
Soms is weten beter dan meten (tenzij je verkeerd zit natuurlijk)

Het discrete Box-Jenkins versus het continue PIRFICT transferruis-model, in praktijk

Jos von Asmuth
Marc F.P. Bierkens
Kees Maas

In dit artikel volgt 'the proof of the pudding' voor wat betreft de vergelijking tussen het volledig statistisch getinte Box-Jenkins-model en het PIRFICT-model waarbinnen bepaalde fysische schematisaties van een systeem aan het model kunnen worden opgelegd. Het is wel aardig dat we aan het eind van deel 1 van dit verhaal geconcludeerd hebben dat het PIRFICT-model in theorie een aantal prettige eigenschappen zou moeten hebben, maar helaas zegt de wetenschapsfilosofie dat theorieën, hoe logisch ze ook mogen klinken, falsifieerbaar moeten zijn in de praktijk. Aldus gaan wetenschappers, volgens een andere wetenschapsfilosoof, naarstig op zoek naar waarnemingen die hun theorieën kunnen staven. Maar zien ze daarbij toevalligerwijze niet situaties over het hoofd die minder mooi kloppen bij de theorie? Een verhaal waarin wetenschappelijk gezien aloude dilemma's zoals zwarte versus witte dozen, analytisch versus numeriek en Popper versus Kuhn het weer eens tegen elkaar opnemen.

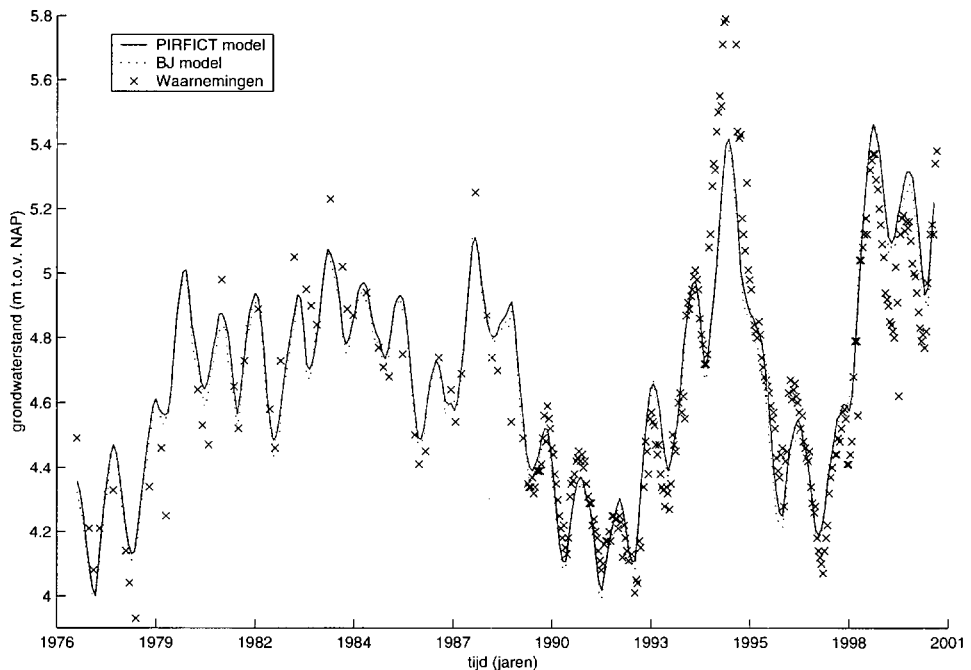
Inleiding

Het toetsen van de prestaties van een model in de praktijk, laat staan het vergelijken van verschillende modellen, is lastiger dan het op het eerste gezicht soms lijkt (zie bijvoorbeeld Konikow en Bredehoeft, 1992; Oreskes et al., 1994; Dekker en Van der Lee, 2001). Je kan natuurlijk, aan de hand van de kalibratieresultaten, beoordelen in hoeverre een model een reeks waarnemingen van een gegeven verschijnsel kan reproduceren, maar je weet al van tevoren dat hoe meer parameters en invoerdata je gebruikt, des te beter een model hierin is. Om op dit effect enigszins te anticiperen zijn de automatische modelselectie criteria ontworpen, die ook in deel 1 aan de orde zijn gekomen. Een betrouwbaarder beeld ontstaat door modelvoorspellingen te toetsen aan gegevens die niet als invoer voor het model hebben gediend, oftewel door het model te valideren. Om zoveel mogelijk uitsluitel te geven is een cross-validatie een goede optie. Hierbij wordt een model tweemaal gekalibreerd en gevalideerd, waarbij de kalibratie- en validatieperiode na de eerste maal omgedraaid worden. Zijn de omstandigheden in de twee perioden verschillend, dan zal een model dat toevallig in de ene toestand beter scoort, de tweede keer niet als beste uit de bus komen. Wil je echter de algehele prestaties van modellen vergelijken, dan is het van belang hoe de modellen onder alle mogelijke omstandigheden presteren, en kun je ook met een beperkt aantal cross-

validaties niet aantonen wat de prestaties van de modellen in andere omstandigheden c.q. bij andere datasets zullen zijn.

