
Meervoudige tijdreeksmodellen en de samenhang in stijghoogtereeksen

Frans van Geer¹ en Wilbert Berendrecht²

Tijdreeksmodellen worden steeds vaker toegepast om de dynamiek in grondwaterstanden en stijghoogten te analyseren. Door de toepassing van meervoudige tijdreeksmodellen kunnen stijghoogtereeksen niet alleen gemodelleerd worden op basis van historische informatie, maar ook op basis van ruimtelijke samenhang. Zo kunnen data van omliggende meetpunten bijdragen aan een betere analyse van een meetreeks en kunnen samenhangende patronen gedetecteerd en geanalyseerd worden. In dit artikel geven we een beschrijving van het meervoudige tijdreeksmodel en passen we dit toe om mogelijke fouten in meetreeksen te detecteren.

Inleiding

Het modelleren van stijghoogtereeksen met een tijdreeksmodel heeft de laatste decennia een grote vlucht genomen. Niet alleen het aantal tijdreeksmodellen van stijghoogtereeksen dat de afgelopen jaren is gemaakt, is indrukwekkend, maar ook het aantal tijdreeksmodelleurs is flink toegenomen. In eerste instantie waren het 'klassieke' Box-Jenkins modellen (Box en Jenkins, 1976; Van Geer e. a. 1988). Sinds de meetfrequentie drastisch is verhoogd als gevolg van het gebruik van automatische loggers, wordt steeds vaker gebruik gemaakt van Kalmanfiltertechnieken (Berendrecht, 2004) of van een op voorhand geformuleerde continue responsfunctie (Von Asmuth, 2012). Tot dusver zijn vrijwel alle toepassingen enkelvoudige tijdreeksmodellen. Dat wil zeggen dat één outputreeks, in ons geval een stijghoogtereeks, wordt verklaard met behulp van één of meer inputreeksen. Berendrecht introduceerde een meervoudig tijdreeksmodel, waarbij ook het onderlinge verband tussen meerdere outputreeksen wordt beschreven (Berendrecht, 2004). Met een meervoudig tijdreeksmodel wordt een stijghoogtereeks niet alleen beschreven met behulp van het eigen verleden van de reeks, maar ook als element van een ruimtelijk patroon. Er zijn vele mogelijkheden om een cluster van tijdreeksen in samenhang te modelleren. Zo beschrijven Camacho e.a. (1987) een 'contemporaneous' ARMA model waarin een variabele gemodelleerd wordt als een lineaire combinatie van zowel eerdere eigen waarnemingen als eerdere waarnemingen van omliggende variabelen. Deutsch en Pfeifer (1981) en Stoffer (1986) beschrijven een variant hierop, het 'space-time' ARMA model, waarbij fysische kennis wordt gebruikt om a-priori de ruimtelijke samenhang te definiëren en daarmee het aantal vrijheidsgraden te reduceren. De essentie van de methode beschreven door

¹ Geologische Dienst Nederland-TNO, Utrecht & Universiteit Utrecht (frans.vangeer@tno.nl)

² Berendrecht Consultancy, Harderwijk (info@berendrechtconsultancy.nl)

Berendrecht (2004) en in dit artikel toegepast, is dat reeksen worden opgesplitst in een of meerdere gezamenlijke patronen en per locatie een specifiek/uniek patroon. Een voordeel ten opzichte van eerdergenoemde methoden is, dat het model veelal aanzienlijk minder modelparameters bevat doordat de samenhang met slechts enkele gezamenlijke patronen wordt beschreven (Zuur e.a., 2003). Daarnaast opent de decompositie in gezamenlijke en specifieke patronen de weg voor een aantal interessante nieuwe toepassingen.

De Geologische Dienst Nederland-TNO wil meer informatie betreffende de stijghoogtereeksen geautomatiseerd via internet beschikbaar maken. Dit betreft onder meer de karakterisatie van de reeks, temporele trends en ruimtelijke patronen. Als onderdeel hiervan heeft De Geologische Dienst Nederland-TNO samen met Berendrecht Consultancy het Python programma METRAN ontwikkeld om de meervoudige tijdreeksmodellering eenvoudig uit te voeren. De ontwikkelde software is ook handmatig te gebruiken, bijvoorbeeld voor het opvullen van gaten in een reeks en het detecteren van extreme waarden ten behoeve van kwaliteitscontrole.

