein zoals beschreven in [Von Asmuth et al., in prep].

- **De grondwaterstand is geen optelsom van losse invloeden meer.** *Menyanthes* toont in plaats hiervan de flux naar het lokale drainage systeem wanneer je met behulp van 'view' de modelresultaten oproept. De effecten van de verschillende invloeden zijn niet meer onafhankelijk van elkaar, en dus niet zomaar te scheiden in deelreeksen. Anderzijds is het wel weer mogelijk om de lokale drainage-middelen aan en uit te zetten, ze langzaam te laten verstoppert, etc.

Op alle plaatsen waar dat relevant is laat *Menyanthes* de eigenschappen van beide responsfuncties zien.

**Let op:** Het kan voorkomen dat het niveau van de drempel boven of onder dat van de grondwaterstanden geschat wordt. In dat geval is de drempel niet effectief, en kunt u beter een 'normaal' lineair tijdreeksmodel gebruiken.

## 4 Tijdreeksanalyse met meerdere verklarende reeksen

### 4.1 Verschillende typen invloedsfactoren

In hoofdstuk 3 is in feite alleen de neerslagrespons onder de loep genomen, maar er zijn natuurlijk allerlei andere factoren die de grondwaterstand kunnen beïnvloeden. Bij het importeren van de data van een bepaalde invloed of verklarende reeks dient u binnen *Menyanthes* op te geven om welk type invloed het gaat. De verschillende typen die het programma onderscheidt zijn: **Precipitation, Evaporation, Well, River, Impulse** en **Box-Jenkins**. Het feit dat een onderscheid gemaakt wordt heeft de volgende redenen:

- **Eenheid van de reeks** - Voor het schattingsproces en de interpretatie van de parameters is de eenheid van een reeks van belang. Bij grondwater-onttrekkingen gaat het bijvoorbeeld om tienduizenden kubieke meters per dag, terwijl het bij neerslag en verdamping millimeterwerk is. Hiermee kan bijvoorbeeld bij de initiële parameterschatting rekening worden gehouden.
- **Teken van de invloed** - De ene reeks heeft nu eenmaal een stijging van de grondwaterstand tot gevolg, terwijl die door een andere invloed altijd daalt. Ook dit soort basale voorkennis wordt gebruikt bij de schatting.
- **Losse waarnemingen of integraal over de tijdstap** - Bij sommige reeksen, bijvoorbeeld onttrekkingen of neerslag, bestaan de gegevens uit een integraal of sommatie van de intensiteit of het debiet over de tijdstap, terwijl het bij andere reeksen, bijvoorbeeld rivierpeilen, gaat om afzonderlijke waarnemingen van het niveau. Bij het interpoleren van de data moet hiermee rekening worden gehouden.
- **Type responsfunctie** - De aard van de respons van een grondwatersysteem verschilt in principe naar gelang de aard van de invloed. In de volgende paragraaf zullen de verschillende responsfuncties de revue passeren.

### 4.2 De responsfunctie en interpretatie van verschillende invloedsfactoren

#### 4.2.1 Het type *PREC* oftewel neerslag(overschot)

In paragraaf 3.4.5 is de Pearson III functie al vlug aan de orde gekomen. We pikken hier de draad weer op, en gaan verder met de fysische betekenis van de parameters daarvan. Onder de knop *Statistics* in het kwadrant *Time Series Models* levert *Menyanthes* voor de impulsrespons van het type *prec* drie parameters op. (Als het model niet-lineair is, zijn er zes parameters. Daar gaan we straks op in). Hier is een bijvoorbeeld:



- input series: HAAMSTEDE type: prec
- parameter 1 ( $M_0$ ): 85.68 stdev : 25
- parameter 2 : 0.0113 stdev : 0.0072
- parameter 3 : 0.921 stdev : 0.29

De eerste parameter is het zogenaamde *nulde moment*  $M_0$ . In de tijdreeksanalyse wordt deze parameter ook wel aangeduid als *gain*. Wiskundig gezien is het nulde moment de oppervlakte van de impulsresponsfunctie.  $M_0$  is gelijk aan parameter  $A$  uit (51). Fysisch gezien geeft  $M_0$  de verhouding weer tussen de uiteindelijke stijghoogte en de intensiteit van de regenval, als het heel lang zou blijven regenen. Die verhouding kan men met recht *drainageweerstand* (in dagen) noemen, maar daarbij moet aangetekend worden dat het een waarde is die peilbuisgebonden is. In de agrohydrologie is het gebruikelijk om de term drainageweerstand te betrekken op de grondwaterstand in het midden van een perceel. Geohydrologen betrekken hem gewoonlijk op de gemiddelde grondwaterstand in een rekencel van een grondwatermodel. De grootte van  $M_0$  hangt sterk af van de grootte van het afstromingsstelsel, maar binnen zo'n stelsel varieert  $M_0$  weer sterk met de plaats van de peilbuis. Als men voor ogen houdt dat  $M_0$  de stationaire stijghoogte is bij een neerslag van 1 m/d, dan heeft men - hoe onzinnig die neerslagintensiteit ook is - wel een gevoel bij de vorm van het  $M_0$ -vlak. Als een peilbuis dicht bij een drainerend middel staat, zoals een sloot of een rivier, zal  $M_0$  slechts enkele dagen bedragen, maar in grote systemen kan hij ver van de waterlopen gemakkelijk waarden van duizenden dagen bereiken. Behalve de grootte van het stelsel speelt ook het doorlaatvermogen een rol.

De tweede parameter is de  $a$  uit (51). Hij zegt iets over de snelheid waarmee het grondwater na een regenbui wegzakt (het *uitputtingsverloop*), dus over het geheugen van het stelsel. De eenheid is 1/d, dus bij een lang geheugen hoort een kleine  $a$ -waarde. In kleine systemen, zoals percelen, bedraagt  $a$  1/(enkele dagen). In grote systemen kan de waarde afnemen tot 1/(enkele jaren). Behalve de grootte van het stelsel spelen echter ook het doorlaatvermogen en de bergingscoëfficiënt een rol. Een correctere maat voor het geheugen van een stelsel is de gemiddelde waarde van de impulsresponsfunctie, die gegeven wordt door:

$$\mu = \frac{n}{a} \quad (52)$$

Hierin is  $n$  de derde parameter uit het rijtje. (Daar gaan we straks op in). Na een tijdsverloop van  $\mu$  dagen is een regenbui voor de helft uitgezakt. Een bijzondere eigenschap van  $\mu$  is dat hij binnen een stelsel niet veel in grootte varieert. Het maakt voor de waarde van  $\mu$  dus niet zoveel uit of een peilbuis dicht bij een drainerende waterloop staat of ver daar vanaf. Daardoor is  $\mu$  een goede parameter om de traagheid van een geheel systeem te

karakteriseren. Voor de goede orde vermelden we dat de standaardafwijking van de impulsresponsfunctie gegeven wordt door

$$\sigma = \frac{\sqrt{n}}{a} \quad (53)$$

Er is geen precies tijdstip aan te geven waarop een regenbui helemaal uit een systeem verdwenen is, maar een praktische maat is  $\mu + 2\sigma$  dagen.

De *derde* parameter is dus de parameter  $n$  uit (51). Hij ligt gewoonlijk in de buurt van 1. Voor het rekenschema van Kraijenhoff van de Leur (figuur 15) is theoretisch aantoonbaar dat  $n$  varieert van 0.7 aan de rand van het systeem tot 1.05 in het midden. De grootte van het systeem of de bodemeigenschappen spelen daarbij geen rol. De enige parameter die invloed op  $n$  heeft, is relatieve de positie van de peilbuis. De *vorm* van het systeem heeft wel enige invloed; voor een cirkelvormig eiland varieert  $n$  van 0.68 aan de rand tot 1.14 in het midden. Als  $n$  gelijk is aan 1, dan is de impulsresponsfunctie een zuivere e-macht. In de tijdreeksanalyse wordt dan wel van een AR(1)-model gesproken. Is  $n$  groter dan 1, dan vertoont het grondwaterstandsverloop naar verhouding weinig pieken. Op plaatsen met  $n$  kleiner dan 1 vertoont de grondwaterspiegel juist een piekig verloop. Maar let op: lange tijdreeksen zien er altijd piekeriger uit dan korte, doordat in het plaatje de tijdschaal sterker gecomprimeerd wordt. Dit is dus vaak gezichtsbedrog. In de praktijk komen we wel waarden van  $n$  tegen die wat kleiner zijn dan 0.68. Waarden groter dan 1.14 komen vaker voor. We hebben de oorzaak nog niet goed geanalyseerd, maar het lijkt aannemelijk dat de dikte van de onverzadigde zone daarin een rol speelt.

**Niet-lineaire modellen** voor neerslag en verdamping zijn gebaseerd op een speciaal geval van de verdeling van Pearson type III, namelijk de zuivere e-macht. Deze is uit (51) te verkrijgen door  $n = 1$  te stellen. Niet-lineaire modellen kennen zes parameters, die weer met de knop *Statistics* in het kwadrant *Time Series Models* tevoorschijn geroepen worden. Voorbeeld:

- input series: HAAMSTEDE type: prec
- parameter 1 (A1): 4.34 stdev: 0.47
- parameter 2: 0.000518 stdev: 0.0015
- parameter 3 (dbase): 2.445 stdev: 4.9
- parameter 4 (A2): 1.711 stdev: 0.32
- parameter 5: 0.02364 stdev: 0.0038
- parameter 6 (thold1): 0.9931 stdev: 0.036

De zesde parameter,  $thold_1$ , behandelden we onder het kopje *Drainagebasis* (paragraaf 5.6.1). Gewoonlijk is dit het niveau van greppels of buisdrainage.

De eerste parameter ( $A_1$ ) komt overeen met  $Aa$  uit (51). Hij heeft op zichzelf geen bruikbare fysische betekenis. Deelt men de eerste parameter door de tweede, dan verkrijgt men  $M_0$ , ofwel de *drainageweerstand* ofwel de *gain*, in tijden waarin het grondwaterniveau beneden  $thold_1$  verwijlt. (De betekenis van deze termen is al bij de behandeling van de lineaire modellen toegelicht).

De tweede parameter is de parameter  $a$  uit (51). Hij bepaalt hoe snel de grondwaterspiegel wegzijgt. Aangezien  $n = 1$  geldt voor het gemiddelde en de standaardafwijking van de impulsresponsfunctie:

$$\mu = \sigma = 1/a \quad (54)$$

(zie ook (52) en (53)). Na verloop van  $\mu$  dagen is een regenbui voor de helft uitgezakt. Het duurt ca  $\mu + 2\sigma = 3/a$  dagen voor een bui praktisch geheel uit het systeem verdwenen is.

De derde parameter is de drainagebasis, die in paragraaf 5.6.1 behandeld wordt. In het geval van niet-lineaire modellen is de drainagebasis vaak moeilijk vast te stellen. Als er merkwaardige getallen uitkomen, hoeft u de modellen nog niet meteen te verwerpen.

De vierde parameter ( $A_2$ ) komt overeen met  $Aa$  uit (51), maar nu in een situatie met het grondwaterniveau bóven  $thold_1$ . De vierde parameter gedeeld door de vijfde geeft weer  $M_0$ , ofwel de *drainageweerstand* ofwel de *gain*, in tijden waarin het grondwaterniveau bóven  $thold_1$  uitkomt. (De betekenis van deze termen is al bij de behandeling van de lineaire modellen toegelicht).

De vijfde parameter is parameter  $a$  uit (51). Hij bepaalt hoe snel de grondwaterspiegel bóven  $thold_1$  wegzijgt. Voor het gemiddelde en de standaardafwijking van de impulsresponsfunctie geldt weer:

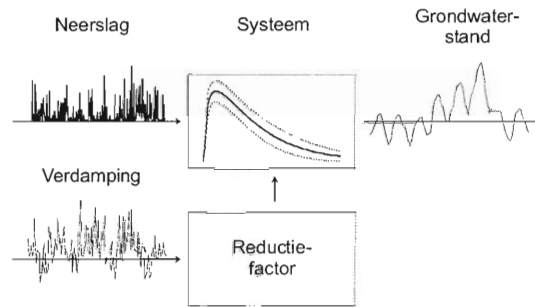
$$\mu = \sigma = 1/a \quad (55)$$

(zie ook (52) en (53)). Na verloop van  $\mu$  dagen is een regenbui voor de helft uitgezakt. Het duurt ca  $\mu + 2\sigma = 3/a$  dagen voor een bui praktisch geheel uit het systeem verdwenen is.

#### 4.2.2 Het type EVAP oftewel verdamping

Bij het bepalen van de invloed van de verdamping op het verloop van de grondwaterstand is als uitgangspunt gekozen dat verdamping dezelfde weg volgt als neerslag, maar met een tegengesteld teken. De consequentie is dat de impulsresponsfunctie van neerslag en verdamping identiek zijn, op een schaalfactor na die we *verdampingsfactor* (*evap.fctr* of *red.fct*) noemen (figuur 22). De schaalfactor is enerzijds nodig omdat de referentieverdamping die

door het KNMI opgegeven wordt geldt voor een vegetatie van kort gras. De referentieverdamping geldt bovendien voor een vegetatie die optimaal van water is voorzien, zodat anderzijds ook verdampingstekorten effect hebben op de verdampingsfactor die ingeschat wordt. Vaak groeit er in de omgeving van de te analyseren peilbuis iets anders, en meestal wordt een deel van de omgeving in beslag genomen door waterlopen, wegen, bebouwing etc. Uit dien hoofde moet men verwachten dat de verdampingsfactor tussen, zeg, 0.5 en 2 ligt. Is hij duidelijk hoger, dan kan dat een aanwijzing zijn dat er nog een andere seizoensafhankelijke invloed in het spel is, die men gemist heeft. Meestal gaat het dan om grondwaterwinning voor de drinkwatervoorziening, voor beregening of voor klimaatregeling. Gemiste beregening kan ook juist tot te lage waarden van de verdamping leiden, evenals peilbeheer.



figuur 22: Onder de aanname dat verdamping dezelfde respons heeft als neerslag wordt alleen een reductiefactor ingeschat.