Omdat het onmogelijk is modellen onder alle denkbare omstandigheden te testen, richten we ons in dit artikel niet zozeer op 'de verkiezing van het beste model', maar eerder op de mechanismen die bij de prestaties van beide modellen een rol spelen. We steunen, met andere woorden, ook bij het behandelen van resultaten in de praktijk toch weer enigszins op theorie. Allereerst illustreren we de gevoeligheid van het Box-Jenkins (BJ)- en het PIRFICT-model voor door de modelleur gemaakte keuzes als modelorde en waarnemingsfrequentie, aan de hand van de kalibratieresultaten van een voorbeeld reeks. Als voorbeeld is die grondwaterstandsreeks gekozen die van 15 geanalyseerde reeksen de traagste respons te zien gaf, omdat juist dan een aantal prettige eigenschappen van het PIRFICT-model zichtbaar wordt. Vervolgens identificeren we één BJ-model voor alle reeksen, en vergelijken de resultaten met die van het PIRFICT model, gekalibreerd op dezelfde data. Voor 6 van de 15 reeksen zijn waarnemingen van vóór de kalibratieperiode beschikbaar en voeren we een validatie uit. Omdat de waarnemingen in die periode slechts op kwartaalbasis beschikbaar zijn kunnen we helaas geen cross-validatie uitvoeren. Om een aantal vermoedens te testen die voortkomen uit de resultaten van de 'real world' reeksen, voeren we als laatste een korte simulatie-studie uit.

De data waarover we beschikken zijn afkomstig van 15 peilbuizen uit het Noord-Hollands Duinreservaat, alle gelegen in de omgeving van Egmond. De selectie van de buizen heeft



Figuur 1: Simulaties van een [10 0 1 1 0] [30.4] Box-Jenkinsmodel en een PIRFICT-model, gekalibreerd op dezelfde grondwaterstandsreeks van peilbuis 19AZW246_1 tussen 1990 en 2001.

plaatsgevonden op het feit dat ze relatief weinig verstoord zijn door de grondwaterwinnings in het gebied (Rolf en Lebbink, 1998), zodat de grondwaterstandreeksen gemodelleerd kunnen worden met het neerslagoverschot als enige invoervariabele. De peilbuizen zijn verder over eenzelfde periode relatief goed bemeten, waarbij de grondwaterstand omstreeks de 14° en 28° van de maand handmatig afgelezen is, zodat de modelresultaten niet beïnvloed worden door een verschil in kalibratieperiode of door de compleetheid van de reeks. Voor het kalibratieproces zijn waarnemingen van de grondwaterstand uit de periode 1-1-1990 tot 1-1-2001 gebruikt, en gegevens over het neerslagoverschot vanaf 1-1-1987. De neerslagdata zijn beschikbaar op dagbasis en zijn door PWN in de buurt van Castricum gemeten, terwijl dagelijkse verdampingsgegevens gebruikt zijn van het KNMI station in De Kooy. Ter illustratie toont figuur 1 een tijd–stijghoogtediagram van de beschikbare waarnemingen van peilbuis 19azw246_1, samen met simulaties van een [10 0 1 1 0] [30,4] BJ-model en een PIRFICT-model van dezelfde data.

Kalibratie van verschillende BJ-modellen en het PIRFICT-model op een voorbeeldreeks