Hoogfrequente reeksen maken het aantrekkelijk om ook in de modellering een kleine tijdstap te gebruiken, bijvoorbeeld een tijdstap van één dag. In Nederland is de correlatie van de stijghoogte tussen opeenvolgende dagen in het algemeen zeer hoog. Bij de 'klassieke' Box-Jenkins modellen zijn dan meestal veel parameters nodig voor een goede beschrijving van de responsfunctie, hetgeen vaak problematisch is bij de schatting van de parameters. Om de problemen bij het schatten van de responsfunctie te ondervangen kan het aantal vrijheidsgraden (het aantal onafhankelijk te schatten parameters) worden teruggebracht. Eén methode om het aantal vrijheidsgraden te reduceren is het gebruik van de Pearson-type-III-verdeling (Von Asmuth, 2012), die in de hierna beschreven METRAN programmatuur is toegepast. De programmatuur is echter zodanig opgezet, dat andere responsfuncties en/of modelvormen eenvoudig zijn toe te voegen. De meervoudige tijdreeksmodellen zijn geschreven in een state-spaceformulering in combinatie met het Kalmanfilter, omdat dat goed bij de probleembeschrijving past en bovenal voordelen biedt bij de schatting van de parameters (Berendrecht, 2004).

Waarom ruimtelijke samenhang mee modelleren?

Stijghoogtereeksen van filters in hetzelfde hydrologische systeem vertonen vaak een overeenkomstig gedrag. Voor veel toepassingen zoals het beoordelen van trends, het detecteren van extremen en het aanvullen van perioden zonder waarnemingen, ligt het voor de hand om van de samenhang in stijghoogtereeksen gebruik te maken. Dat is de reden dat in de tweede helft van de vorige eeuw zogenoemde stambuizen zijn aangewezen. In termen van tijdreeksmodellen kun je een stambuis zien als inputreeks voor de andere stijghoogtereeksen. De veronderstelling was dat de stambuizen 'niet beïnvloed' zijn en daarmee als referentie zouden kunnen dienen voor andere stijghoogtereeksen. In de praktijk bestaan er in Nederland feitelijk geen onbeïnvloede stijghoogtereeksen. Hierdoor is het in de meeste gevallen lastig om harde conclusies te trekken over structurele veranderingen, ook al zijn de tijdreeksmodellen met een stambuis als inputreeks in wiskundige zin vaak uitstekend. Bovendien wordt telkens een één-op-één-relatie beschreven, zonder formele ruimtelijke context.

Een meervoudig tijdreeksmodel beschouwt een aantal stijghoogtereeksen als cluster van gelijkwaardige reeksen. Het basisidee is het detecteren van ruimtelijk samenhangende patronen en locatie specifieke variaties. De hier beschreven meervoudige modellering sluit aan bij de situatie van stijghoogtereeksen in Nederland. Veel stijghoogtereeksen zijn sterk beïnvloed door het neerslagoverschot. Dit geldt zeker voor de freatische grondwaterstand. Daarom modelleren we eerst alle stijghoogtereeksen univariaat met het neerslag en verdamping als inputreeksen. Vervolgens beschouwen we de residureeksen van de univariate modellering als cluster, waarin we met een Dynamic Factor Model ruimtelijk samenhangende patronen detecteren, de Common Dynamic Factors (Geweke, 1977; Stock en Watson, 2010). Wat er uiteindelijk per meetlocatie overblijft, zijn de Specific Dynamic Factors, die geen verband vertonen met het neerslagoverschot en ook niet samenhangen met de andere residureeksen in het cluster. Uiteraard vertoont ook de reactie van de stijghoogte op het neerslagoverschot een ruimtelijk patroon, zie bijvoorbeeld Van Geer en Zuur (1997), maar dit valt buiten de scope van dit artikel.

Naast de toepassingsmogelijkheden van de enkelvoudige tijdreeksmodellen, bieden meervoudige tijdreeksmodellen extra mogelijkheden. De Common Dynamic Factors geven een beeld van de ruimtelijk samenhangende patronen die niet direct uit neerslagoverschot (of andere verklarende grootheden) verklaard kunnen worden. Dit kunnen bijvoorbeeld tweede-orde-effecten zijn van het neerslagoverschot, maar ook andere invloeden, zoals beregening, waterhuishoudkundige maatregelen enz. De vorm van een bepaalde Common Dynamic Factor is voor alle stijghoogte reeksen gelijk, maar de meervoudige modellen geven per reeks aan hoe sterk die Common Dynamic Factor aanwezig is. Dit levert een beeld op hoe een bepaald patroon ruimtelijk in sterkte toe- of afneemt. Als we bijvoorbeeld een cluster van stijghoogtereeksen van filters uit hetzelfde boorgat beschouwen, geven de Common Dynamic Factors een indicatie in hoeverre er mogelijk hydraulische kortsluitingen in de verticaal zijn. De Specific Dynamic Factor van een reeks is de variatie van de stijghoogte die, gegeven het model, op geen enkele wijze is te relateren aan de bekende drijvende krachten of aan omliggende stijghoogtereeksen.