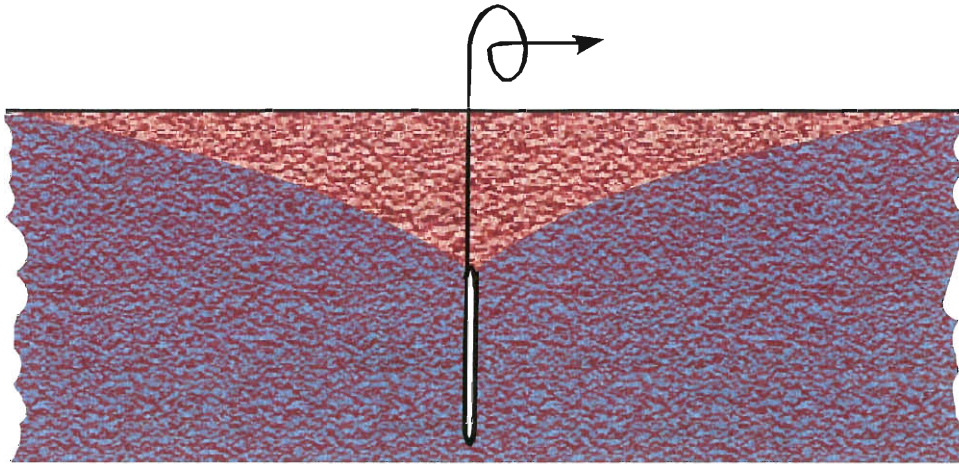
De verdampingsfactor die Menyanthes berekent is weer te vinden onder de knop *Statistics* in het kwadrant *Time Series Models*. Voorbeeld:

- input series: Vlissingen\_evap type: evap
- parameter 4 (red fct): 1.99 stdev: 0.44

U treft de verdampingsfactor (*evap.fctr*) ook rechts van het preview-venster aan, in het kwadrant *Time Series Models*.

#### 4.2.3 Het type WELL oftewel onttrekkingen

De responsen op andere oorzaken dan neerslag of verdamping laten zich vaak minder goed modelleren met de verdeling van Pearson type III. Daarom kent Menyanthes ook invloedsreeksen (*explanatory series*) van de typen *well* en *river*.



figuur 23: Hydrologisch schema van een onttrekking in een homogeen watervoerend pakket volgens Hantush.

De pompproefformule van Hantush (*Hantush en Jacob, 1955*) staat model voor de impulsrespons van het type *well*. Het rekenschema betreft een pompput met een lang filter in een aquifer met doorlatendheid  $kD$  en bergingscoëfficiënt  $S$ , die afgedekt is door een bergingsloze aquitard met weerstand  $c$  (figuur 23). Een standaard pompproef levert in een peilbuis echter geen impulsrespons op, maar een staprespons. De impulsrespons is daarvan per definitie de afgeleide naar de tijd, waarbij het debiet van de put 1 gesteld is:

$$\theta(t) = -\frac{1}{4\pi kDt} \exp\left(-\frac{r^2 S}{4kDt}\right) \exp\left(-\frac{t}{cS}\right) \quad (56)$$

Hierin is

- $\theta(t)$  = impulsrespons (1/m)
- $kD$  = doorlaatvermogen van de aquifer (m)
- $c$  = weerstand tegen voeding (d)
- $S$  = bergingscoëfficiënt (-)

Deze formule blijkt doorgaans ook bij een bodemopbouw die duidelijk afwijkt van figuur 23 in staat te zijn om het verloop van de grondwaterstand in een peilbuis goed te beschrijven. Maar de fysische interpretatie deugt dan natuurlijk niet meer. Daarom werkt Menyanthes met abstracte parameters:

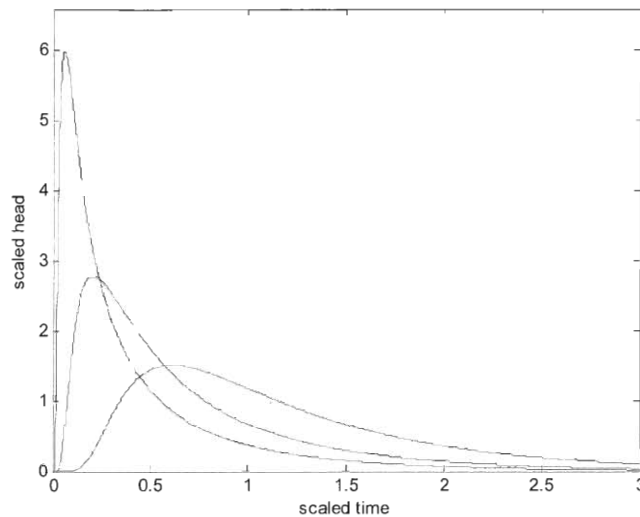
$$a = \frac{r}{2\sqrt{kDc}}, \quad b = \frac{1}{\sqrt{cS}}, \quad c' = \frac{1}{4\pi kD} \quad (57)$$

zodat:

$$\theta(t) = -\frac{c'}{t} \exp\left(-\frac{a^2}{b^2 t}\right) \exp(-b^2 t) \quad (58)$$

Net zoals bij de verdeling van Pearson type III is dit een familie van krommen, waaruit Menyanthes tijdens het modelleren het meest geschikte familielid selecteert. In figuur 24 worden een paar familieleden getoond.

**Betekenis van de parameters die Menyanthes berekent:**



figuur 24: Enkele voorbeelden van impulsresponsfuncties van het type Well.

Onder de knop *Statistics* in het kwadrant *Time Series Models* levert Menyanthes

voor de impulsrespons van het type *well* drie parameters op. Voorbeeld:

- input series: Manderheide type: well
- parameter 5 : 0.154 stdev : 0.0242
- parameter 6 : 0.0797 stdev : 0.0124
- parameter 7 : 2.81e-005 stdev : 7.88e-006

De *eerste parameter* (hier parameter 5 genoemd) is  $a$  uit (58). Ruwweg geeft hij de afstand van de peilbuis tot de *well* weer, als fractie van de invloedsdiameter van de winning. Hij zal meestal tussen 0 en 1 à 2 liggen. Aangezien de afstand tot de *well* bekend is, kan hieruit *globaal* de invloedsdiameter van de winning bepaald worden, maar deze verwachting is gebaseerd op het rekenschema van figuur 23. Dit aspect hebben we nog niet grondig bestudeerd.

De *tweede parameter* (hier parameter 6) is  $b$  uit (58). Hij zegt iets over de tijd die een onttrekking nodig heeft om volledig door te werken, dus over het geheugen van het systeem. Een correctere maat voor het geheugen van het systeem is de gemiddelde waarde van de impulsresponsfunctie, die gegeven wordt door

$$\mu = \frac{a}{b^2} \frac{K_1(2a)}{K_0(2a)} \quad (59)$$

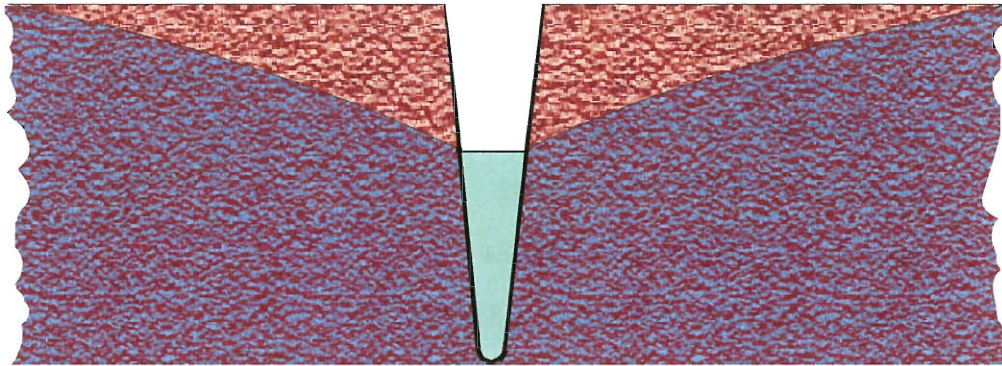
$K_0$  en  $K_1$  zijn gemodificeerde Besselfuncties. Ze zijn beschikbaar in Matlab als `besselk(0,2a)` en `besselk(1,2a)`.  $\mu$  geeft de tijd aan die een onttrekking nodig heeft om de helft van zijn eindverlaging te bereiken.

De *derde parameter* (hier parameter 7) is  $c'$  uit (58). Hij heeft iets te maken met het doorlaatvermogen van de aquifer (zie (57)), maar deze interpretatie is weer gebaseerd op het rekenschema van figuur 23. Voorzichtigheid is dus geboden. Interessanter is de  $M_0$ , dat is de *gain* ofwel de stationaire invloed van een onttrekking van  $1 \text{ m}^3/\text{d}$ . Hij wordt gegeven door:

$$M_0 = -c' K_0(2a) \quad (60)$$

De stationaire invloed  $M_0$  van de *well* is niet onder de knop *Statistics* te vinden, maar hij wordt vermeld in het export-rapport; dat is een .csv-bestand dat uitgevoerd wordt door in het kwadrant *Time Series Models* op de knop *Export* te drukken.

#### 4.2.4 Het type River ofwel oppervlaktewaterpeil (polder-, rivier- of zeepeil)



figuur 25: Hydrologisch schema van oppervlaktewater in een homogeen watervoerend pakket volgens Bruggeman.

De impulsrespons van het type *river* heeft dezelfde geohydrologische achtergrond als die van het type *well*. Er wordt echter niet uitgegaan van radiale stroming, maar van 1-dimensionale stroming. Verder wordt de fluctuatie niet veroorzaakt door een zeer kortdurende onttrekking, maar door een zeer kort durende verhoging van een rivierpeil. De staprespons voor deze situatie staat bekend als Polderfunctie. Hij is beschreven in (Bruggeman, 1999) De impulsrespons is per definitie de afgeleide van deze functie naar de tijd, waarbij de inhoud van de puls 1 gesteld is:

$$\theta(t) = \frac{1}{\sqrt{\pi \frac{4kDt^3}{x^2 S}}} \exp\left\{-\frac{x^2 S}{4kDt} - \frac{t}{cS}\right\} \quad (61)$$

Hierin is

- $\theta(t)$  = impulsrespons (1/d)
- $kD$  = doorlaatvermogen van de aquifer (m)
- $c$  = weerstand tegen voeding (d)

$S$  = bergingscoëfficiënt (-)

Menyanthes werkt ook hier met abstracte parameters:

$$a = \frac{x}{2\sqrt{kDc}}, \quad b = \frac{1}{\sqrt{cS}} \quad (62)$$

Voor optimale flexibiliteit is nog een schaalfactor  $c'$  toegevoegd, zodat:

$$\theta(t) = \frac{c'}{\sqrt{\pi \frac{b^2}{a^2} t^3}} \exp\left\{-\frac{a^2}{b^2} \frac{1}{t} - b^2 t\right\} \quad (63)$$

Net als bij de verdeling van Pearson type III is dit een familie van krommen, waaruit Menyanthes tijdens het modelleren het meest geschikte familielid selecteert. In figuur 26 worden een paar familieleden getoond.

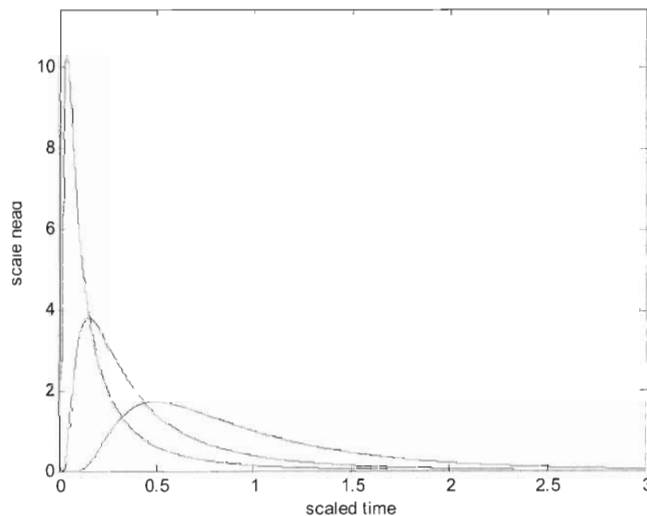
#### Betekenis van de parameters die Menyanthes berekent

Onder de knop *Statistics* in het kwadrant *Time Series Models* levert Menyanthes voor de impulsrespons van het type *river* drie parameters op. Voorbeeld:

- input series: SambeekBoven type: river
- parameter 5 : 0.4825 stdev : 3.1e+003
- parameter 6 : 0.9883 stdev : 6.9e+002
- parameter 7 : 1.578 stdev : 9.7e+003

De *eerste parameter* (hier parameter 5 genoemd) is parameter  $a$  uit (63). Ruwweg geeft hij de afstand van de peilbuis tot de *river* weer, als fractie de invloedsafstand van de rivier. Normaal gesproken zal hij tussen 0 en 1 à 2 liggen. Aangezien de afstand tot de rivier bekend is, kan hieruit *globaal* de invloedsafstand bepaald worden, maar deze verwachting is gebaseerd op het rekenschema van figuur 25. Dit aspect hebben we nog niet grondig bestudeerd.

De *tweede parameter* (hier parameter 6) is parameter  $b$  uit (63). Hij zegt iets over de tijd die een peilverandering nodig heeft om volledig door te werken,



figuur 26: Enkele voorbeelden van impulsresponsfuncties van het type *River*.



dus over het geheugen van het systeem. Een correctere maat voor het geheugen van het systeem is de gemiddelde waarde van de impulsresponsfunctie, die gegeven wordt door:

$$\mu = \frac{a}{b^2} \quad (64)$$

$\mu$  geeft de tijd aan die een plotselinge verlaging van het rivierpeil nodig heeft om ter plaatse van de peilbuis de helft van de eindverlaging te bereiken.

De *derde parameter* (hier parameter 7) is  $c'$  uit (63). Dit is een schaalfactor, die verder geen interessante fysische betekenis heeft. Wel interessant is  $M_0$ , dat is de *gain* ofwel de stationaire invloed van een plotselinge verandering van het rivierpeil 1 m. Hij wordt gegeven door:

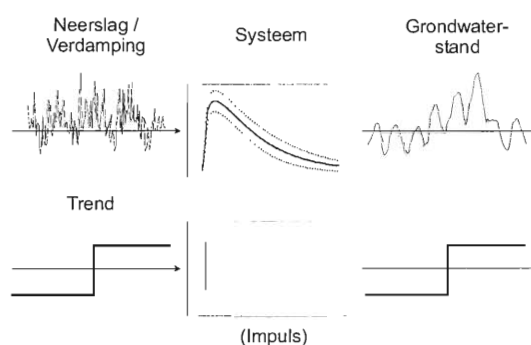
$$M_0 = c' \exp(-2a) \quad (65)$$

De stationaire invloed  $M_0$  van de *river* is niet onder de knop *Statistics* te vinden, maar hij wordt vermeld in het export-rapport; dat is een .csv-bestand dat uitgevoerd wordt door in het kwadrant *Time Series Models* op de knop *Export* te drukken.