Om een BJ-model te kunnen kalibreren moet allereerst de modelorde daarvan geïdentificeerd worden. Als voorbeeld voor de rest van de buizen zijn de resultaten van een aantal verschillende BJ-modellen en van het PIRFICT-model uitgewerkt voor peilbuis 19AZW246_1. Uit fysische overwegingen beschouwen we hier alleen BJ-modellen met orde $[nb \ 0 \ 1 \ 1 \ nk] [th=tp]$, i.e. modellen waarbij het transfer- en ruisgedeelte van het model één auto-regressieve parameter, een vertragingstijd en een variabel aantal moving-average parameters bevat. Voor het aantal moving-average parameters nb kiezen we daarbij 5, 10 en 20, deels afhankelijk van de waarnemingsfrequentie, wat het bereik van waarschijnlijke modelordes wel omvat. De tijdsdiscretisatie voor het BJ-model kiezen we in overeenstemming met de waarnemingsfrequentie die beschikbaar is, en zodanig dat de respons van het systeem adequaat beschreven kan worden met een hanteerbaar aantal parameters. Voor deze reeksen komt dit neer op een discretisatie th van 15,2, 30,4 of 60,8 dagen, zodanig dat er precies 24, 12 of 6 waarnemingen in een jaar beschikbaar zijn. Omdat BJ-modellen equidistante reeksen nodig hebben als invoer, en aangezien de waarnemingen omstreeks de 14° en 28° van elke maand gedaan zijn en er her en der waarnemingen missen, moet de reeks nog equidistant gemaakt en aangevuld worden. Een equidistante reeks vanaf 1-1-1990 met een waarneming om de 15,2 dagen verkrijgen we door de grondwaterstanden lineair te interpoleren tussen de twee dichtstbijzijnde waarnemingen. Waarnemingen van het neerslagoverschot zijn beschikbaar op dagbasis, zodat we hier een equidistante reeks verkrijgen door de dagelijkse sommen verder te sommeren voor elke tijdsstap.

Omdat zowel de invoer- als uitvoerreeks veranderen door de interpolatie- en sommatieoperaties die nodig zijn voor de BJ-modellen, kalibreren we ter vergelijking een PIRFICT-model voor elke dataset die aldus verkregen is. Om hiertussen onderscheid te kunnen maken labelen we de PIRFICT-modellen met $[th \ tp]$ en met (non)equidistant, wat echter in dit geval geen verschil in orde weergeeft zoals bij BJ-modellen, maar een verschil in de gebruikte dataset. De modelresultaten geven op deze manier een indicatie van de prestaties van een reeks van BJ-modellen, van hun afhankelijkheid van door de gebruiker gedefiniëerde keuzes als modelorde en waarnemingsfrequentie en van de prestaties van het PIR-

FICT-model, gebruikmakende van dezelfde data. Dit betekent wel dat waarschijnlijk geen van de BJ-modellen het optimale resultaat zal laten zien, maar dat de algehele prestatie van het BJ-model afgeleid moet worden uit de resultaten van de verschillende modelordes en uit hun gevoeligheid voor de gemaakte keuzes.

De kalibratieresultaten van de BJ-modellen en het PIRFICT-model zijn weergegeven in tabel 1, waarbij de verschillende modellen worden vergeleken aan de hand van een aantal criteria. Allereerst worden de ‘Root Mean Squared Error (RMSE)’ en ‘Root Mean Squared Innovation (RMSI)’ vermeld, die een maat vormen voor respectievelijk de fout van het transfermodel en de variantie van het ruisproces. Omdat de waarnemingsfrequentie en de interpolatieoperaties de schatting van de variantie van een tijdreeks, en daarmee ook de RMSE van het transfermodel, kunnen beïnvloeden, wordt ook het percentage verklaarde variantie gegeven, dat gedefinieerd is als:

$$EVP = \frac{\sigma_{h(t)}^2 - \sigma_{n(t)}^2}{\sigma_{h(t)}^2} * 100\% \quad (1)$$

Dat het bovengenoemde effect ook daadwerkelijk optreedt blijkt wel uit het feit dat voor het PIRFICT-model de 60,8-daagse data de hoogste RMSE, maar juist de laagste verklaarde variantie oplevert. Een logische manier om de resultaten van verschillende modellen te vergelijken lijkt het gebruik van automatische modelorde selectie criteria, zoals Akaike’s Information Criterion en Final Prediction Error criterion. Omdat deze criteria voor het bepalen van de juiste modelorde echter de innovatie-variantie of hun likelihoodfunctie gebruiken, die beide beïnvloed worden door de waarnemingsfrequentie, kunnen deze niet gebruikt worden voor het vergelijken van modellen waarbij verschillende waarnemingsfrequenties of datasets gebruikt zijn.