Hoeveel variatie van een reeks met Common Dynamic Factors wordt beschreven is uiteraard afhankelijk van de samenstelling van het cluster waarmee de modellering wordt uitgevoerd. Bij een goede samenstelling van een cluster stijghoogtereeksen is het te verwachten dat de variaties van de Specific Dynamic Factor in veel gevallen klein is en in elk geval kleiner dan de variatie van de residureeks van de enkelvoudige modellering. Extreme waarden en trends in de Specific Dynamic Factor komen uitsluitend voor in de betreffende reeks. Hierdoor zijn meetfouten zoals een individuele uitschieter of een weglappend referentieniveau veel makkelijker te detecteren dan met de residureeks van het univariate tijdreeksmodel. Bovendien kan ook een schatting gemaakt worden hoe de reeks gecorrigeerd zou moeten worden. Overigens zijn natuurlijk niet alle lokale extremen ook meetfouten!

De Common Dynamic Factors maken het ook mogelijk om perioden zonder waarnemingen aan te vullen, ook in de periode vóórdat de betreffende meetreeks begon. Voor dit laatste moeten er voor die perioden uiteraard wel andere stijghoogtereeksen uit het cluster beschikbaar zijn. Ook kunnen de Common Dynamic Factors worden gebruikt om een cluster stijghoogtereeksen te simuleren, waarbij niet alleen de statistieken van de individuele reeksen correct worden gesimuleerd, maar ook de ruimtelijke samenhang tussen de stijghoogtereeksen.

Formulering van het meervoudig tijdreeksmodel

Bij een enkelvoudig tijdreeksmodel wordt een meetreeks van de stijghoogte gesplitst in een deterministisch deel, een ruiscomponent, een referentie niveau en eventueel een meetfout:

$$y_t = d_t + n_t + \mu + \varepsilon_t \quad (1)$$

Waarbij

- y_t : de reeks met de metingen van de stijghoogte
- d_t : het deterministische deel, dat verklaard kan worden uit één of meer inputreeksen
- n_t : de ruiscomponent, ofwel het onverklaarde deel
- μ : het referentie niveau, ook bekend als het lokale drainageniveau
- ε_t : de meetfout

Het deterministische deel kan worden beschreven met responsfuncties:

$$d_t = \sum_{j=1}^p \left[\sum_{i=0}^{\infty} \theta_{j,i} u_{j,t-i} \right] \quad (2)$$

Waarbij

- $\theta_{j,i}$: de waarde van de responsfunctie voor inputreeks j op tijdstip $t-i$
- p : het aantal inputreeksen (verklarende variabelen)
- $u_{j,t}$: inputwaarde (verklarende variabele) op tijdstip

Evenals Von Asmuth (2012) definiëren wij de responsfunctie als een Pearson-type-III-verdeling:

$$\theta_{j,i} = A \frac{a^n (i - b)^{n-1} \exp\{-a(i - b)\}}{\Gamma(n)} \quad (3)$$

Waarbij

- A, a, n en b : parameters, die geschat moeten worden
- $\Gamma(n)$: Gamma functie

Voor de ruiscomponent is gekozen voor het veelgebruikte AR(1)-model:

$$n_t = \phi n_{t-1} + \eta_t \quad (4)$$

Waarbij

- ϕ : de autoregressieve coëfficiënt van het ruisdeel
- η_t : een witte ruis reeks

Als meerdere stijghoogtereeksen binnen het zelfde hydrologische systeem met het hierboven gedefinieerde enkelvoudige model gemodelleerd worden, blijkt dat de ruiscomponenten van de verschillende modellen vaak duidelijke overeenkomsten ver-

tonen. Dit betekent dat er klaarblijkelijk ruimtelijke samenhang aanwezig is die niet met de deterministische delen verklaard kan worden.

We kunnen van deze samenhang gebruik maken door de reeksen simultaan te modelleren met een meervoudig tijdreeksmodel, waarbij we de ruiscomponenten splitsen in ruimtelijk samenhangende patronen, de Common Dynamic Factors, en variaties die uitsluitend in één tijdreeks aanwezig zijn, de Specific Dynamic Factor:

$$n_{i,t} = \left[\sum_{j=1}^M \lambda_{i,j} x_{c,j,t} \right] + x_{s,i,t} \quad (5)$$

Waarbij

- $n_{i,t}$: ruiscomponent van tijdreeks i
- $x_{c,j,t}$: de Common Dynamic Factor j op tijdstip t
- M : het aantal Common Dynamic Factors
- $\lambda_{i,j}$: het gewicht, ofwel de schaalfactor, van Common Dynamic Factor j voor reeks i
- $x_{s,i,t}$: de Specific Dynamic Factor van tijdreeks i op tijdstip t

Voor de overzichtelijkheid beperken wij ons in de verdere formules tot slechts één Common Dynamic Factor. De uitbreiding naar meerdere Common Dynamic Factors is rechttoe-rechtaan. Omdat zowel de Common Dynamic Factor als de Specific Dynamic Factor veelal beschreven kunnen worden met een AR(1)-model, kan vergelijking (5) worden geschreven als:

$$n_{i,t} = \lambda_i [\phi_c x_{c,t-1} + \eta_{c,j,t}] + [\phi_{s,i} x_{s,i,t-1} + \eta_{s,i,t}] \quad (6)$$