#### 4.2.5 De typen Impulse en BOXJEN

In het modelidentificatie-gedeelte van *Menyanthes* kunt u ook kiezen voor de typen Impulse en BOXJEN. Dit zijn twee bijzondere functies:

**Impulse** – De impulsrespons is hier zelf ook een impuls. Dat betekent dat de invoer door het model niet vervormd wordt, maar afgezien van een schalingsfactor, direct opgeteld wordt bij de uitvoer ofwel de simulatie van de grondwaterstand. Het responstype Impulse kun je bijvoorbeeld gebruiken om te schatten of er een trendmatig verloop in de grondwaterstand aanwezig is, zonder dat deze trend nog eens vervormd wordt (zie figuur 27).



figuur 27: Wanneer de impulsrespons een impuls is, komt een invoerreeks zonder vervorming uit het systeem en wordt direct bij de grondwaterstand opgeteld.

**BOXJEN** – Wanneer u dit type kiest gebruikt *Menyanthes* een Box-Jenkins of ARMA transferfunctie voor de invoerreeks. Voordeel daarvan is dat de vorm van de respons-functie niet opgelegd wordt, maar vrijelijk door het model bepaald kan worden, weliswaar afhankelijk van het aantal moving average parameters waarvoor gekozen wordt. Het betreft hier een ietwat

vereenvoudigde, maar voor hydrologische doeleinden vaak doeltreffende versie van een ARMA model, waarbij het aantal AR parameters vastgelegd is op 1. Voor verdere achtergronden over ARMA modellen verwijzen we u naar paragraaf 2.6.3.

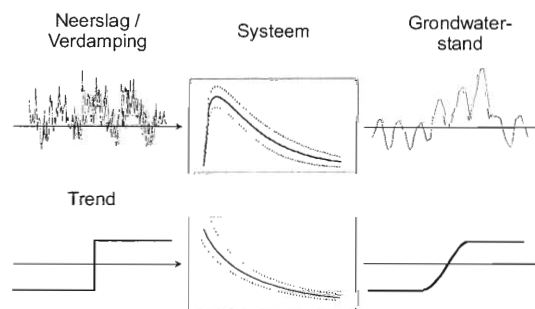
#### 4.3 De effecten van hydrologische ingrepen: interventieanalyse

Een belangrijke toepassing van tijdreeksanalyse is het inschatten van de effecten van hydrologische ingrepen op de grondwaterdynamiek. Dit noemt men interventieanalyse, en komt ook ter sprake in paragraaf 6.3.3. Het kan daarbij gaan om het dempen of graven van een sloot, het kappen of omvormen van bos, het verhogen van het stuwpeil in een sloot, etc. Dit soort maatregelen hebben echter een verschillende uitwerking op de grondwaterstand en/of het grondwatersysteem. In een tijdreeksanalyse komt dit tot uitdrukking in de manier waarop je deze ingrepen meeneemt in de analyse. Qua aanpak kun je drie manieren onderscheiden:

- **Type 1:** een plotselinge verandering van de (actuele) verdamping, door bijvoorbeeld bosomvorming. Een dergelijke verandering kan gemodelleerd met een zogenaamde 'staptrend'. De respons van het systeem op een staptrend in de verdamping is hetzelfde als die van de verdamping zelf. Het enige dat geschat hoeft te worden is een 'schalingsfactor', ofwel de mate waarin de verdamping veranderd is.

➔ **Aanpak:** Voeg een staptrend toe als verklarende reeks in het model identificatie scherm, met type **Evap**

- **Type 2:** een plotselinge verandering van een andere factor, waarvan het exacte verloop in de tijd niet voorhanden is. Te denken valt bijvoorbeeld aan een stuwpeil dat opgezet, maar verder niet bemeten is. In dat geval moet het programma een afzonderlijke respons voor het veranderde stuwpeil inschatten.



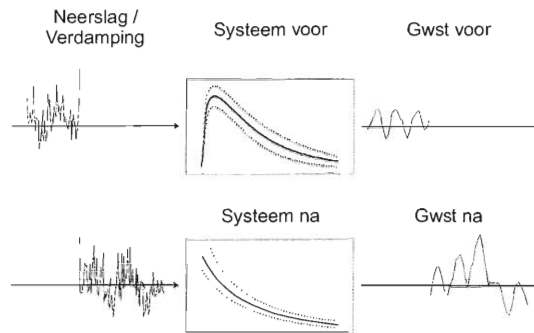
figuur 28: Bij een verandering van een verder onbemeten invloed moet de hele respons van het systeem ingeschat worden.

➔ **Aanpak:** Voeg een staptrend toe als verklarende reeks in het model identificatie scherm, met type **BoxJen, Prec, Well of River**, al naar gelang het type factor dat veranderd is (figuur 28).

- **Type 3:** een plotselinge verandering van het systeem zelf, bijvoorbeeld als er een sloot gegraven of gedempt is. In dat geval zal de drainageweerstand voor de ingreep anders zijn dan erna, en dus is ook de respons van het systeem anders voor en na de ingreep.

→ **Aanpak:** Splits de grondwaterreeks op in een reeks voor en een reeks na de ingreep, en fit hierop afzonderlijke tijdreeksmodellen. Via simulatie kan de grondwaterdynamiek van beide systeemtoestanden vergeleken worden

figuur 29.



figuur 29: Bij een verandering van het systeem zelf moeten voor en na de ingreep aparte modellen gefit worden.

#### 4.4 Trendanalyses

Ook zonder dat men precies weet of de grondwaterstand veranderd is en waarom, kan men op zoek gaan naar dalende of stijgende trends in bepaalde reeksen. Binnen Menyanthes zijn op dit moment twee typen trends voorhanden:

- **Lineaire trend** – Bevindt er zich een lineair dalende of stijgende trend in de grondwaterstand? Voeg een lineaire trend toe via **add** in het modelidentificatie scherm, en kies vervolgens de periode waarover deze trend zich voordoet. Door meerdere lineaire trends te combineren kan bijvoorbeeld ook gezocht worden naar het optreden van een significante trendbreuk, of verandering van de trend.
- **Stap trend** – Naast de toepassingen zoals beschreven onder 4.3 kan een stap trend ook gebruikt worden om in te schatten of onbekende invloeden, d.w.z. andere invloeden dan die reeds in het model aanwezig zijn, een verandering van de gemiddelde grondwaterstand hebben veroorzaakt. Voeg een Step\_trend toe via **add** in het modelidentificatie scherm, en geef vervolgens de datum op waarop de verandering zich voor heeft gedaan.

*Menyanthes* zal automatisch **Impulse** als responsfunctie kiezen. Door deze responsfunctie te kiezen zorgt u ervoor dat de trend behandeld wordt als trend in de grondwaterstand, en niet als trend in een invoerreeks die eerst nog getransformeerd wordt door het systeem (zie figuur 27).

**Let op:** Het effect van andersoortige trends of patronen kan meegenomen worden door deze als invoerreeks in te voeren en mee te nemen in de analyse.

## 4.5 Het ruismodel

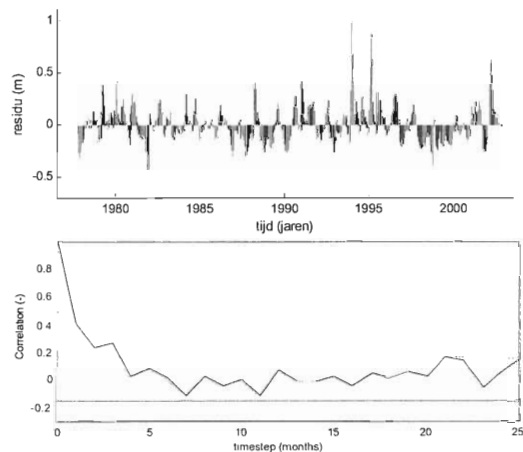
### 4.5.1 De residuen van het transfermodel

Een belangrijke component van tijdreeks- of transferruis-modellen wordt gevormd door het zogenaamde ruismodel. In het transfer-model worden de invoerreeksen 'getransformeerd' met behulp van een transfer- of impulsresponsfunctie, zodanig dat ze gezamenlijk zo goed mogelijk de waargenomen grondwaterstanden verklaren. De invoerreeksen zullen echter nooit de waargenomen grondwaterstanden voor 100% kunnen verklaren, al is het maar omdat waarnemingen per definitie behept zijn met fouten en onnauwkeurigheden. Wanneer je dus simulatie en waarnemingen van elkaar aftrekt, ontstaat vanzelf een nieuwe tijdreeks met modelfouten, de zogenaamde **residue**reeks. Deze residuereeks wordt bij tijdreeksanalyse gemodelleerd met een apart, zogenaamd ruismodel. De redenen hiervoor hebben te maken met het begrip **autocorrelatie**.

### 4.5.2 Wat is autocorrelatie?

In tegenstelling tot wat vaak (bij benadering) het geval is bij gewone regressie, zijn de waarnemingen bij tijdreeksanalyse niet willekeurig verdeeld rond de regressie- of simulatielijn. In de residuen zit nog een bepaalde 'schommeling', of autocorrelatie (figuur 30 boven). Het begrip autocorrelatie wil niet meer zeggen dan dat een parameter gecorreleerd is met zichzelf, maar dan op een ander tijdstip. Zet je bijvoorbeeld steeds de grondwaterstand van vandaag uit tegen die van gisteren, dan verschijnt een lineair verband, vanwege het **geheugen** of de **traagheid** die in het

grondwatersysteem zit. Bereken je op een dergelijke manier voor verschillende tijdstappen de mate van correlatie, en zet je deze twee tegen elkaar uit in een grafiek, dan ontstaat een zogenaamde **autocorrelatie-functie** (figuur 30 onder). In de grafiek zijn ook de 95% betrouwbaarheidsintervallen weergegeven, die aangeven of de correlatie significant afwijkt van 0 (zie voor de wiskundige beschrijving hiervan paragraaf 2.3.2).



figuur 30: Typisch verloop van een residuereeks. De autocorrelatie-functie geeft aan dat de autocorrelatie significant is voor de kleinere tijdstappen.

### 4.5.3 Belang van het ruismodel

De belangrijkste functie van het ruismodel is dat het de autocorrelatie verwijdert uit de residuereeks. Dit proces wordt ook wel 'wissen' genoemd, en het resultaat **witte ruis**. Of de ruis daadwerkelijk wit geworden is kan gecontroleerd worden met behulp van de autocorrelatie-functie. De redenen van het toepassen van een ruismodel zijn onder andere:

- **Leveren van informatie over nauwkeurigheid en onzekerheid.** Deze informatie is nuttig bij het beoordelen van modellen en het nemen van beslissingen.
- **Zuivere schatting van betrouwbaarheidsintervallen.** Bij transferruismodellen vindt de schatting van de parameters en hun covariantiematrix plaats op basis van de ruis, en niet de residuen. Dat is nodig om dat achter de berekening daarvan de aanname schuilt dat de fouten ongecorrleerd en normaal verdeeld zijn. Bevat de ruis nog autocorrelatie, dan wordt de variantie daarvan te hoog ingeschat.
- **Nauwkeuriger voorspelling van missende of toekomstige grondwaterstanden.** Naast het transfermodel, dat een voorspelling van de grondwaterstand doet op basis van de invoerreeksen, kan ook het ruismodel een voorspelling doen van het residu, met behulp van de autocorrelatie. De som van deze voorspellingen is beter dan de voorspelling op basis van het transfermodel alleen.
- **Betere simulatie van de duurlijn en/of de frequentieverdeling van de grondwaterstand.** In lijn met het hierboven geschrevene bevat ook het ruismodel informatie over het gedrag van de grondwaterspiegel, en wel datgene wat niet door de invoerreeksen verklaard kan worden. Met behulp van het ruismodel komen met name de extremen van de duurlijn en/of de frequentieverdeling beter in beeld.
- **Weging van waarnemingen met onregelmatige frequentie.** Waarnemingen die vlak na elkaar zijn gedaan zijn niet onafhankelijk, maar bevatten deels dezelfde informatie over de toestand van het systeem. Het ruismodel neemt deze correlatie weg, en weegt de individuele elementen van de ruisreeks, ofwel de **innovaties**, naar de variantie ervan. Daarbij is de variantie van de innovaties een functie van de tijdstap (zie e.g. (Von Asmuth en Bierkens, in press.; Bierkens et al., 1999a)).

# 5 Het proces van tijdreeksmodellering

## 5.1 Inleiding

Voor uiteenlopende doeleinden willen onderzoekers over een model kunnen beschikken dat het gedrag van een variabele in de tijd adequaat beschrijft. Een model kan bijvoorbeeld worden gebruikt voor een voorspelling naar een of meerdere toekomstige tijdstippen (forecasting). Tijdreeksmodellen kunnen ook worden gebruikt voor simulatie van langere reeksen dan de waargenomen reeks. Verder worden tijdreeksmodellen vaak gebruikt om diverse componenten die van invloed zijn op het dynamische gedrag van een variabele van elkaar te scheiden en de bijdrage van deze componenten aan de variatie te kwantificeren. Dit is bijvoorbeeld het doel van interventieanalyse. In dit hoofdstuk zullen een aantal algemene, praktische aspecten van tijdreeksmodellering worden behandeld.

## 5.2 In den beginne: de gegevens

### 5.2.1 *Opschonen van de dataset*

Voor je aan de slag gaat met de tijdreeksanalyse is het verstandig om allereerst de beschikbare grondwaterstandsreeksen individueel te bekijken en door te nemen. Zogenaamde 'outliers' of afwijkende (aantoonbaar foute) waarnemingen dienen hierbij uit de reeksen verwijderd te worden, en ook al die reeksen die niet bruikbaar en/of kwalitatief onvoldoende zijn. Een waarneming kan als outlier worden bestempeld als:

- Deze duidelijk ver buiten het normale fluctuatiepatroon valt (dat is doorgaans al visueel goed te beoordelen);
- Als het bovendien om een eenmalige waarneming gaat. Een werkelijk sterk afwijkende grondwaterstand zou vertraagd terug moeten keren naar het normale patroon, zodat je ook tussenliggende waarnemingen verwacht;
- Als de uitschieter bovendien niet te vinden is in de andere filters van de peilbuis, of de omgeving.