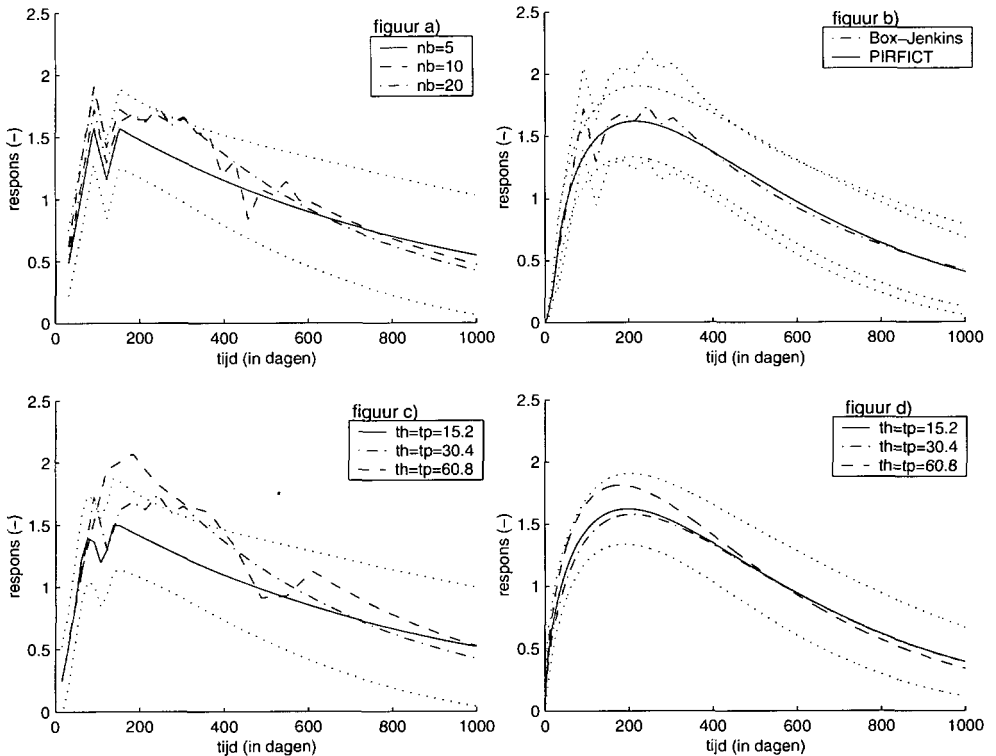
Tabel 1: Resultaten van BJ-modellen met verschillende ordes en PIRFICT-modellen gekalibreerd op dezelfde data van peilbuis 19AZW246_1.

| Box-Jenkins-modellen | | | | | |
|----------------------|--------------------|--------------|--------------|--------------|--------------------|
| Modelorde | [5 0 1 1 0] [30.4] | [10 0 1 1 0] | [20 0 1 1 0] | [20 0 1 1 0] | [5 0 1 1 0] [60.8] |
| [nb nc nd nf nk] | | [30.4] | [30.4] | [15.2] | |
| [th = tp] | | | | | |
| RMSE (cm) | 17,82 | 16,39 | 17,08 | 16,58 | 17,33 |
| RMSI (cm) | 8,61 | 8,42 | 8,06 | 7,04 | 10,41 |
| EVP (%) | 82,35 | 85,08 | 83,79 | 84,83 | 83,59 |
| PIRFICT-model | | | | | |
| RMSE (cm) | | 16,25 | | 16,41 | 17,43 |
| RMSI (cm) | | 9,03 | | 7,27 | 10,52 |
| EVP (%) | | 85,32 | | 85,14 | 83,39 |

Uit de gegevens van tabel 1 blijkt allereerst dat de verschillen tussen de twee modellen wat fit betreft niet erg groot zijn. De laagste RMSE en hoogste verklaarde variantie worden weliswaar bereikt door het PIRFICT-model, maar de variatie die de verschillende waarnemingsfrequenties en interpolaties met zich meebrengt is groter dan de verschillen met het BJ-model. Het optimum voor het BJ-model lijkt te liggen bij het [10 0 1 1 0] [30,4] model,

zodat we voornamelijk dit als de juiste modelorde accepteren. De RMSI en dus de innovatievariantie is zoals verwacht inderdaad afhankelijk van de tijdstap tussen twee waarnemingen van de uitvoerreeks. Voor $th = 15,2$ ligt de RMSI in de orde van 7 tot 8 cm, terwijl de RMSI 9 tot 10 cm is voor $th = 60,8$. Opmerkelijk is dat de RMSI van de BJ-modellen af blijkt te nemen met het aantal parameters, terwijl de RMSE een optimum vertoont voor het $[10\ 0\ 1\ 1\ 0]$ [30,4] model. Dit fenomeen kan wellicht verklaard worden uit het feit dat de modellen gekalibreerd worden door de RMSI en niet de RMSE te minimaliseren, waardoor juist de RMSI vermindert wanneer extra parameters worden toegevoegd. Een verschil tussen het PIRFICT- en BJ-model dat duidelijk naar voren komt is dat de RMSI van het PIRFICT-model structureel hoger is.