Waarbij

- ϕ_c : de autoregressieve coëfficiënt van de Common Dynamic Factor
- $\eta_{c,t}$: een witte ruis reeks van de Common Dynamic Factor
- $\phi_{s,i}$: de autoregressieve coëfficiënt van de Specific Dynamic Factor van tijdreeks i
- $\eta_{s,i,t}$: een witte ruis reeks van de Specific Dynamic Factor van tijdreeks i

Voor N stijghoogtereeksen hebben we $(N+M)$ ruiscomponenten. Om de parameters te kunnen schatten gebruiken we een Kalmanfilter. Hiertoe wordt vergelijking (6) in een state-spaceformulering geschreven, waarbij ook de meetfout wordt meegenomen. We definiëren de toestand op het tijdstip t als een vector met als elementen de Specific Dynamic Factors en de Common Dynamic Factor van N tijdreeksen. Gecombineerd met vergelijking (1) wordt de meetvergelijking:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}_t = \begin{bmatrix} n_1 \\ \vdots \\ n_N \end{bmatrix}_t + \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_t + \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_N \end{bmatrix}_t + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}_t, \quad \begin{bmatrix} n_1 \\ \vdots \\ n_N \end{bmatrix}_t = \begin{bmatrix} \mathbf{I}_N & \lambda_1 \\ & \vdots \\ & \lambda_N \end{bmatrix} \begin{bmatrix} x_{s,1} \\ \vdots \\ x_{s,N} \\ x_c \end{bmatrix}_t \quad (7)$$

Waarbij

- $\mathbf{I}_N$: een $(N \times N)$ eenheidsmatrix

De toestandsvergelijking kan geschreven worden als:

$$\begin{bmatrix} x_{s,1} \\ \vdots \\ x_{s,N} \\ x_c \end{bmatrix}_t = \begin{bmatrix} \phi_{s,1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \phi_{s,N} & 0 \\ 0 & \cdots & 0 & \phi_c \end{bmatrix} \begin{bmatrix} x_{s,1} \\ \vdots \\ x_{s,N} \\ x_c \end{bmatrix}_{t-1} + \begin{bmatrix} \eta_{s,1} \\ \vdots \\ \eta_{s,N} \\ \eta_c \end{bmatrix}_t \quad (8)$$

Voor de schatting van de parameters $(\phi_{s,1}, \dots, \phi_{s,N}, \phi_c)$ en de varianties van de witte ruis reeksen $(\eta_{s,1,t}, \dots, \eta_{s,N,t}, \eta_{c,t})$ maken we gebruik van de loglikelihood functie in combinatie met het Kalmanfilter (Berendrecht, 2004). De gewichten $(\lambda_1, \dots, \lambda_N)$ worden geschat op basis van eigenwaardedecompositie van de correlatiematrix van de ruiscomponenten.

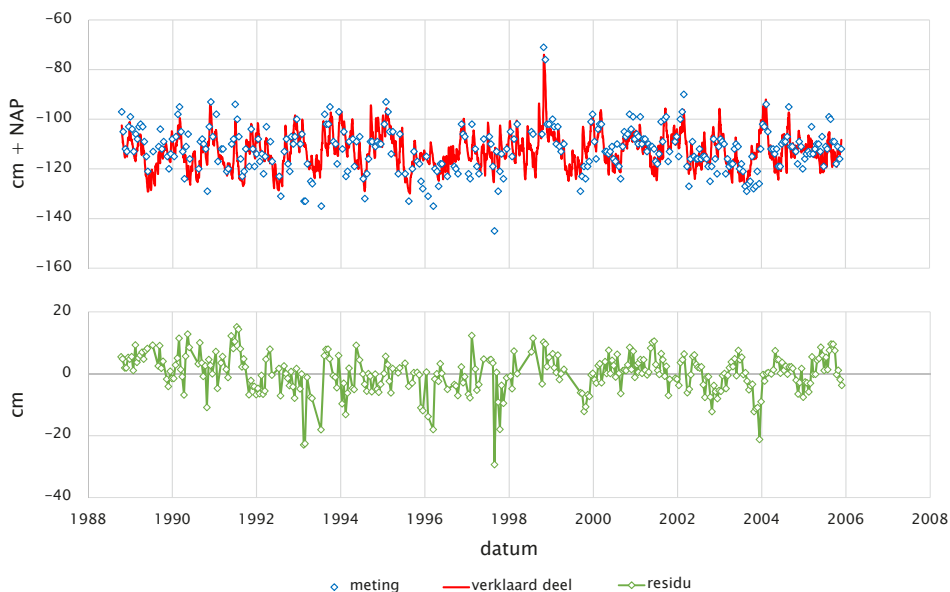
Praktijkvoorbeeld: detectie mogelijke meetfouten

In dit artikel demonstreren we de werking van het meervoudig tijdreeksmodel door één praktijktoepassing uit te werken, namelijk die van het controleren van meetreeksen op mogelijke meetfouten. We doen dit aan de hand van de meetreeksen van vijf onder elkaar liggende meetfilters met identificatie B21B0214001-005. Het meetpunt ligt even ten westen van Zwartsluis in Overijssel. De filters bevinden zich op een diepte tussen de 7 en 53 meter beneden maaiveld.