Meestal zijn niet alle reeksen in een dataset bruikbaar voor een bepaalde studie. De reeksen moeten namelijk voldoende informatie bevatten en (in ieder geval een deel voor en tijdens) de periode van eventueel te analyseren maatregelen omvatten. Voor het al dan niet meenemen van een reeks kun je de volgende criteria hanteren:

- De reeksen moeten een minimale lengte hebben (afhankelijk van de traagheid van het systeem);
- De reeksen moeten een minimale meetfrequentie hebben (wederom afhankelijk van de traagheid van het systeem);
- Bij het analyseren van een ingreep moet de meetreeks zowel een afdoende periode met metingen voor als na de uitgevoerde maatregel bevatten;

### 5.2.2 Aanlooperperioden en beginvoorwaarden

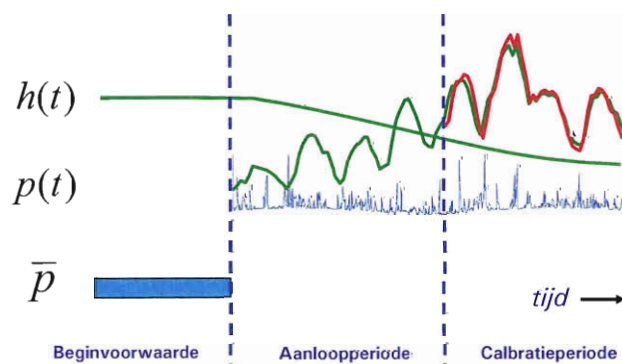
Vanwege het geheugen van een grondwatersysteem moet rekening gehouden worden met de bijdragen van de verklarende reeksen vóór de eerste grondwaterstand. De eerste grondwaterstand is immers de resultante van alles wat er voor die grondwaterstandsdatum aan neerslag, verdamping, onttrekking, e.d., voorgevallen is. Hierbij moet onderscheid worden gemaakt in:

- 1) **De aanlooperperiode** – de periode met gegevens van de verklarende reeksen voordat er grondwaterstandsmetingen zijn.
- 2) **De beginvoorwaarde** – de waarde van een verklarende reeks vóórdat er data van die reeks beschikbaar is.

De lengte van de aanlooperperiode is idealiter zo lang mogelijk. Echter, hoe langer de aanlooperperiode is, des te meer data moet er verwerkt worden en des te langer de rekentijd. Menyanthes kapt daarom verklarende reeksen automatisch af op de volgende manier:

- 1) De kortste reeks is maatgevend. Als er van een bepaalde reeks geen data beschikbaar is voor een bepaalde datum, worden alle reeksen op die datum afgekapt. De data zijn ontoereikend om voor die datum alle afzonderlijke bijdragen aan de grondwaterstand te berekenen.
- 2) Zijn alle verklarende reeksen over een voldoende lange periode vóór de eerste grondwaterstand aanwezig, dan kapt Menyanthes deze af op een lengte van 3000 dagen, oftewel een dikke 8 jaar, voor de eerste grondwaterstandsdatum af. Dit is voor Nederlandse grondwatersystemen ruim voldoende, zeker omdat er ook met een beginvoorwaarde rekening gehouden wordt.

Naast de aanlooperperiode kent Menyanthes dus ook automatisch een beginvoorwaarde toe. Dit om te compenseren voor het feit dat het uiteindelijke, stationaire niveau ten gevolge van een factor als de neerslag zich maar langzaam instelt. Bij factoren die van nature fluctueren rondom een min of meer een constant gemiddelde, zoals neerslag, verdamping en oppervlaktewater neemt Menyanthes de gemiddelde waarde van die factor (Figuur 31). Bij factoren waar dit niet het geval is, zoals grondwaterwinning en trends, is dit nul.



Figuur 31: Rol van de beginvoorwaarde, aanlooperperiode en de calibratieperiode bij het maken van een tijdreeksmodel

**Let op:** Omdat grondwaterwinningen in werkelijkheid niet (altijd) nul geweest zullen zijn voor de eerst beschikbare daadwerkelijk gemeten hoeveelheid, verdient het aanbeveling om zelf een beginvoorwaarde toe te kennen in de invoerdata, volgens het meest waarschijnlijke scenario. Dit om aanloopeffecten te voorkomen.

### 5.2.3 Over het interpoleren en 'sampelen' van de data

Wanneer diver- of andere hoogfrequente data gebruikt worden, of als een ARMA model geschat dient te worden kan het nodig zijn om de frequentie van de waarnemingen te verlagen en/of deze te interpoleren. Dit zijn twee verschillende en op zich zelf staande acties:

- 1) Het reduceren van de meetfrequentie ofwel 'sampelen' van de data betekent dat er per tijdstap slechts één van de metingen wordt gebruikt, namelijk diegene die het dichtst bij de tijdstap zelf ligt. Het tijdstip van de meting en de gemeten waarde worden daarbij niet verandert. De rest van de waarnemingen wordt niet meegenomen in de verdere analyse. Dit leidt tot een verkorting van de rekentijd, maar in beginsel ook tot minder nauwkeurige modeluitkomsten.
- 2) Het interpoleren van de metingen zorgt ervoor dat er een nieuwe reeks verkregen wordt met een waarde op elke, vaste tijdstap, die verkregen wordt door middel van lineaire interpolatie. Dit is een voorwaarde voor het kunnen schatten van ARMA of Box-Jenkins-modellen, maar zorgt ook bij de overige modeltypen (Prec, Evap, Well, River) voor een verkorting van de rekentijd. Anderzijds kunnen door interpolatie de statistische eigenschappen van de oorspronkelijke data 'verminkt' raken.

Omdat het model en de simulaties wiskundig gezien 'continu' blijven, kan het aanpassen en interpoleren van de invoerreeksen al dan niet los of samen met de grondwaterstandsreeksen gebeuren. Als **Sample GWL series** aangevinkt is betekent dit dat **ook** de grondwaterstandsreeksen aangepast en eventueel geïnterpoleerd zullen worden.

## 5.3 Modelidentificatie, calibratie en verificatie

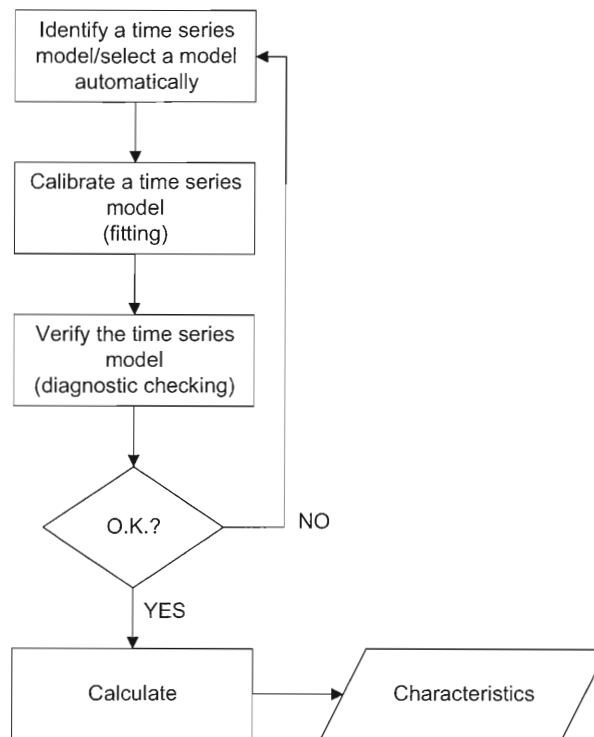
Box en Jenkins (1976) onderscheidden drie fasen in modelconstructie:

1. Identificatie;
2. Estimation (kalibratie);
3. Diagnostic checking (verificatie).

Deze zogenaamde Box-Jenkinsprocedure is schematisch weergegeven in figuur 32.



In de identificatiefase wordt geanalyseerd welk stochastisch model het beste past bij de karakteristieken van de tijdreeks. Identificatie begint met het bestuderen van een tijdreeksgrafiek. Hieruit kan blijken of er een seizoencomponent aanwezig is of een andere trend. Het kan nuttig zijn om een reeks te filteren zodat er een gladder (smooth) beeld ontstaat en trends in het gemiddelde niveau beter zichtbaar worden. Zogenaamde box-and-whisker-grafieken kunnen een beeld geven van het gedrag van de variantie in de tijd. Verder kan worden beoordeeld of er een transformatie van de



figuur 32: Stroomschema van de Box-Jenkinsprocedure voor tijdreeksmodellering

data nodig is, teneinde bij benadering te kunnen voldoen aan de veronderstelling van normaal verdeelde residuen met constante variantie. Een andere hulpmiddel bij identificatie is de sample ACF voor univariate tijdreeksmodellen. Deze grafiek geeft een idee van het type stochastisch proces dat kan worden verondersteld (AR of MA) en de orde van het proces. De identificatiefase bij TFN-modellering is samengevat in paragraaf 2.8.3. In de calibratiefase worden de modelparameters geschat met behulp van een optimalisatiealgoritme, op basis van een kleinste kwadraten criterium of een maximum likelihood criterium. In deze syllabus wordt hier niet verder op ingegaan.

In de verificatiefase wordt gecontroleerd of aan de modelveronderstellingen is voldaan. Dit gebeurt voornamelijk aan de hand van analyse van de residuen: een tijdreeksplot en de residuele ACF. In het geval van een transfer-ruismodel wordt ook gekeken naar de CCF voor de invoer en de residuen. Ook zijn er diverse toetsen voor de aanwezigheid van autocorrelatie of cross-correlatie.

#### 5.4 Automatische criteria voor modelselectie

Automatische modelselectie is een objectief en reproduceerbaar alternatief voor de procedure van identificatie en calibratie. Er wordt hierbij een grote set kandidaatmodellen samengesteld die variëren in complexiteit. Vervolgens worden deze modellen gecalibreerd en wordt het selectiecriterium berekend. Dit selectiecriterium is als volgt samengesteld:

*automatisch selectiecriterium = afwijking t.o.v. data + complexiteit.*

Hoe kleiner het criterium, hoe beter. Er bestaan diverse automatische selectiecriteria. De bekendste zijn Akaike's Informatiecriterium (AIC) en Bayes Informatiecriterium (BIC). De laatste selecteert doorgaans minder complexe modellen dan de eerste.

Als eenmaal een model geselecteerd is, dan moet worden gecontroleerd of aan de modelveronderstellingen is gedaan (verificatie). Het is van belang dat de set kandidaatmodellen zorgvuldig wordt samengesteld. Een te grote set kan tot de selectie van een 'gelegenheidsfit' leiden. Enerzijds kan inzicht in het fysische proces kan helpen bij de samenstelling van de set kandidaatmodellen. Anderzijds kan bestaand inzicht een beperking zijn bij het verkrijgen van nieuw inzicht in het proces, als de set kandidaatmodellen al te rigide wordt beperkt.

## 5.5 Beoordelen van modellen vanuit statistisch perspectief

### 5.5.1 Verificatie of validatie?

Bij *verificatie* wordt gecontroleerd of een model beantwoordt aan de modelveronderstellingen. Deze controles worden vaak toegepast op de residuen (ruis) van het tijdreeksmodel. Aan de hand van een tijdreeksplot van de residuen en de residuele ACF wordt gecontroleerd of de residuen een witte-ruisproces vormen, dat wil zeggen onafhankelijk van elkaar zijn en een constante variatie in de tijd hebben. Dit soort controles kunnen zowel door visuele beoordeling plaatsvinden als door meer objectieve statistische toetsen. Bij transfer-ruismodellen wordt ook gecontroleerd of de residuen gecorreleerd zijn met de verklarende reeks. Bij verificatie wordt ook beoordeeld of het model goed bij de waarnemingen past: *goodness of fit*. Een absolute maat voor *goodness of fit* is de residuele variantie,  $\hat{\sigma}_e^2$  (het dakje geeft aan dat deze geschat is uit de gefitte residuen). Bij een transfer-ruismodel zal men ook geïnteresseerd zijn in het verschil tussen de transfer-component en de waarnemingen, dat gekwantificeerd kan worden met  $\hat{\sigma}_N^2$ . Een relatieve maat voor de *goodness of fit* is het percentage verklaarde variantie: het percentage van de totale variatie van de doelvariabele dat kan worden verklaard met het transfer-model. Voor het transfer-ruismodel in vergelijking (27) kan het percentage verklaarde variantie worden berekend met:

$$\text{perc.verkl.variantie} = \left( \hat{\sigma}_Z^2 - \frac{\hat{\sigma}_N^2}{\hat{\sigma}_Z^2} \right) \times 100. \quad (66)$$

*Verificatie* kan worden gezien als een wetenschappelijke beoordeling van een model. Gebruikers van een model zullen echter geïnteresseerd zijn in een praktische beoordeling: doet het model wat het moet doen? Het zou bijvoorbeeld kunnen dat een model niet voldoet aan de modelveronderstellingen, maar dat dit geen ongewenste consequenties heeft bij het gebruik van het model. Het omgekeerde kan ook het geval zijn: een model voldoet theoretisch perfect en past prima bij de data, maar blijkt slecht te kunnen voorspellen omdat het 'overfit' is.

*Validatie* geeft antwoord op de vraag of een model in de praktijk bruikbaar is. Bij validatie worden de resultaten van voorspelling of simulatie vergeleken met de waarnemingen. Validatie kan als alternatief voor verificatie (diagnostic checking) worden gehanteerd. Als de afwijkingen t.o.v. de waarnemingen kleiner zijn dan een van tevoren bepaalde grens dan kan het model worden geaccepteerd voor verdere toepassing. Validatie biedt de mogelijkheid om de selectie van een model af te stemmen op het uiteindelijke doel van het model (simulatie of voorspelling, schatten van bepaalde karakteristieken). Als het doel van het model bijvoorbeeld is om extreme situaties te simuleren dan zou juist op de gesimuleerde minima en de maxima kunnen worden gevalideerd. Er moet natuurlijk wel een onafhankelijke validatieset beschikbaar zijn. Bij cross-validatie worden twee subsets van een reeks onderscheiden: de kalibratieset en de validatieset. Kalibratie en validatie worden vervolgens net zolang herhaald tot alle data voor beide zijn gebruikt. Belangrijke validatiematen zijn:

1. de gemiddelde fout als maat voor de systematische afwijking of 'bias':

$$ME = \frac{1}{n} \sum_{i=1}^n e_i \quad (67)$$

met  $n$  het aantal waarnemingen en  $e_i$  het verschil tussen waarneming en voorspelling;

2. de standaardafwijking van de fouten als maat voor de toevallige afwijking of inverse precisie:

$$SDE = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (e_i - ME)^2} \quad (68)$$

hoe kleiner de  $SDE$  hoe preciezer het model;

3. de root mean squared error als maat voor de totale afwijking tussen model en waarnemingen, of inverse nauwkeurigheid:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (69)$$

hoe kleiner de  $RMSE$  hoe nauwkeuriger het model.