De parameterwaarden van beide modellen kunnen we het beste vergelijken door de geschatte BR- en transfer-functies tegen elkaar af te zetten. Figuur 2a) toont de transfer-functies van drie $[nb\ 0\ 1\ 1\ 0]$ [30,4] BJ-modellen met $nb = 5, 10$ and 20 . Uit de betrouwbaarheidsintervallen valt op te maken dat de resultaten van een BJ-model significant beïnvloed worden door de modelorde (wat een tijdreeksmodelleur niet zal verbazen) aangezien de transfer-functies van het $nb = 10$ en 20 model deels net buiten het interval van het $nb = 5$ model liggen. Ook kan men uit deze figuur afleiden dat een aantal van 10 MA-parameters



Figuur 2: Blokrespons- en transfer-functies van verschillende TR-modellen van peilbuis 19AZW246_1, met in figuur a) drie $[nb\ 0\ 1\ 1\ 0]$ [30,4] BJ-modellen met $nb = 5, 10$ en 20 , in figuur b) een $[10\ 0\ 1\ 1\ 0]$ [30,4] BJ- en een PIRFICT-model van dezelfde data, in c) en d) respectievelijk drie $[10\ 0\ 1\ 1\ 0]$ [$th\ tp$] BJ- en drie [$th\ tp$] PIRFICT-modellen met $th = tp = 15,2, 30,4$ en $60,8$. De gestippelde lijnen geven steeds het 95% betrouwbaarheidsinterval van een van de modellen weer (in a) het $nb = 5$ in b) beide in c) en d) het $th = tp = 30,4$ model).

juist voldoende is om het eerste deel van de respons van het systeem goed te kunnen beschrijven, terwijl de staart adequaat gemodelleerd kan worden met één enkele AR-parameter. Figuur 2b) laat zien dat de geschatte responsfunctie van het PIRFICT-model en het BJ-model nagenoeg overeenkomen, met als belangrijkste verschil dat de BJ-transferfunctie deels onregelmatig is, terwijl de PIRFICT-respons glad is. Dit is consistent met het feit dat de RMSE's niet veel van elkaar verschillen. In figuur 2c) is de orde van het BJ-model, in de meer traditionele zin van het woord, constant gehouden terwijl de waarnemingsfrequentie is gevarieerd, resulterend in drie $[10\ 0\ 1\ 1\ 0]$ $[th=tp]$ BJ-modellen met $th = tp = 15,2, 30,4$ en $60,8$. Ook hier is te zien dat de resultaten significant beïnvloed worden door de keuze van de waarnemingsfrequentie, aangezien deze interfereert met de modelorde c.q. de vorm van de transfer-functie in werkelijke tijd. Tot slot tonen we ook de IR-functies van drie PIRFICT-modellen die op dezelfde equidistante data gekalibreerd zijn als het BJ-model, dus waarbij $th = tp = 15,2, 30,4$ en $60,8$. Afgaande op deze figuur blijkt de schatting van de modelparameters voor de eerste twee waarnemingsfrequenties nagenoeg identiek, terwijl het model klaarblijkelijk moeite heeft om het begin van de IR-functie goed te schatten voor $th = tp = 60,8$ (alhoewel de functie wel binnen het betrouwbaarheidsinterval valt). Dit effect heeft waarschijnlijk te maken met het toenemende tijdsinterval en afnemende aantal waarnemingen, aangezien het BJ $[60,8]$ model het begin ook te hoog inschat.

Kalibratie en validatie van 15 c.q. 6 grondwaterstandsreeksen

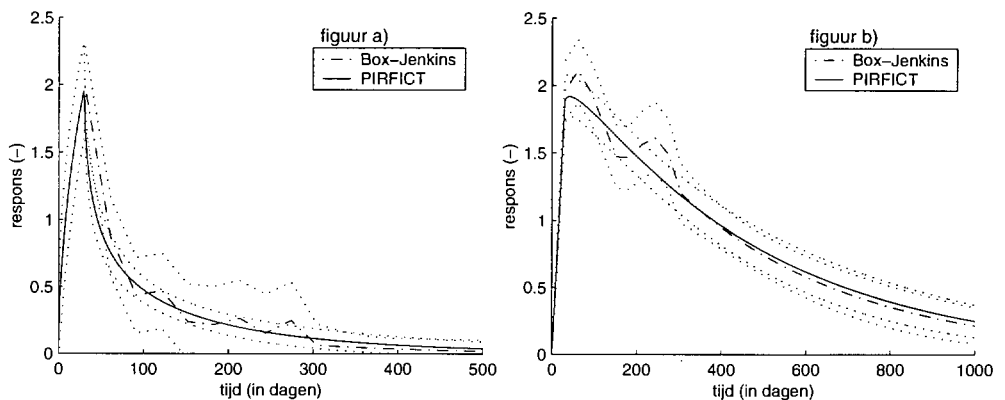
Voor een bredere vergelijking tussen het Box-Jenkins- en PIRFICT-model zijn beide modellen gekalibreerd op grondwaterstandsreeksen van 15 peilbuizen met waarnemingen uit de periode 1-1-1990 tot 1-1-2001. Voor de vergelijkbaarheid en objectiviteit voeren we niet voor elke peilbuis apart een iteratieve model-identificatieprocedure uit om de meest geëigende modelorde (in ons geval gedefinieerd door nb , nk en th) van het BJ-model te kiezen. Als alternatief kalibreren we een $[10\ 0\ 1\ 1\ 0]$ $[30,4]$ BJ-model op alle tijdreeksen, aangezien dit de transfer-functie genoeg MA-parameters en dus vrijheid geeft om zelfs de traagste respons van de 15 reeksen aardig te kunnen benaderen. De respons-functies van de BJ en PIRFICT-modellen blijken voor alle peilbuizen dezelfde grote lijn te volgen, wat aangeeft dat het PIRFICT-model in ieder geval voor deze reeksen de respons van (relatief) snelle, gemiddelde en trage systemen adequaat kan schatten (figuur 2b,3a en b). Voor systemen met een snelle of gemiddelde respons zou het BJ-model waarschijnlijk ook met een kleiner aantal parameters kunnen volstaan, zodat deze reeksen meer dan nodig ge'overfit' zullen worden.