Zoals beschreven in de vorige paragraaf hebben we allereerst de vijf reeksen gemodelleerd met dagcijfers van neerslag en verdamping als verklarende variabelen. Omdat binnen de DINO-omgeving de analyses veelal automatisch zullen worden uitgevoerd, worden vooraf extreme uitbijters ten behoeve van de parameterschatting verwijderd. Hierbij kwalificeren we arbitrair een meetwaarde als uitbijter, indien de waarde een afwijking van het gemiddelde heeft van meer dan vijfmaal de standaardafwijking. Met name bij korte reeksen kunnen deze uitbijters namelijk een onevenredig grote invloed hebben op de schattingen.

Afbeelding 1 geeft voor het onderste filter de resultaten van de eerste stap. De resultaten voor de overige filters zijn vergelijkbaar. De reeksen kunnen voor 53-60% verklaard worden uit neerslag en verdamping. Aangezien de focus van dit artikel ligt op de beschrijving van de Common en Specific Dynamic Factors en niet zozeer op het uit neerslagoverschot verklaarde deel, zullen we niet verder ingaan op de resultaten van de univariate transfermodellen, maar deze resultaten als uitgangspunt nemen voor de verdere analyse.

De tweede stap in de meervoudige analyse is het vaststellen hoeveel van de variatie in de ruiscomponenten beschreven kan worden met één of meerdere Common Dynamic Factors. Uit deze analyse blijkt dat één Common Dynamic Factor voldoende is om de gezamenlijkheid in de ruiscomponent te modelleren en dat met deze factor 90% van de variatie in de ruiscomponenten beschreven kan worden. Tabel 1 toont per meetreeks de geschatte waarde van de factorgewichten $\lambda_1, \dots, \lambda_N$ uit vergelijking (7). Deze gewichten geven aan in welke mate een Common Dynamic Factor doorwerkt in een reeks. In de tabel zijn de gewichten min of meer gelijk, wat aangeeft dat de Common Dynamic Factor overal nagenoeg even sterk doorwerkt. Indien een gewicht relatief klein zou zijn, dan zou dit betekenen dat deze ruiscomponent nauwelijks overlap



Afbeelding 1: Modelresultaat voor B21B0214005 met neerslag en verdamping als verklarende variabelen

heeft met de overige reeksen, of andersom geredeneerd, dat het een hoge unieke waarde heeft en dus veel informatie aan het cluster toevoegt.

In Tabel 1 is naast de factorgewichten ook de uniciteit per meetreeks aangegeven. Deze wordt berekend als:

$$\text{uniciteit}_i = \frac{\text{var}(x_{s,i})}{\text{var}(n_i)} \times 100\% \quad (9)$$

Waarbij

$\text{var}(x_{s,i})$: de variantie van de Specific Dynamic Component van tijdreeks i

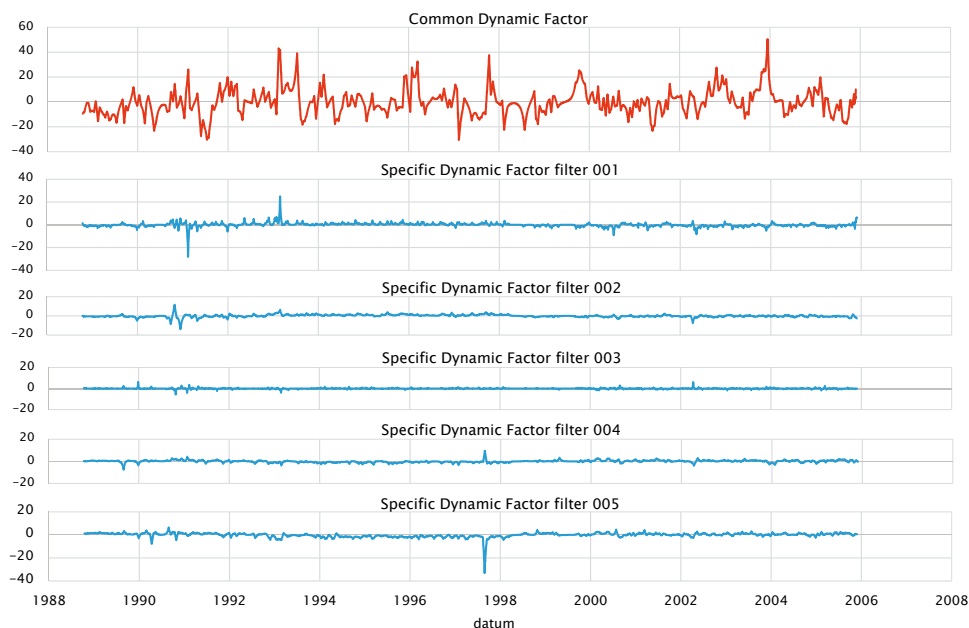
$\text{var}(n_i)$: de variantie van de ruiscomponent van tijdreeks i

Tabel 1: Factorgewichten en uniciteit voor de filters B21B0214001-005.