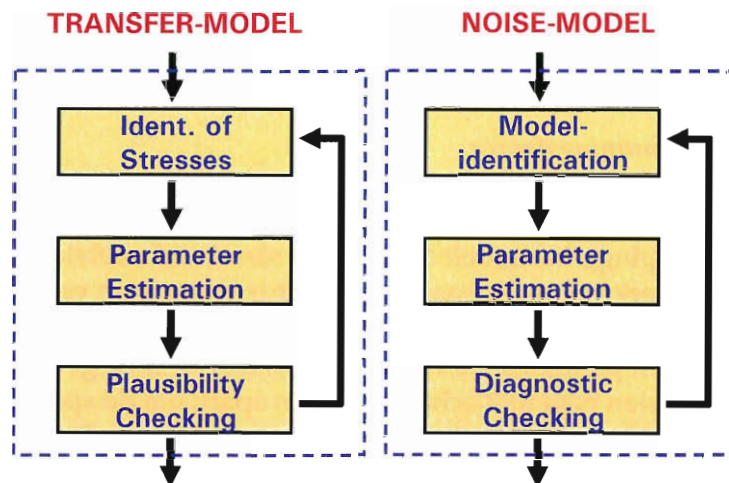
## 5.6 Beoordelen van modellen vanuit fysisch-hydrologisch perspectief

In het voorgaande zijn de validatiematen aan de orde gekomen waarmee je kunt kwantificeren hoe goed het model in statistische termen presteert, en de verificatiematen waarmee je kunt toetsen of al dan niet aan de statistische uitgangspunten is voldaan (witte, normaal verdeelde ruis, stationariteit, cross-correlatie meer tussen invoer en residuen, etc.). Naast deze statistische beoordeling moeten de resultaten ook in fysisch-hydrologisch opzicht beoordeeld en naar waarde geschat worden. Met het aantonen van een

statistisch significante relatie tussen twee factoren, staat daarmee nog niet vast of deze relatie ook daadwerkelijk causaal. Je kunt slechts waarschijnlijk maken dat de gevonden relatie causaal is c.q. ook in fysisch-hydrologische zin bestaat. We gebruiken in dit kader de term *plausibiliteit*, en hiermee ontstaat een aparte, fysisch gebaseerde procedure voor het transfer- of deterministische gedeelte van het model, naast de modelidentificatieprocedure conform Box-Jenkins voor het ruisgedeelte (figuur 33). Voor het hydrologisch interpreteren van de resultaten kun je de volgende middelen gebruiken:

- De betekenis en fysisch plausibele range van parameterwaarden, met in het bijzonder:
  - De drainagebasis
  - De verdampingsfactor
  - De stationaire invloed van de verschillende factoren
  - De traagheid van de verschillende factoren
- Ruimtelijke samenhang van de resultaten

In hoofdstuk 4 zijn de betekenis en mogelijke range van parameterwaarden en van de verschillende typen invloedsfactoren al aan de orde gekomen, en hoe daar de stationaire invloed en traagheid van af te leiden. We bespreken hieronder de drainagebasis, verdampingsfactor en ruimtelijke samenhang van de resultaten.



figuur 33: Alternatieve procedure voor TFN-modellering, fysisch gebaseerd voor het transfergedeelte en conform Box-Jenkins voor het ruisgedeelte

#### 5.6.1 De drainagebasis

De drainagebasis (*dng base, local drainage base*) is het niveau dat de grondwaterspiegel zou bereiken als het hele afstromingsgebied lange tijd verstoken zou blijven van de invloeden die men bij de tijdreeksmodellering betreft. Het is dus geen objectief niveau. Als, er bijvoorbeeld, alleen gemodelleerd wordt met neerslag en verdamping, dan is de drainagebasis het niveau dat uiteindelijk bereikt wordt als het heel lang niet regent of verdampt. Is er een grondwaterwinning in de buurt die niet meegemodelleerd is, dan ziet men in de drainagebasis de verlagingskegel terug. Neemt men de grondwaterwinning wél mee, dan vertoont de drainagebasis géén verlagingskegel. Vaak zal de drainagebasis het niveau zijn dat bepaald

wordt door het gemiddelde peil van de ontwateringsmiddelen. U vindt de drainagebasis onder de knop *Statistics* in het kwadrant *Time Series Models*.  
Voorbeeld:

- Local drainage base: 0.788 m

U treft de drainagebasis (*dng base*) ook rechts van het preview-venster aan, in het kwadrant *Time Series Models*. Voor niet-lineaire modellen worden op die plaats twee getallen vermeld. Niet-lineaire modellen kennen namelijk (momenteel) twee drainagebases. De tweede drainagebasis vindt u ook onder de knop *Statistics* in het kwadrant *Time Series Models*, onder de naam *thold1*.  
Voorbeeld:

- parameter 6 (thold1): 0.9931 stdev : 0.036

Doorgaans is dit het niveau van greppels of buisdrains. Omdat men meestal wel een idee heeft van de ligging van de drainagebasis, is de drainagebasis één van de getallen waaraan snel te zien is of een tijdreeksmodel verworpen moet worden. Maar let op: niet-lineaire modellen hebben vaak moeite om de locale drainagebasis te schatten. Een onwaarschijnlijk getal betekent dan nog niet direct dat er iets mis is met het model. De drainagebasis correleert sterk negatief met de *gain* (die in de volgende paragraaf behandeld wordt): een te lage drainagebasis gaat gepaard met een te hoge gain, en omgekeerd.

### 5.6.2 *De verdampingsfactor*

Waar in het verleden de meeste tijdreeksanalisten en programmatuur gebruik maakten van het neerslagoverschot als invoerreeks (neerslag – potentiële verdamping), behandelt *Menyanthes* standaard neerslag en verdamping als aparte reeksen, weliswaar met dezelfde respons en vermenigvuldigt met een reductiefactor voor de verdamping (de zogenaamde verdampingsfactor). Deze is in paragraaf 4.2.2 natuurlijk al aan de orde gekomen, maar we behandelen hem hier echter nog eens apart, om de speciale rol te benadrukken die deze kan vervullen in de beoordeling van de plausibiliteit van een model. Een belangrijke reden hiervan is dat de verdampingsreeks een sterk seizoensverloop heeft, terwijl dat bij de neerslag niet of nauwelijks het geval heeft. Bovendien hebben veel andere natuurlijke factoren ook een sterk seizoensmatig karakter (beregening, drinkwaterontrekking, etc.). Door neerslag en verdamping apart te houden kun je aan de verdampingsfactor direct zien of er problemen zijn met de correlatie tussen dit soort factoren.

### 5.6.3 *Ruimtelijke samenhang van de resultaten*

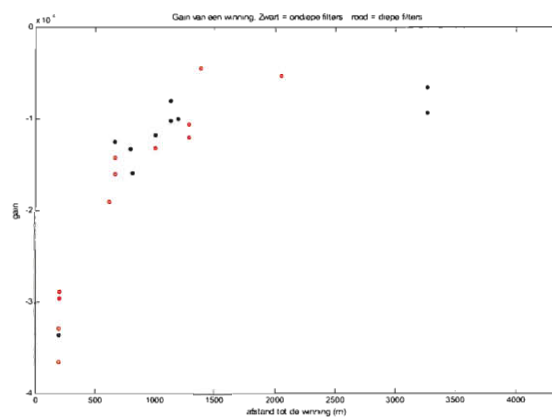
Door de geautomatiseerde verwerking die binnen *Menyanthes* mogelijk is maakt het qua inspanning en kosten niet veel uit hoeveel grondwaterstandreeksen je tegelijk analyseert. Hoewel tijdreeksmodellen in principe ‘puntmodellen’ zijn, leveren dergelijke resultaten bij het dichte netwerk van peilbuizen dat we in Nederland kennen toch een ruimtelijk beeld. Dit geeft vaak een verfijnder beeld op dan de resultaten van een individuele reeks: terwijl de resultaten van een individuele reeks op zich best plausibel kunnen zijn, kunnen de resultaten in relatie tot die in de nabije omgeving er toch uitspringen als outliers.

Voor grondwateronttrekkingen is bijvoorbeeld de gain op te vatten als de stationaire verlaging van de grondwaterstand in meters per eenheid onttrokken debiet (in  $1 \text{ m}^3/\text{d}$ ). De gain van een grondwaterwinning is afhankelijk van de positie van het peilbuisfilter dat geanalyseerd wordt, en van de geohydrologische eigenschappen van het grondwatersysteem. Enkele factoren die hierop betrekking hebben zijn:

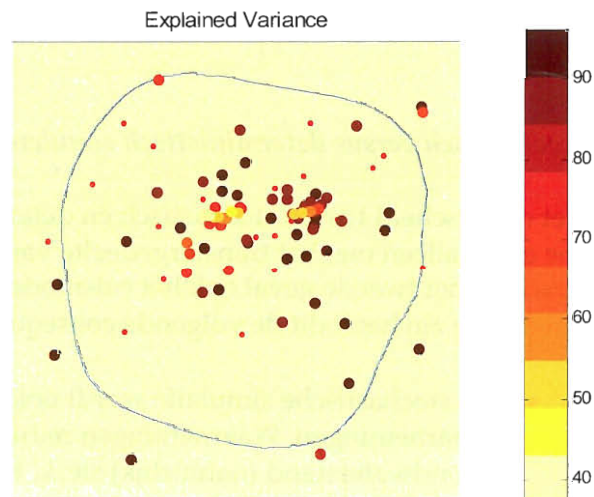
- Het al dan niet freatisch zijn van het grondwater
- De afstand van de winning tot de peilbuis
- De afstand van de peilbuis tot een voedende grens
- De weerstand van het watervoerende pakket waaruit onttrokken wordt
- De weerstand van eventuele afsluitende lagen tussen winning en peilbuis-filter

Aan de ene kant geven de resultaten van een tijdreeksanalyse via de gain en de andere parameters dus informatie over de eigenschappen van het grondwatersysteem. Aan de andere kant kan informatie over het grondwatersysteem helpen om plausibele van niet-plausibele relaties scheiden. Wanneer men bijvoorbeeld de afstand van een winning tot de geanalyseerde peilbuizen uitzet tegen de gain daarvan, zou een onttrekkingskegel zichtbaar moeten worden als de omgeving geohydrologisch min of meer homogeen is (figuur 34).

Een andere optie die veel inzicht op kan leveren is het plotten van de resultaten (parameters, validatiematen, etc.) op de locatie van de geanalyseerde filters. De plot in figuur 35 geeft een ruimtelijk overzicht van de verklaarde varianties op de plaats van de geanalyseerde reeksen van een



figuur 34: De afstand van winning tot peilbuis uitgezet tegen de gain geeft een onttrekkingskegel als het grondwatersysteem ruimtelijk homogeen is.



figuur 35: Plot van de verklaarde variantie op de plaats van de geanalyseerde reeksen. Duidelijk zichtbaar zijn twee 'strengen' met minde goede resultaten.

waterwingebied in Duitsland. Duidelijk zichtbaar zijn twee 'strengen' met minder goede resultaten. In dit geval was de verklaring daarvan snel gevonden: als verklarende variabele was het verloop van de totale onttrekking meegenomen, terwijl de afzonderlijke winputten, op de plek van de twee strengen, individueel worden aangestuurd.

## 5.7 Het simuleren van grondwaterstanden

### 5.7.1 Toepassingen

Wanneer u een tijdreeksmodel geschat heeft, ligt in feite de relatie tussen invoerreeksen en grondwaterstanden vast. Op basis hiervan kunt u de grondwaterstand simuleren in periodes dat deze niet gemeten is, zolang over deze periode de gegevens van de invoerreeksen maar voorhanden zijn. De mogelijkheid om grondwaterstanden te simuleren kent verschillende toepassingen:

- **Opschonen van gemeten reeksen** – Door middel van simulatie kunt u gaten in een reeks op vullen, reeksen verlengen en eventuele uitbijters detecteren en vervangen.
- **Vaststellen van de gevolgen van een ingreep** – In sommige gevallen moet bij een hydrologische ingreep, de situatie voor en na de ingreep apart worden gemodelleerd (zie paragraaf 4.3). Door voor beide periodes de grondwaterstand te simuleren kan een goede vergelijking gemaakt worden van de toestand voor en na de ingreep.
- **Scenario berekeningen** – Wanneer u geïnteresseerd bent in de grondwaterstand bij een bepaald scenario, bijvoorbeeld een bepaald onttrekkingsregime, kunt u de gemeten waarden behorende bij die verklarende reeks vervangen door die van het gekozen scenario, en op basis daarvan simuleren. Bij tijdreeksmodellen met niet-lineariteit kunt u bovendien de lokale drainage middelen aan- of uitzetten.

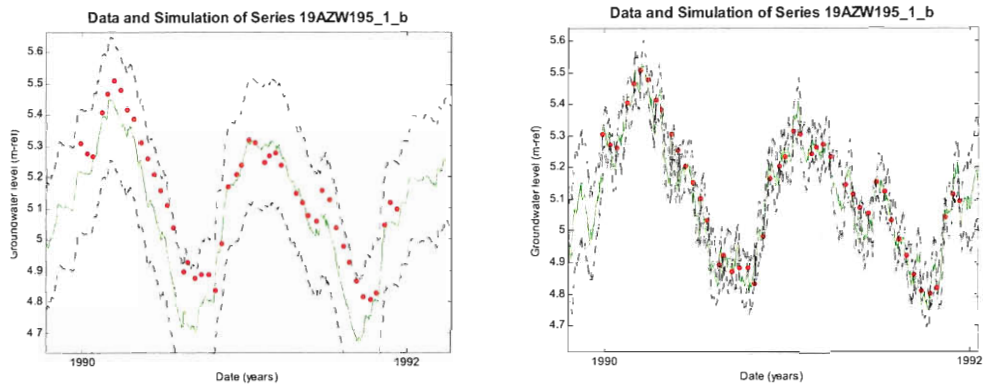
### 5.7.2 Stochastisch versus deterministisch simuleren

Het onderscheid tussen stochastisch en deterministisch simuleren is dat in het ene geval alleen met het transfergedeelte van het model wordt gesimuleerd, terwijl in het tweede geval ook het ruismodel daarbij in wordt gezet. In praktische zin heeft dit de volgende consequenties:

- Bij stochastische simulatie wordt ook rekening gehouden met de waarnemingen. Waarnemingen reduceren de onzekerheid over de grondwaterstand (natuurlijk) sterk. Hoe dichter je in de tijdreeks nadert naar het tijdstip waarop een waarneming beschikbaar is, hoe kleiner de onzekerheid op dat tijdstip wordt, wat tot uitdrukking komt in een variabel voorspellingsinterval (figuur 36).



- Een meer of minder groot deel van de variatie in de grondwaterspiegelfluctuaties wordt verklaard door het ruismodel. Dit betekent dat bij een deterministische simulatie, de variantie van de fluctuaties wordt onderschat. Om een goede schatting te maken van de kansen op extremen, herhalingskansen, frequentieverdelingen etc. van de grondwaterstand is het van belang het ruismodel mee te nemen.