De gemiddelde kalibratie- en validatieresultaten van de gemodelleerde peilbuizen zijn

| | BJ [10 0 1 1 0] [30.4] | PIRFICT [30.4 30.4] equi. | PIRFICT [30.4 1] non-equi. |
|-------------|---------------------------|------------------------------|-------------------------------|
| RMSE (cm) | 12,42 | 12,56 | 12,50 |
| RMSI (cm) | 7,20 | 7,65 | 7,89 |
| EVP (%) | 88,66 | 88,37 | 88,36 |
| V-RMSE (cm) | 15,76 | 15,64 | 15,38 |

Tabel 2: Gemiddelde kalibratie- en validatieresultaten van het BJ- en het PIRFICT-model voor 15 en 6 peilbuizen.

weergegeven in tabel 2. Naast de RMSE, RMSI en het percentage verklaarde variantie staat in deze tabel ook de gemiddelde validatie-RMSE van de 6 peilbuizen waarvoor gegevens van vóór 1-1-1990 beschikbaar waren. Een tijd-stijghoogte diagram met waarnemingen van peilbuis 19AZW246_1 en simulaties van een $[10\ 0\ 1\ 1\ 0]$ $[30,4]$ BJ-model en een equidistant $[30,4\ 30,4]$ PIRFICT-model voor zowel de kalibratie- en validatieperiode is weergegeven in figuur 1. De figuur laat duidelijk zien dat de waarnemingsfrequentie van de peilbuis niet constant is, maar is veranderd van 4 keer per jaar naar gemiddeld 24 keer per jaar aan het einde van 1989. Omdat de waarnemingsfrequentie in de validatieperiode een stuk lager is dan de frequentie van de simulaties, zijn voor de validatie de gesimuleerde waarden geïnterpoleerd naar de waarnemingsdatum en niet andersom, zoals in de kalibratieperiode. De simulaties van het PIRFICT-model hoeven niet geïnterpoleerd te worden, omdat de gesimuleerde grondwaterstanden daar continu zijn. Voor alle modellen zijn de resultaten met toepassing van een vertragingstijd gegeven omdat deze zowel de kalibratie- als validatieresultaten duidelijk bleek te verbeteren, het meest nog (logischerwijze) voor het PIRFICT-model dat gebonden is aan een vaste vorm. De vertragingstijd is, op basis van een handmatige model-identificatieprocedure, gekozen als $nk = [12\ 6\ 0\ 0\ 0\ 3\ 1\ 13\ 13\ 6\ 6\ 8\ 1\ 5\ 0]$ dagen voor de 15 peilbuizen, en is toegepast door de neerslagreeks te verschuiven over de tijdsas. Door de invoerreeks, en niet de responsfunctie, te verschuiven wordt het mogelijk om vertragingstijden toe te passen die kleiner zijn dan het discrete tijdsinterval, waardoor de vertragingstijd voor zowel de BJ- als de PIRFICT-modellen gelijk gehouden kon worden.