Filter identificatie	Factorgewicht λ	Uniciteit
B21B0214001	0,43	27%
B21B0214002	0,45	10%
B21B0214003	0,46	3%
B21B0214004	0,46	6%
B21B0214005	0,44	18%

Uit de tabel blijkt dat er tussen de filters nog redelijk wat verschil is in de uniciteit. Filter 003 heeft de laagste waarde van 3%, terwijl filter 005 en 001 een waarde hebben van respectievelijk 18% en 27%. Om meer inzicht hierin te verkrijgen toont Afbeelding 2 de Common Dynamic Factor en de vijf Specific Dynamic Factors. De afbeelding laat duidelijk zien dat er in het algemeen nauwelijks nog variatie in de

Specific Dynamic Factors zit, behalve een enkele uitschieter. Deze uitschieters zijn het grootst in reeks 001 en 005, wat gelijk de relatief hoge uniciteit van deze reeksen verklaart. Deze uitschieters zijn in het oorspronkelijke residu (zie Afbeelding 1 voor het residu van filter 005) niet of nauwelijks waar te nemen en zijn ook niet als 'statistisch significante' uitbijter aan te duiden. In de Specific Dynamic Factor vallen deze uitbijters echter overduidelijk op en kunnen ook statistisch als uitbijter worden beschouwd met een waarde van ruim meer dan 5 maal de standaardafwijking, voor enkele zelfs 10-20 maal de standaardafwijking.



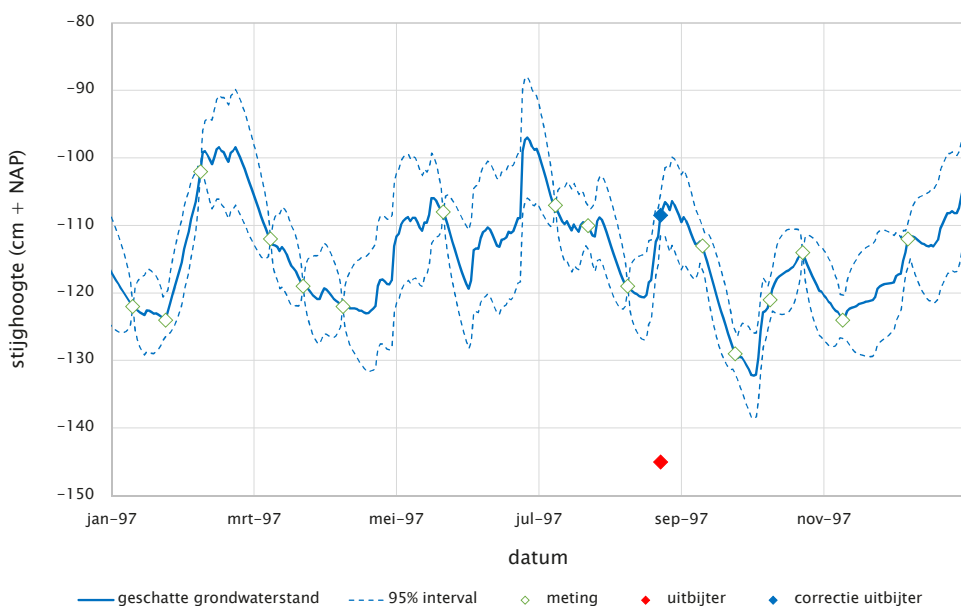
Afbeelding 2: Common Dynamic Factor en Specific Dynamic Factors voor de vijf meetfilters

Voor de meetwaarden die als uitbijter (mogelijke meetfout) worden gekwalificeerd, kan met behulp van de informatie uit de verklarende variabelen neerslag en verdamping, het eigen ruisproces (Specific Dynamic Factor model) én de informatie uit de andere meetfilters (via het Common Dynamic Factor model), een schatting worden gemaakt van de meest waarschijnlijke grondwaterstand/-stijghoogte op dat tijdstip, inclusief een schatting van een betrouwbaarheidsinterval. Het verschil met een enkelvoudige model is dat we, bovenop een schatting van het verklaarde deel op tijdstip t en een voorspelling van de ruiscomponent, nu ook op tijdstip t metingen van de andere locaties kunnen gebruiken. De informatie uit deze metingen kunnen we via de Common Dynamic Factor direct gebruiken bij het schatten van de grondwaterstand. Vergeleken met een univariate schatting resulteert dit in zowel een betere schatting als in een kleinere betrouwbaarheidsband.