*figuur 36: Verschil tussen deterministische en stochastische simulatie. In het tweede geval zijn de betrouwbaarheidsintervallen variabel, en kleiner als er metingen beschikbaar zijn.*





## 6 Monitoring en tijdreeksmodellering

### 6.1 Inleiding

In dit hoofdstuk gaan wij in op een aantal aspecten die betrekking hebben op de monitoringaspecten van tijdreeksanalyse. In dit hoofdstuk komen niet alleen aspecten van waarnemen in de tijd, maar ook van waarnemen in ruimte en ruimte en tijd aan de orde. Paragraaf 6.2 geeft een korte introductie in monitoring, gebaseerd op het handboek 'Sampling for Natural Resource Monitoring' van J.J. de Gruijter, D.J. Brus, M.F.P. Bierkens en M. Knotters, dat in 2006 zal verschijnen. Dit boek gaat in op de 'sampling-' of steekproefaspecten van monitoring. Het boek beoogt praktische ondersteuning te bieden bij beslissingen over de monitoringstrategie die moet worden gevolgd. Een belangrijk uitgangspunt hierbij is de planning van het gehele monitoringtraject in één keer, inclusief statistische verwerking van de gegevens. 'Begin aan het einde, en redeneer dan terug'. Wij staan stil bij de verschillende doelstellingen van monitoring, het ontwerpen van een monitoringstrategie en geven een kort overzicht van strategieën in ruimte, tijd en ruimte-tijd. In paragraaf 6.3 gaan wij in op het waarnemen in de tijd ten behoeve van tijdreeksmodellering. Wij gaan in op de frequentie en de lengte die tijdreeksen moeten hebben, om een proces nauwkeurig te kunnen beschrijven met een tijdreeksmodel. Vervolgens gaan wij in op het benodigd aantal waarnemingen om een gemiddelde met een bepaalde nauwkeurigheid te schatten. Ten slotte bespreken wij hoe het aantal waarnemingen kan worden bepaald dat nodig is om met een bepaalde betrouwbaarheid te beoordelen of er een trend aanwezig is.

### 6.2 Monitoring in ruimte, tijd en ruimte en tijd

#### 6.2.1 *Waarom monitoren?*

Om uiteenlopende redenen worden variabelen in de natuurlijke omgeving gemonitord, en op basis van monitoring worden allerlei beslissingen genomen. De Gruijter e.a. (2006) onderscheiden de volgende categorieën:

1. Status monitoring. Hierbij wordt de toestand van iets bepaald en gevolgd in de tijd. Deze informatie kan nodig zijn voor het ontwikkelen van beleid, voor wetenschappelijk onderzoek, voor regelgeving etc. Voorbeelden zijn het monitoren van het klimaat, de uitstoot van broeikasgassen etc.;
2. Trend- of effectmonitoring. Dit heeft tot doel te beoordelen of een gebeurtenis of maatregel effect heeft op de natuurlijke omgeving. Voorbeelden zijn het monitoren van het effect van drinkwaterwinning op de grondwaterstand, het effect van verdrogingsbestrijding op de grondwaterstand in een natuurgebied etc.;
3. Compliance of regulatory monitoring. Deze vorm van monitoring heeft tot doel te beoordelen of iets aan een bepaalde norm voldoet,

bijvoorbeeld of de grond- en oppervlaktewaterkwaliteit beantwoordt aan de normen van de Europese Kaderrichtlijn Water.

Als vierde categorie kan hier wellicht aan worden toegevoegd monitoring dat als doel heeft iets op te sporen, bijvoorbeeld een puntbron van verontreiniging.

#### 6.2.2 *Het ontwerpen van een monitoringstrategie*

Bij monitoring kunnen in chronologie de volgende activiteiten worden onderscheiden:

- planning van het veldwerk, inclusief beslissingen over de hoeveelheid monsters, de locaties en tijdstippen waarop ze worden genomen, gegeven het doel van de monitoring en de financiële en praktische randvoorwaarden;
- uitvoering van het veldwerk;
- laboratoriumbepalingen;
- opslaan van gegevens;
- (statistische) verwerking van gegevens;
- rapportage.

Ondanks de logische volgorde vormt de bovenstaande lijst geen goede basis voor de planning van een monitoringproject, omdat de keten van activiteiten inclusief de statistische verwerking niet als geheel wordt gepland. De Gruijter e.a. (2006) adviseren niet alleen de dataverzameling te plannen, maar de gehele keten van activiteiten inclusief statistische verwerking. Zij stellen voor een schema te ontwikkelen met als principe 'begin aan het eind, en redeneer dan terug'.

Het zou te ver voeren om in deze handleiding de ontwikkeling van een schema voor monitoring uitvoerig te behandelen. Bij enkele belangrijke aspecten willen wij hier echter graag stilstaan. (De Gruijter *et al.*, 2006) benadrukken het onderscheid tussen strategieën die op modellen zijn gebaseerd en strategieën die op steekproefontwerpen zijn gebaseerd (model-based en design-based strategieën). De keuze tussen beide strategieën hangt af van de informatie die gewenst is. Als bijvoorbeeld het doel van monitoring is de fluctuatie van de grondwaterstand in het heersende hydrologische regime te karakteriseren, dan ligt een modelgebaseerde strategie voor de hand. Is echter het doel te controleren of de grondwaterkwaliteit in een bepaald jaar aan een bepaalde norm voldeed, dan is een ontwerpgebaseerde strategie meer op zijn plaats. Is er een kaartje met gebiedsdekkende voorspellingen van bijvoorbeeld de gemiddeld hoogste of laagste grondwaterstand gewenst, dan wordt vaak voor een modelgebaseerde strategie gekozen. Moet bijvoorbeeld de oppervlaktefractie van een gebied worden geschat waarin de Europese nitraatnorm wordt overschreden, dan is een ontwerpgebaseerde strategie een goede keus.

Naast onderscheid tussen ontwerp- en modelgebaseerde strategieën kan onderscheid worden gemaakt tussen de types gewenste informatie: 'globale

grootheden' of 'locale grootheden'. Voorbeelden van globale grootheden zijn een gemiddelde van een gebied of een periode, een totaal of een fractie. Voorbeelden van locale grootheden zijn de concentratie van een bepaalde stof op een bepaalde locatie of op een bepaald tijdstip.

### **6.2.3 *Strategieën voor monitoring in ruimte***

Er zijn veel strategieën beschikbaar om steekproeven in de ruimte uit te voeren, waarbij onderscheid kan worden gemaakt tussen 'ontwerpgebaseerde' en 'modelgebaseerde' strategieën ten behoeve van het schatten of voorspellen van globale en locale grootheden. Ontwerpgebaseerde strategieën worden vooral toegepast om globale grootheden te schatten, zoals ruimtelijke gemiddeldes en oppervlaktefracties. Bekende voorbeelden van ontwerpgebaseerde strategieën zijn de enkelvoudige aselechte steekproef en de gestratificeerde aselechte steekproef. Bij de enkelvoudige aselechte steekproef heeft iedere locatie evenveel kans geloot te worden. Bij een gestratificeerde aselechte steekproef ook, maar hier wordt binnen zogenaamde strata of deelgebieden geloot. Op deze manier kunnen uitspraken per deelgebied worden gedaan, kan de efficiëntie van de steekproef worden verhoogd om dat de variantie wordt gereduceerd, en kan meer geografische spreiding worden verkregen. Een meer geavanceerde steekproef is de tweetrapssteekproef. Hierbij worden eerst primaire eenheden geloot, bijvoorbeeld deelgebieden, en daarna binnen deze primaire eenheden de secundaire eenheden, bijvoorbeeld monsterlocaties. In de eenvoudigste vorm worden de primaire en secundaire eenheden geloot met enkelvoudige aselechte steekproeven. Een tweetrapssteekproef wordt toegepast als er veel kosten zijn gemoeid met het reizen naar de monsterlocaties. Het is bij steekproeven mogelijk om gebruik te maken van hulpinformatie, bijvoorbeeld informatie uit bestaande meetnetten. We spreken dan van regressieschatters.

Er is nog een groot aantal andere steekproeven, die hier niet behandeld worden. De Gruijter e.a. (2006) geven een beslisboom om een keuze te maken uit de verschillende steekproefontwerpen.

Modelgebaseerde methoden worden gebruikt bij zowel het schatten van globale grootheden zoals oppervlaktefracties en gebiedsgemiddelden als bij het ruimtelijk voorspellen of simuleren van waarden op puntlocaties. Hiervoor zijn vele methoden uit de geostatistiek beschikbaar. Er wordt gebruik gemaakt van een model dat de ruimtelijke samenhang beschrijft (een variogram of co-variogram).

### **6.2.4 *Strategieën voor monitoring in tijd***

Op enkele uitzonderingen na kunnen steekproefstrategieën die in de ruimte worden toegepast ook in de tijd worden uitgevoerd. Er zijn wel enkele praktische punten waar rekening mee gehouden moet worden. De voorbereiding en planning van het veldwerk is heel belangrijk. Als tijdstippen afvallen door bijvoorbeeld verlof of feestdagen, dan moeten vóórdat de meetperiode start nieuwe tijdstippen worden geloot. Gebeurt dit tijdens de meetperiode, dan zou er een oververtegenwoordiging van

waarnemingen aan het einde van de periode ontstaan. Door ziekte e.d. kunnen ook tijdstippen afvallen. Daarom is het verstandig het aantal tijdstippen ruim te kiezen.

#### **6.2.5 *Strategieën voor monitoring in ruimte-tijd***

Ontwerpgebaseerde steekproeven voor ruimte-tijd kunnen worden verdeeld in statische, synchrone en roterende. Bij statische steekproeven wordt eerst een aantal locaties geselecteerd. Per locatie worden tijdstippen geselecteerd waarop wordt bemonsterd. Bij synchrone steekproeven wordt eerst een aantal tijdstippen geselecteerd en vervolgens per tijdstip een aantal locaties geloot. Als een aantal van deze locaties gelijk is aan die van een eerder tijdstip dan spreken we over een roterend design. Als alle tijdstippen voor alle locaties gelijk zijn (en natuurlijk andersom: als voor alle locaties op dezelfde tijdstippen wordt bemonsterd) spreken we van een statisch-synchroon design.

De strategieën voor monitoring in ruimte-tijd kunnen ontwerpgebaseerd, modelgebaseerd of een combinatie van beide zijn. Een voorbeeld van een ontwerpgebaseerde benadering is een strategie voor compliance monitoring van de gemiddelde concentraties N-totaal en P-totaal in het oppervlaktewater op melkveebedrijven in het zomerhalfjaar. Het gaat hier om een ruimte-tijdgemiddelde; dit gemiddelde mag niet een bepaalde norm van de Europese Kaderrichtlijn Water overschrijden. Er wordt getoetst of de norm overschreden wordt (Knotters, 2005). Bij modelgebaseerde benaderingen worden methoden zoals space-time kriging en Kalmanfiltering toegepast. Een praktische toepassing is bijvoorbeeld het in ruimte en tijd volgen van de verspreiding van een verontreiniging in het grondwater.

### **6.3 Monitoring voor tijdreeksmodellen**

#### **6.3.1 *Het schatten van modelparameters***

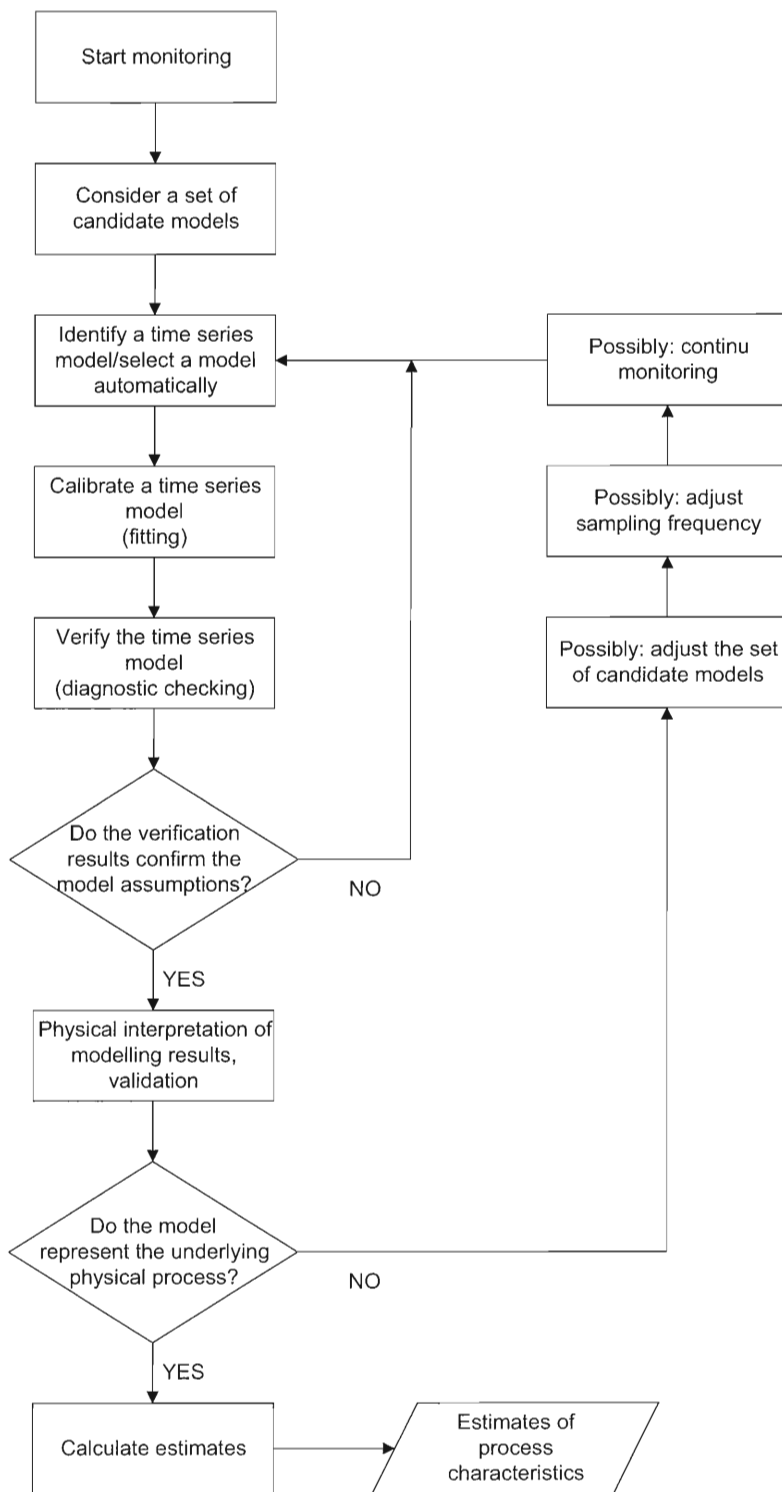
In deze paragraaf gaan wij in op het ontwikkelen van monitoringstrategieën voor het verkrijgen van parameters die een algemene beschrijving geven van de temporele variatie van een fenomeen. Deze parameters worden ook wel systeemparameters genoemd; als monitoring gericht is op het schatten van systeemparameters dan spreken we van status-monitoring. Statistische gevolgtrekking (Eng. inference) is mogelijk door te veronderstellen dat de temporele variatie het resultaat is van een stochastisch proces, dat kan worden beschreven met een tijdreeksmodel. De belangrijkste vraag in deze paragraaf is daarom: Hoe moeten waarnemingen in de tijd zijn verdeeld, om tijdreeksmodellen te verkrijgen die de dynamiek van een fysisch fenomeen adequaat beschrijven? Het beantwoorden van deze vraag begint met het maken van veronderstellingen over het stochastische proces, waarvan de meest toegepaste zijn samengevat in hoofdstuk 2.