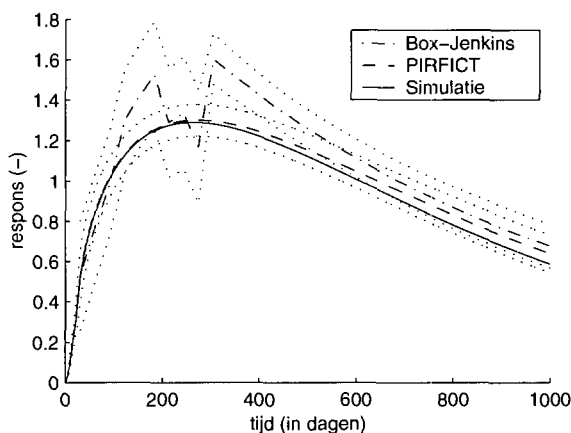
Figuur 3: Transfer-functies van een $[10\ 0\ 1\ 1\ 0]$ $[30,4]$ BJ model en blokrespons van een PIRFICT-model gekalibreerd op dezelfde data van peilbuizen 19AZL5038_1 en 19AZW195_1. De gestippelde lijnen geven de 95% betrouwbaarheidsintervallen weer.

De gemiddelde RMSE van de 15 reeksen van het $[10\ 0\ 1\ 1\ nk]$ $[30,4\ 30,4]$ BJ-model komt, met een verschil van slechts 0,8 millimeter, nagenoeg overeen met de RMSE van het niet-equidistante $[30,4\ 1]$ PIRFICT-model. In 7 van de 15 peilbuizen was de fit van het PIRFICT-model beter, terwijl bij 8 peilbuizen het BJ-model de beste resultaten te zien gaf. Tegelijkertijd echter, is de gemiddelde RMSI van het BJ-model ook hier beduidend lager. Voor een deel wordt de lagere RMSI waarschijnlijk veroorzaakt door de interpolatie-operaties die uitgevoerd zijn om de niet-equidistante reeksen equidistant te maken, aangezien het interpolatie-proces de willekeurige variatie c.q. ruis voor een deel uitmiddelt. Een aanwijzing hiervoor is het feit dat het PIRFICT-model voor de equidistante reeks een RMSI oplevert die 2,4 millimeter lager is dan de RMSI van de niet-equidistante reeks. Omdat de

RMSI van de equidistante BJ- en PIRFICT-modellen gemiddeld nog 4,5 millimeter van elkaar verschillen, zal het verschil waarschijnlijk deels ook te wijten zijn aan de verschillende modelstructuur. Zoals eerder geconcludeerd en ook te zien is in figuren 3a en b, komt de PIRFICT-blokrespons in grote lijn overeen met de transferfunctie van het BJ-model, waarbij de laatste echter een onregelmatig verloop rond de gladde continue curve te zien geeft. Omdat de RMSE van het BJ- en PIRFICT-model ongeveer gelijk zijn, zal het onregelmatige verloop van de BJ-transferfunctie wellicht deels toevallige cross-correlaties tussen de residu-reeks en de neerslagoverschotsreeksen fitten, wat de hoeveelheid ruis of RMSI die nodig is om de residureeks te modelleren verlaagt. Het vermoeden dat het onregelmatige patroon van de BJ-transferfunctie deels toevallig is wordt ondersteund door de lagere validatie-RMSE van het PIRFICT-model.

Een korte simulatie-studie

Afgaande op bovenstaande resultaten lijkt het BJ-model wat gevoeliger te zijn voor toevallige correlaties in de residureeks. Omdat we echter geen cross-validatie hebben kunnen uitvoeren, en niet zeker weten wat de respons-functies 'in werkelijkheid' waren, doen we nog een korte simulatie-studie waarbij we een reeks genereren door een neerslagoverschotreeks te convolueren met een zelf gekozen impuls-respons, in de vorm van een PearsonIII-functie ($A = 1500$, $\alpha = 0,002$, $n = 1,5$). Hieraan voegen we een residu-reeks toe, die we verkrijgen door een reeks van normaal verdeelde random getallen (witte ruis, $\sigma = 2,5$ cm op dagbasis) te convolueren met een wederom zelf gekozen, exponentiële ruis-IR-functie ($\alpha = 25$). Van de aldus verkregen reeks weten we precies wat de parameters zijn die we met onze tijdreeksanalyse terug zouden moeten vinden. De transferfuncties van een [10 0 1 1 0] [30,4] BJ-model en een PIRFICT-model zijn, naast de werkelijke respons, te zien in figuur 4. De geschatte parameters van het ruismodel zijn respectievelijk [$\alpha = 29,6$, $\sigma_a = 2,39$ cm] en [$\alpha = 27,9$, $\sigma_a = 2,43$ cm]. De transfer-functie van het BJ-model blijkt, net als bij de analyses van de werkelijke reeksen, een onregelmatig patroon te vertonen waarbij de functie deels zelfs buiten het betrouwbaarheidsinterval van de PIRFICT-respons ligt. Bij vergelij-



Figuur 4: Transfer-functies van een [10 0 1 1 0] [30,4]BJ-model en een PIRFICT-model gekalibreerd op dezelfde gesimuleerde reeks. De doorgetrokken lijn geeft de werkelijke respons en de gestippelde lijnen de 95% betrouwbaarheidsintervallen weer.

king met de 'werkelijkheid' is de afwijking van het PIRFICT-model over de gehele linie iets minder groot dan met het BJ-model.