Voor het verkrijgen van een optimale schatting kunnen we gebruik maken van het Kalmanfilter. Indien we een meting verder terug in de tijd analyseren en er reeds recentere metingen beschikbaar zijn, kunnen we ook de Kalman-smoother gebruiken. In tegenstelling tot het Kalmanfilter maakt de Kalman-smoother ook gebruik van de

meetinformatie ná tijdstip t voor een toestandschatting op tijdstip t . Degenen die geïnteresseerd zijn in meer informatie over de Kalman-smoother verwijzen we naar Berendrecht (2004) en Jazwinski (1970).

Afbeelding 3 toont een illustratie van uitbijtercorrectie. Het betreft hier de uitbijter die we in de Specific Dynamic Factor van filter 005 hebben waargenomen op 28 augustus 1997; zie de onderste figuur van Afbeelding 2. Op deze datum heeft de Specific Dynamic Factor een waarde van 18 keer de standaardafwijking. In de afbeelding is de gemeten waarde, welke is gekwalificeerd als uitbijter, weergegeven tezamen met een schatting van de gecorrigeerde stijghoogte op basis van het meervoudige tijdreeksmodel. Tevens wordt een schatting van het bijbehorend 95%-betrouwbaarheidsinterval getoond. De gemeten waarde van 145 cm - NAP wordt hiermee gecorrigeerd naar een geschatte stijghoogte van 108 cm - NAP met een standaardafwijking van 3 cm.



Afbeelding 3: Correctie van uitbijter met behulp van meervoudig tijdreeksmodel

Deze methode voor het corrigeren van een uitbijter kan ook rechtstreeks worden toegepast voor het opvullen van een of meerdere missende waarnemingen in een reeks. Ook dan wordt informatie uit omliggende c.q. boven-/onderliggende meetfilters gebruikt om via één of meerdere Common Dynamic Factors de grondwaterstand of stijghoogte te schatten.

Het is overigens uitdrukkelijk niet de bedoeling van de Geologische Dienst Nederland-TNO om automatische correcties uit te gaan voeren of missende waarnemingen aan te vullen. Bovenstaande berekeningen kunnen wel als ondersteuning dienen bij kwaliteitscontrole. De beslissing ten aanzien van correcties ligt principieel bij de bronhouder van de stijghoogtemetingen.

Discussie

De belangrijkste reden voor de Geologische Dienst Nederland-TNO om programma's te ontwikkelen voor meervoudig tijdreeksmodellen, inclusief de meervoudige modellering, is het ontwikkelen van informatieproducten betreffende het karakter van de stijghoogtereeksen uit DINO. Dit zullen ten dele bekende karakteristieken zijn, zoals GHG, GLG, GT's enz., maar ook karakteristieken die informatie geven over trends in de tijd, responskarakteristieken en ruimtelijke samenhang c.q. uniciteit of informatieve waarde van een meetlocatie. Daarbij gaat het primair om een geautomatiseerde toepassing, waarbij de informatie op soortgelijke wijze uitgeleverd kan worden als de huidige stijghoogte informatie op DINO-loket. De inhoud en presentatievorm van deze informatieproducten is nog niet volledig uitgekristalliseerd. Belanghebbenden en geïnteresseerden zijn dan ook van harte welkom om mee te denken wat er uiteindelijk als informatieproduct via de website van de Geologische Dienst Nederland-TNO beschikbaar kan komen.

Het gebruik van een tijdreeksmodel in een geautomatiseerde workflow geeft geen garantie dat dit ook het best mogelijke model is. De criteria waarop een tijdreeksmodel in het geautomatiseerde gebruik al dan niet wordt geaccepteerd en in hoeverre een gebruiker daar invloed op kan uitoefenen is momenteel onderwerp van nader te onderzoeken. Het blijft uiteraard altijd mogelijk om de modellering op projectbasis handmatig uit te voeren.

Naast het ontwikkelen van informatieproducten kunnen meervoudige tijdreeksmodellen een rol spelen in de kwaliteitscontrole van stijghoogtereeksen. In het rapport "Kwaliteitsborging grondwaterstands- en stijghoogtegegevens Systematiek en methodiek voor datakwaliteitscontrole (QC)" (Von Asmuth en Van Geer, 2015) is de fase 'plausibiliteitscontrole' gedefinieerd. Hierbij gaat het erom te beoordelen in hoeverre een waarneming past bij het karakter van de reeks zelf én in het ruimtelijke beeld van de stijghoogte in de omgeving. De Specific Dynamic Factor lijkt daarbij een geschikt hulpmiddel, omdat extremen in de Specific Dynamic Factor niet overeenkomen met het patroon in andere reeksen in de cluster. Het daadwerkelijke besluit of het al dan niet een fout betreft, blijft altijd ter beoordeling van een expert en dient niet te worden geautomatiseerd.