In paragraaf 2.5 werd al opgemerkt dat stochastische processen data beschrijven. Dat wil zeggen dat de stochastische processen die zijn beschreven in paragraaf 2.6 tot en met 2.8 het dynamische gedrag van

fysische fenomenen beschrijven, gegeven de waarnemingsfrequentie en de lengte van de monitoringperiode. Er moet dus worden gekozen voor een waarnemingsfrequentie en een lengte van de monitoringperiode waarbij het mogelijk is om het dynamische gedrag van een fysisch fenomeen adequaat te beschrijven met een tijdreeksmodel. Met adequaat bedoelen we hier dat het model beantwoordt aan het doel waarvoor het is gemaakt, zoals voorspellingen en simulaties.

Box en Jenkins (1976) geven een procedure van identification, fitting (kalibratie) en diagnostic checking (verificatie) van tijdreeksmodellen (zie paragraaf 5.3). In het stroomschema van figuur 37 is deze procedure samengevat en uitgebreid met beslissingen over de lengte van de monitoringperiode en de waarnemingsfrequentie. Als een alternatief voor modelidentificatie kan automatische modelselectie worden toegepast, met criteria zoals Akaike's Information Criterion (AIC) en Bayes Information Criterion (BIC). De procedures van modelidentificatie of automatische modelselectie, kalibratie en verificatie leveren modellen op die de data adequaat beschrijven. De volgende stap is om te beoordelen in hoeverre het onderliggende fysische proces adequaat is beschreven met het model. Dit kan door middel van een fysische interpretatie van het model. Daarnaast is het raadzaam om indien mogelijk het model te valideren, dat wil zeggen dat met behulp van onafhankelijke waarnemingen wordt getoetst of het model aan zijn doel beantwoordt. Zowel de resultaten van fysische interpretatie als van validatie kunnen niet alleen aanleiding geven tot verbeteringen aan het model, maar ook tot verlenging van de monitoringperiode en bijstelling van de waarnemingsfrequentie. Als voorbeeld geven we de volgende twee situaties, zonder daarmee volledig te willen zijn:

1. De autoregressieve parameter van een eerste orde autoregressief model heeft een hoge waarde. De validatie levert grote systematische fouten op. In dit geval bestrijkt de monitoringperiode mogelijk de correlatielengte niet volledig, of, in het geval van een dynamische samenhang tussen twee variabelen is de monitoringperiode korter dan de responstijd. In deze situaties zou de monitoring moeten worden voortgezet zij totdat tenminste de correlatielengte of de responstijd volledig omvat;
2. Alhoewel autoregressieve samenhang werd verwacht op basis van inzichten in het fysische systeem leverde de modellering geen significant autoregressief verband op. De validatieresultaten indiceren grote willekeurige fouten. In dit is het waarnemingsinterval mogelijk langer dan de correlatielengte, of, in het geval van een dynamische samenhang tussen twee variabelen, is het waarnemingsinterval langer dan de responstijd.



figuur 37: Stroomschema voor de procedure van het identificeren, kalibreren en verifiëren van een tijdreeksmodel, uitgebreid met aspecten van monitoring

### 6.3.2 Waarnemingsfrequentie voor het schatten van trends

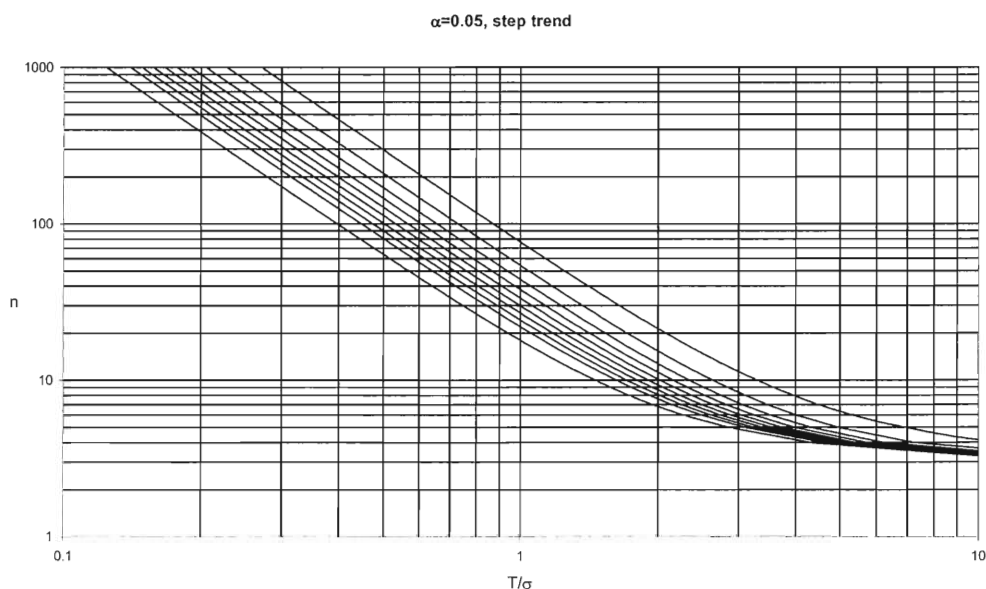
Wij vatten een trend op als een component van een tijdreeksmodel. Als we ons beperken tot equidistante waarnemingen in de tijd bij het opsporen van trends, dan is het van belang om de juiste waarnemingsfrequentie te kiezen. De trend kan zowel een stap- als een lineaire trend betreffen. Het opsporen van een trend is te benaderen als een toetsingsprobleem, met  $H_0$ : een trend is afwezig in het onderliggende proces, en  $H_1$ : een trend is aanwezig.

Als we ten onrechte concluderen dat een trend aanwezig is, dan maken we een fout van de eerste soort. De kans op deze fout mag niet groter zijn dan de onbetrouwbaarheidsdrempel  $\alpha$ . Vaak wordt een kleine onbetrouwbaarheidsdrempel  $\alpha=0.05$  gekozen. Als we ten onrechte aannemen dat er geen trend aanwezig is, dan maken we een fout van de tweede soort. De kans op deze fout,  $\beta$ , mag ook niet te groot zijn.  $1-\beta$  noemen we het onderscheidingsvermogen of de *power* van de toets. Deze hangt af van het aantal waarnemingen, van de gekozen waarde voor  $\alpha$ , van het kleinste verschil dat men nog wil aantonen en van de variantie van de populatie die wordt onderzocht.

Het model voor een staptrend ziet er als volgt uit:

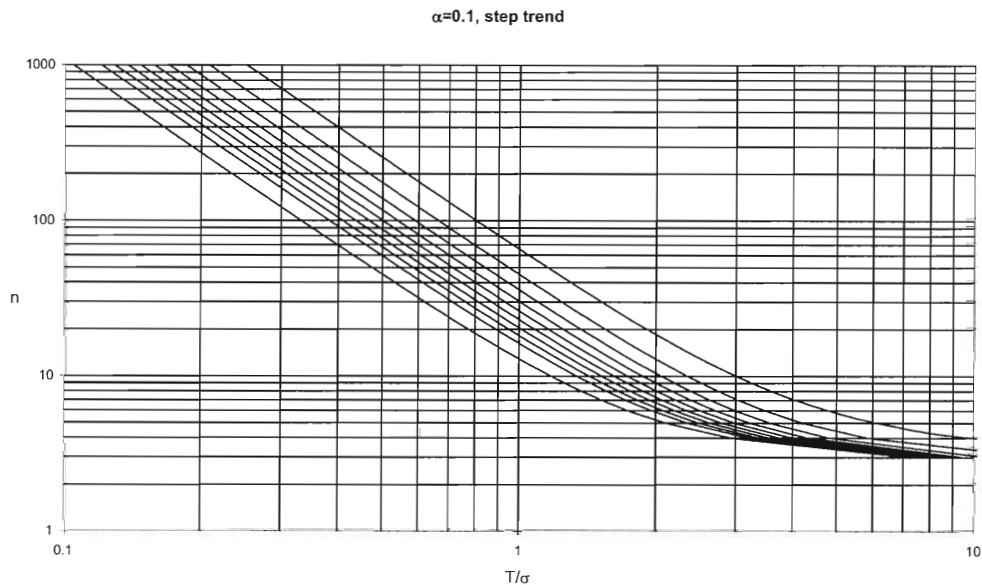
$$\begin{aligned} Z_i &= \mu_1 + \varepsilon_i, \text{ als } t \leq n/2 \\ Z_i &= \mu_2 + \varepsilon_i, \text{ als } t > n/2 \end{aligned} \quad (70)$$

waarin  $\varepsilon_i$  een onafhankelijke en gelijk verdeelde reeks vormt met variantie  $\sigma_\varepsilon^2$ . Met een Student's  $t$ -toets maakt een keuze tussen  $H_0: \mu_1 = \mu_2$  en  $H_1: \mu_1 \neq \mu_2$ . Figuur 38 en 39 geven het benodigd aantal waarnemingen aan bij verschillende waarden voor  $\beta$ , onbetrouwbaarheidsdrempels  $\alpha$  en verhoudingen tussen de trend die nog aangetoond moet worden en standaardafwijking van de populatie (genormaliseerde trends).





figuur 38: Benodigd aantal waarnemingen uitgezet tegen genormaliseerde staptrendvoor  $\sigma=0.05$ ,  $\beta=0.5, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1, 0.05$  en  $0.01$  (van links naar rechts).  $n=n_1+n_2$ ,  $T=|\mu_1-\mu_2|$  en  $\sigma=\sigma_\varepsilon$

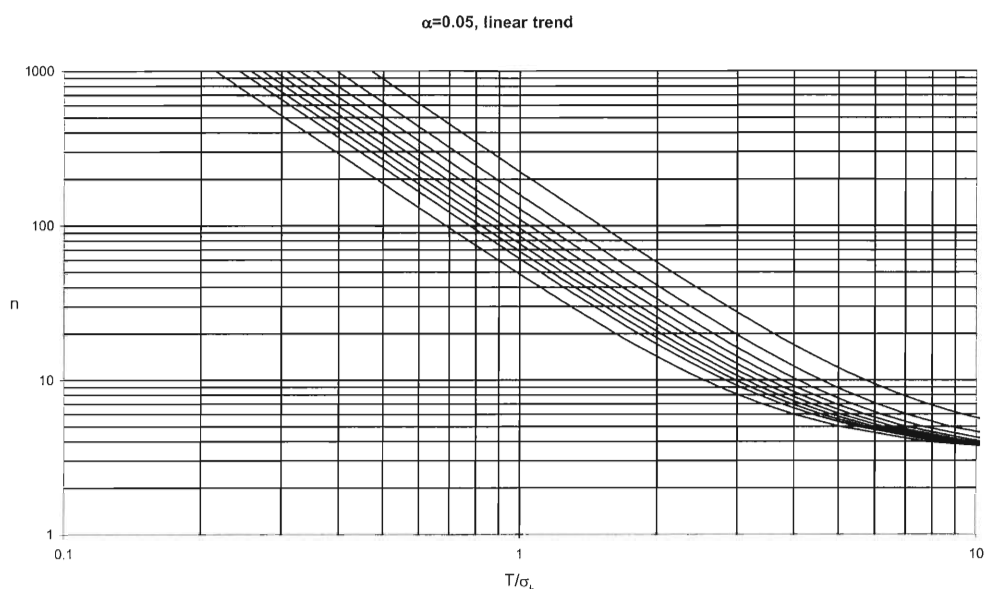


figuur 39: Benodigd aantal waarnemingen uitgezet tegen genormaliseerde staptrendvoor  $\sigma=0.10$ ,  $\beta=0.5, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1, 0.05$  en  $0.01$  (van links naar rechts).  $n=n_1+n_2$ ,  $T=|\mu_1-\mu_2|$  en  $\sigma=\sigma_\varepsilon$

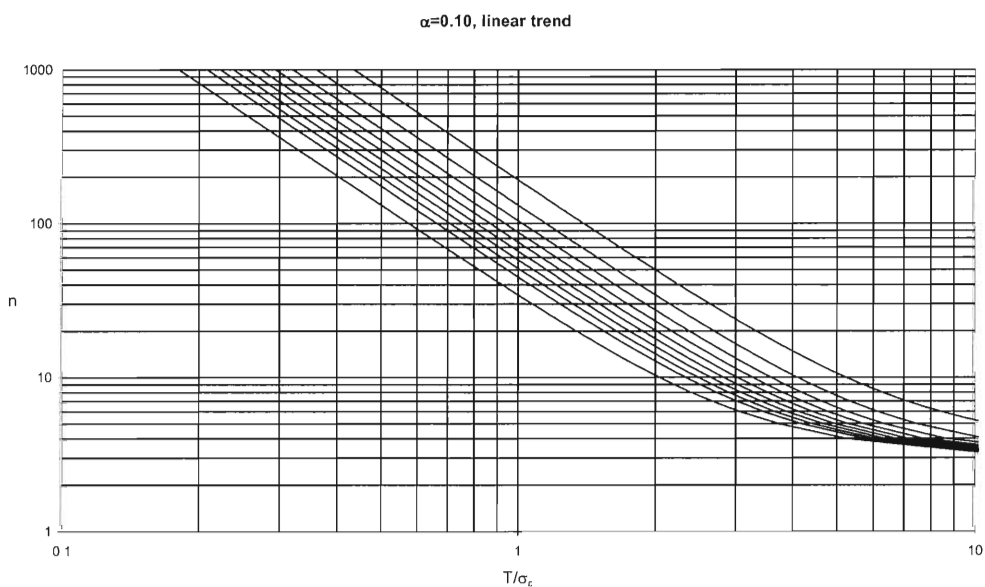
Een lineaire trend kan als volgt worden gemodelleerd:

$$Z_i = \varepsilon_i + i\tau + \gamma \quad (71)$$

waarin  $\varepsilon_i$  een onafhankelijk en gelijk verdeelde reeks vormt met variantie  $\sigma_\varepsilon^2$ ,  $\tau$  de lineaire trendcoëfficiënt is en  $\gamma$  het basisniveau is. De  $H_0$ -hypothese is  $\tau = 0$  en de alternatieve hypothese  $H_1$  is  $\tau \neq 0$ . Met een Student's  $t$ -toets kan een keuze gemaakt worden tussen  $H_0$  en  $H_1$ . Figuur 40 en 41 geven het benodigd aantal waarnemingen voor verschillende waarden van  $\alpha$  en  $\beta$ , en verschillende verhoudingen tussen de trend die nog moet kunnen worden aangetoond en de standaardafwijking van  $\varepsilon$ .



figuur 40: Benodigd aantal waarnemingen uitgezet tegen genormaliseerde lineaire trend voor  $\sigma=0.05$ ,  $\beta=0.5, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1, 0.05$  en  $0.01$  (van links naar rechts),  $T$  is de trendmagnitude en  $\sigma=\sigma_\varepsilon$



figuur 41: Benodigd aantal waarnemingen uitgezet tegen genormaliseerde lineaire trend voor  $\sigma=0.10$ ,  $\beta=0.5, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1, 0.05$  en  $0.01$  (van links naar rechts),  $T$  is de trendmagnitude en  $\sigma=\sigma_\varepsilon$

Tot nu toe zijn we ervan uitgegaan dat  $\varepsilon_i$  onafhankelijk en gelijk verdeeld is. Als er sprake is van correlatie in de reeks  $\varepsilon_i$ , dan zijn er meer waarnemingen nodig om voor trend te toetsen. Om dit zogenaamde equivalente aantal

waarnemingen te bepalen zijn formules en grafieken beschikbaar (De Gruijter e.a., 2006).