Discussie en conclusies

De uitgevoerde analyses op de voorbeeldreeks bevestigen de verwachting dat de resultaten van het Box–Jenkins-model afhankelijk zijn van de modelorde en de waarnemingsfrequentie, terwijl het PIRFICT-model hier niet gevoelig voor is. Voor beide modellen blijft natuurlijk wel de voorwaarde staan dat de meetreeks voldoende waarnemingen dient te bevatten (zowel qua aantal als dichtheid) om de parameters goed te kunnen schatten. De prestaties van het PIRFICT-model blijken voor alle 15 peilbuizen vergelijkbaar te zijn met die van het BJ-model. Een eerste indicatie hiervoor is dat de geschatte respons van het PIRFICT-model doorgaans binnen het betrouwbaarheidsinterval van het BJ-model valt. Ten tweede bleek de kalibratie-RMSE van het PIRFICT-model gemiddeld slechts 0,8 millimeter hoger dan dat van het BJ-model, terwijl de validatie-RMSE van het PIRFICT-model juist 3,8 millimeter lager was. Dit fenomeen zou, alhoewel de verschillen klein zijn, goed toegeschreven kunnen worden aan de structuur van het BJ-model. Omdat de transferfunctie van BJ-modellen in principe elke vorm aan kan nemen, is deze ook vrij om toevallige cross-correlaties tussen de invoerreeks en de residureeks te fitten, wat resulteert in een lagere kalibratie-RMSI en een deels onregelmatig patroon in de transferfunctie. Het toevallige karakter van dit patroon kan verklaren waarom het PIRFICT-model een lagere validatie-RMSE te zien gaf. Vanwege de lage waarnemingsfrequentie in de periode voor 1-1-1990 kon helaas geen cross-validatie uitgevoerd worden, wat al te stellige conclusies uitsluit.

Bij een korte studie op basis van een gesimuleerde reeks kon wel worden vastgesteld dat het PIRFICT-model een schatting opleverde die dichter bij de gesimuleerde werkelijkheid zat. Alhoewel het misschien op het eerste gezicht triviaal lijkt dat het PIRFICT-model 'zichzelf' beter reproduceerde dan het BJ-model, is het dat bij nadere beschouwing toch niet. Allereerst komt uit het simulatie-experiment naar voren dat de onregelmatigheden in het patroon van BJ-transferfuncties inderdaad deels op toevalligheid berusten. Daarnaast toont de studie ook een principe aan dat het mogelijk maakt om uitspraken te doen over de prestatie van beide modellen in andere situaties. Door de keuze van een voorgedefinieerde IR-functie wordt de respons die het PIRFICT-model schat 'gedwongen' bepaalde fysische wetmatigheden te volgen. Dit vermindert aan de ene kant de gevoeligheid van het model voor toevallige correlaties in de data en levert daardoor betere schattingen op (in dat geval is dus 'weten beter dan meten'), maar kan aan de andere kant de resultaten negatief beïnvloeden als de fysische aannamen niet blijken te kloppen of te grof zijn (dus als 'je verkeerd zit'). Voor de reeksen die geanalyseerd zijn bleek de Pearson-III-verdeling, ofwel een systeem van gekoppelde reservoirs, een goede beschrijving van de respons van de grondwaterspiegel op neerslag te kunnen geven. De praktijk zal moeten uitwijzen of andere functies of combinaties van functies in andere situaties betere resultaten opleveren. Vanwege de flexibiliteit van de Pearson-III-verdeling mag echter verwacht worden dat deze functie in alle gevallen tenminste de belangrijkste karakteristieken van de respons van grondwatersystemen goed benadert.

Dankwoord

De auteurs willen bij deze PWN, en met name Harry Rolf, danken voor het beschikbaar stellen van de gegevens en het uitvoeren van test-tijdreeksanalyses met behulp van het programma TRG. Dank gaat bovendien uit voor het beschikbaar stellen van tijd en ruimte in de aanloop van dit onderzoek om met hun methode, programmatuur en gegevens vertrouwd te raken.

Literatuur

Rolf, H.L.M. en Lebbink, J. J. (1998) Hydrologische effecten als gevolg van de reductie van de duinwaterwinning in het Noord-Hollands Duinreservaat; Provinciaal Waterleidingbedrijf Noord-Holland, Heemskerk.