De Common Dynamic Factors geven aan of er ruimtelijke patronen in de stijghoogte zijn te onderkennen, zowel in horizontale zin als in verticale zin. De schalingsfactor geeft daarbij aan in hoeverre er een verloop is in de sterkte van deze patronen. Ruimtelijke patronen kunnen inzicht geven in verwachte of onverwachte fenomenen in de stijghoogte, bijvoorbeeld een signalering van een trend in de stijghoogte. Ook bevatten de Common Dynamic Factors informatie over de hydraulische eigenschappen van de ondergrond, met name de aan- of afwezigheid van weerstanden. Op welke wijze deze informatie benut kan worden is nog een punt van nader onderzoek.

Conclusies

Het gegeven voorbeeld en eerdere studies (Berendrecht, 2004) laten zien dat de toepassing van meervoudige tijdreeksmodellen op clusters van stijghoogtereeksen een meerwaarde kan leveren bij de analyse van de stijghoogtedynamiek. Hoe groot die meerwaarde is hangt af van de samenstelling van de clusters en is onderwerp van

verdere studie. Door de splitsing van de residueeksen in Common en Specific Dynamic Factors ontstaat een beeld welke variaties ruimtelijk samenhangen en welke locatie specifiek zijn. De ruimtelijk samenhangende patronen kunnen worden gebruikt in een verder analyse betreffende de respons van de stijghoogte op bekende en onbekende oorzaken. Bovendien bevatten de ruimtelijke patronen ook aanwijzingen ten aanzien van de hydraulische eigenschappen van de ondergrond en kunnen de Common Dynamic Factors gebruikt worden voor het opvullen van perioden zonder waarnemingen. De Specific Dynamic Factor geeft de lokale variaties van de stijghoogte die op geen enkele wijze samenhangen met de inputreeksen en de overige stijghoogtereeksen. Zoals in dit artikel is geïllustreerd, is de Specific Dynamic Factor bij een goede samenstelling van de cluster een geschikt hulpmiddel voor plausibiliteitstesten bij kwaliteitsborging. Bovendien geven de Specific Dynamic Factors ook aan welke stijghoogtereeksen het meeste overlap vertonen met omliggende stijghoogtereeksen en dus het meeste redundante informatie bevatten. Hiervan kan gebruik gemaakt worden bij meetnetoptimalisaties, of bij de prioriteitstelling voor het installeren van automatische dataloggers.

Literatuur

Berendrecht, W.L. (2004) State space modelling of groundwater fluctuations; Ph.D thesis TUD. ISBN 90-9018342-6.

Box, G.E.P. en Jenkins, G.M., (1976) *Time series analysis, forecasting and control. Revised Edition*; Holden-Day, San Francisco.

Camacho, F., McLeod, A. en Hipel, K. (1987) *Multivariate contemporaneous ARMA model with hydrological applications Stochastic Hydrology and Hydraulics*; Springer-Verlag, vol 1, pag 141-154.

Deutsch, S.J. en Pfeifer, P.E. (1981) Space-Time ARMA Modeling with Contemporaneously Correlated Innovations; *Technometrics*, vol 23, pag 401-409.

Geweke, J.F. (1977) The dynamic factor analysis of economic time series models; in Aigner, D. en Goldberger, A. (eds.), *Latent Variables in Socioeconomic Models*, pag 365-383, Amsterdam.

Jazwinsky, A.H. (1970) *Stochastic processes and filtering theory*; Ac. Press, New York.

Stock, J. en Watson, M. (2010) Dynamic Factor Models; in Clements, M.P. en Henry, D.F. (eds.), *Oxford Handbook of Economic Forecasting*, Oxford University Press.

Stoffer, D.S. (1986) Estimation and identification of space-time ARMAX models in the presence of missing data; *J. Am. Stat. Assoc.*, vol 81, pag 762-772.

Van Geer, F.C., Baggelaar, P.K. en Defize, P.R. (1988) Toepassing van tijdreeksanalyse op meetreeksen van de stijghoogte; *H₂O*, vol 2, nr 16 pag 438-442.

Van Geer, F.C. en Zuur, A.F. (1997) An extension of Box Jenkins transfer/noise models for spatial interpolation of groundwater head series; *J. Hydrol.*, vol 192 pag 65-80.

Von Asmuth, J. (2012), Groundwater System Identification, through Time Series Analysis; Ph.D thesis TUD. ISBN 978-90-5155-079-5.

Von Asmuth, J. en Van Geer, F.C. (2015) Kwaliteitsborging grondwaterstands- en stijg-hoogtegegevens Systematiek en methodiek voor datakwaliteitscontrole (QC); KWR 2015.004

Zuur, A.F., Fryer, R.J., Jolliffe, I.T., Dekker, R. en Beukema, J.J. (2003) Estimating common trends in multivariate time series using dynamic factor analysis; *Environmetrics*, vol 14, pag 665-685