### 6.3.3 *Interventiemodellen*

De modellen die in de vorige paragraaf zijn gepresenteerd zijn relatief eenvoudig; naast de trend zijn er geen andere invloeden. Het kan echter zijn dat er naast de trend andere invloeden zijn die kunnen worden beschreven met een transfer-ruismodel. Het transfer-ruismodel uit vergelijking (27) kan dan worden uitgebreid met een trendcomponent tot een interventiemodel:

$$Z_t = Z_t^* + I_t + N_t \quad (72)$$

waarin  $I_t$  de interventiecomponent is:

$$I_t = \delta_1 I_{t-1} + \delta_2 I_{t-2} + \dots + \delta_p I_{t-p} + \omega_0 S_t - \omega_1 S_{t-1} - \omega_2 S_{t-2} - \dots - \omega_q S_{t-q} \quad (73)$$

met  $S_t = 0$  tot het moment van de interventie en 1 vanaf de interventie. De interventiecomponent kan verschillende vormen hebben. Bij een eenvoudige staptrend heeft alleen  $\omega_0$  een waarde en zijn de overige parameters in (39) gelijk aan 0. Voor interventiemodellering geldt het zelfde als wat beschreven is in paragraaf 6.3.1. Bij een staptrend is het belangrijk om ernaar te streven dat voor en na de interventie ongeveer evenveel waarnemingen zijn verricht. Als na de interventie het effect langzaam wegebt, dan is het belangrijk om na de interventie voldoende lang door te meten zodat de periode dat er nog een effect merkbaar is volledig is bestreken. Grafieken zoals in figuur 12 tot en met 15 zijn voor interventiemodellen niet eenvoudig te construeren.

## Practicum 1

### 'kennismaking met *Menyanthes* en tijdreeksmodellen'

Welkom in de wereld die *Menyanthes* heet! Dit eerste practicum heeft vooral ten doel om vertrouwd te raken met de 'besturing' van *Menyanthes*, de import-export van het programma, het door peilbuisdata heen 'browsen' met de listboxen en via selectie met de muis in het kaartje. De tweede opdracht, vooral voor mensen die *Menyanthes* al gebruiken, bestaat uit het handmatig fitten van een tijdreeksmodel, om zelf eens kennis te maken met de analyses die *Menyanthes* later geautomatiseerd voor je uit zal voeren.

#### Opdracht 1: Grondwaterstandsreeksen in de boswachterij Appelscha

1. Importeer een selectie van Grondwaterstandsreeksen, direct afkomstig uit DINO-loket.
2. Laad de topografische achtergrond met waterlopen.
3. Loop door de data via het listboxje. Waar zitten fouten of afwijkingen in de reeksen?
4. Loop door de data via het kaartje en de dwarsdoorsneden. Zoek naar fouten door meerdere reeksen tegelijk te selecteren.
5. Hoe zit het systeem in elkaar? Waar bevinden zich schijnspiegelsystemen? Hoe stroomt het grondwater globaal? Tip: Gebruik ook de optie 'Plot GW Level' via het 'spatial analysis' wereldbolletje en dan 2D/3D tools. Deze functies zijn overigens nog 'Under Construction'.

#### Opdracht 2: Fit tijdreeksmodel handmatig

6. Ga binnen *Menyanthes* naar help-> demo's en start de 'Convolution Demo' op.
7. Druk op de knop 'Plot Convolution'
8. De convolutie wordt uitgevoerd, en u krijgt een eerste iteratie ofwel schatting van uw grondwaterstandreeks in de onderste grafiek.
9. Varieer vervolgens de parameters 'A' en 'a' van de gebruikte exponentiele responsfunctie, en zorg ervoor dat uw modelsimulatie zo goed mogelijk over de metingen heen valt. Hoe groot is uw uiteindelijke minimale RMSE of modelfout?



## **Practicum 2**

### **‘Vergelijking van het Box-Jenkins- en PIRFICT-model’**

Dit practicum is gebaseerd op vijftien tijdreeksen van de grondwaterstand. De reeksen komen van een selectie van peilbuizen van het Noord-Hollands duinreservaat, in de buurt van Egmond aan Zee. Dit deel van de duinen is relatief onbeïnvloed door de grondwaterwinningen in het gebied, zodat de peilbuizen gemodelleerd kunnen worden met neerslag en verdamping als enige verklarende variabelen. De resultaten zijn eerder gepubliceerd onder de titel ‘Transfer function noise modeling in continuous time using predefined impulse response functions’ (Von Asmuth, Bierkens & Maas, 2002, Water Resources Research, 38, 12, 23\_1-23\_12.).

#### **Opdracht 1: Tijdreeksmodellering met het Box-Jenkins-model**

1. Kies een reeks en modelleer deze met behulp van het Box-Jenkins-model. Gebruik daarbij verschillende modelordes (aantal MA parameters, bijvoorbeeld 5, 10, 20).
2. Sla de resultaten op, en vergelijk deze op de volgende criteria:
  - Verklaarde variantie
  - Blokrespons-functie
3. Welke aantal MA parameters geeft de beste resultaten?
4. Waarom zien de responsfuncties van de verschillende BJ-modellen er verschillend uit?

#### **Opdracht 2: Tijdreeksmodellering met het PIRFICT-model, en vergelijking met het BJ-model**

5. Neem dezelfde reeks en modelleer deze met behulp van PIRFICT -model.
6. Vergelijk de Blokrespons-functie met die van het beste Box-Jenkins model.
7. Waarin verschillen de responsfuncties, en waarom?

#### **Opdracht 3: Invloed van de ‘Sample frequency’**

8. Varieer nu de ‘Sample frequency’ van het Box-Jenkins model, en vergelijk de responsfuncties. Waarom zijn ze verschillend?
9. Varieer de ‘Sample frequency’ van het PIRFICT-model, en vergelijk de responsfuncties. Wat gebeurt er?
10. Vergelijk de Blokresponsfuncties van het PIRFICT-model op verschillende frequenties, en dan de eerste x dagen daarvan. Verklaar de verschillen.
11. Vergelijk de Impulsresponsfuncties van het PIRFICT-model op verschillende frequenties. Zijn er verschillen, en zo ja, waarom?



**Aanvullende opdrachten:**

12. Vergelijk de resultaten van het PIRFICT model met vaste en variabele tijdstappen (interpolate aan/uit)
13. Kies een andere reeks, die er op het oog duidelijk anders uit ziet.
  - a. Wat is hier de optimale modelorde van het BJ model?
  - b. Wat zie je aan de responsfuncties? Welke reeks heeft de snelste reactie?
  - c. In hoeverre kun je de reactiesnelheid van het systeem al op het oog inschatten, en hoe komt dat tot uitdrukking in de fluctuaties van de grondwaterstand?





## **Practicum 3 'Metingen versus modelresultaten'**

Dit practicum is gebaseerd op een enkele peilbuisreeks (B24F0047), wederom afkomstig uit de duinen, en ditmaal gelegen in het Nationaal Park Zuid-Kennemerland. Ook deze reeks is relatief onbeïnvloed door andere factoren dan neerslag en verdamping. Het bijzondere van dit practicum is dat we deze reeks hebben opgeknipt in verschillende deelreeksen van oplopende lengte (1-20 jaar). We hebben er bovendien voor gekozen om met 'lopende' deelreeksen te werken, waarbij het 'window' telkens een jaar verschuift. We hebben dus 30 jaar reeksen van één jaar, 29 van twee jaar, 28 van drie jaar, etc., oftewel in totaal 410 deelreeksen.

**N.B.: Plak de belangrijkste resultaten en grafieken in een Powerpoint-document zodat je de resultaten eventueel aan het eind voor de groep kunt toelichten / bediscussieren.**

### **Opdracht:**

1. Maak tijdreeksmodellen voor alle jaren van de meetreeks afzonderlijk.
2. Maak met elk aldus verkregen tijdreeksmodel een simulatie van de grondwaterstand en sla deze op.
3. Bekijk de gesimuleerde reeksen. Wat valt je op?
4. Exporteer de modelresultaten naar Excel
5. In de uitvoer wordt per model de  $M0$ , de stationaire invloed van de neerslag, en de geschatte standaardafwijking daarvan gegeven. Verwijder op basis daarvan de schattingen met de grootste standaardafwijkingen uit de set van modellen.
6. Bekijk nogmaals de gesimuleerde reeksen, maar dan alleen van die modellen die je nog overhebt.
7. Bereken de GXG op basis van de meetgegevens, én op basis van de gesimuleerde reeksen. Bij welke van de twee is de variatie groter?
8. Herhaal het bovenstaande voor de reeksen van 2 en 3 jaar lang. Wat gebeurt er met de variatie in de simulaties en de GXG-schattingen als de reeksen langer worden?

### **Aanvullende opdrachten:**

1. Herhaal het bovenstaande voor de reeksen met nog grotere lengte.
2. Bepaal de standaardafwijking van de GxG-schatting voor de verschillende lengtes van de meetreeks, zowel voor de schattingen op basis van de metingen alleen, als voor de schattingen op basis van de gesimuleerde reeksen.
3. Maak een grafiek van de standaardafwijking van de GxG-schatting als functie van de lengte van de meetreeks. Wat is je conclusie? Gebruik je liever alleen meetwaarden, of liever tijdreeksanalyseresultaten op basis van de meetwaarden?



## **Practicum 4**

### **‘Het effect van ingrepen in de waterhuishouding op de grondwaterstand’**

Dit practicum is gebaseerd op twee tijdreeksen van grondwaterstanden die zijn geanalyseerd in het artikel ‘Physical basis of time series models for water table depths’ (Martin Knotters en Marc Bierkens, 2000, Water Resources Research 36(1): 181-188).

**N.B.: Plak de belangrijkste resultaten en grafieken in een Powerpoint-document zodat je de resultaten eventueel aan het eind voor de groep kunt toelichten / bediscussieren.**

#### **Case 1: Intensivering van de drainage**

In de loop van 1995 en 1996 is in de buurt van Lemelerveld (Berkendijk, gemeente Heino) de drainage geïntensiveerd, met als doel wateroverlast voor de landbouw te beperken. In het gebied staat buis 27FL0008, waarin halfmaandelijks de grondwaterstand wordt waargenomen.

##### **Opdracht:**

1. Lees de gegevens van buis 27FL0008 in van het bestand 27FL0008.xls. Lees de neerslag- en verdampingsdata van De Bilt in.
2. Beoordeel de grondwaterstandsreeks visueel. Denk je zo op het oog dat vanaf 1995/1996 de grondwaterstand gedaald is als gevolg van de ingrepen in de waterhuishouding?
3. Modelleer het effect van de ingrepen. Dat doe je met ‘Create Model’. Je kunt hier een staptrend als verklarende reeks toevoegen. We weten niet de precieze datum waarop de ingrepen effectief werden en kiezen 01/01/1996. We kiezen voor de dagfrequentie van neerslag en verdamping bij het modelleren. Voorts kiezen we voor een lineair model.
4. Welk effect heeft de ingreep op de grondwaterstand gehad? Is dit effect significant?
5. Bewaar het model.

#### **Case 2: Wateriaanvoer**

In de buurt van Lettele (Butersdijk, gemeente Diepenveen) is in de loop van 1985 een plan voor waterconservering uitgevoerd. In het gebied staat buis 27HL0023, die al lang wordt waargenomen. Omdat de buis enkele keren herplaatst en schoongemaakt is, zijn de data vanaf 1 januari 1982 betrouwbaar.

##### **Opdracht:**

1. Lees de gegevens van de buis in vanaf 27HL0023.XLS. Beoordeel de buis visueel? Wat valt je op aan de meetfrequentie?
2. Modelleer het effect van de ingreep. Kies als interventiedatum 1 januari 1986. Kies een dagfrequentie voor neerslag en verdamping en maak gebruik van een lineair model.



3. Welk effect heeft de ingreep op de grondwaterstand gehad? Is dit effect significant?
4. Bewaar het model.

**Aanvullende opdrachten:**

1. Hoe goed passen de modellen bij de waarnemingen, in absolute en relatieve zin?
2. Maak een reeks 27FL0008\_b (before), voor het gedeelte vóór de ingreep (met Edit en Select). Modelleer dit deel (zonder interventie).
3. Simuleer (deterministisch) de grondwaterstand met dit model. Bewaar de simulatie, en vergelijk deze met de waarnemingen. Hoe goed past het model bij de waarnemingen na de interventie?
4. Voor een dergelijke analyse ook uit voor 27HL0023, door een reeks 27HL0023\_a (after) vanaf de interventie te selecteren.