



Bijeenkomst: NHV discussiemiddag tijdreeksanalyse – modelbeoordeling  
Vergadering nummer: 1  
Secretaris: Koen van der Hauw, Arjen Roelandse  
Datum: 2013-06-27  
Stuknummer: 1

Aanwezig: CC:

Zie bijlage

## **NHV Discussiemiddag Tijdreeksanalyse - Modelbeoordeling**

Locatie: Kantoor Houdering De Bilt, Grontmij, 14:00 – 16:30

### **Programma**

14.00 - 14.10 Opening; scope van de middag  
14.10 - 14.20 Inleiding Martin Knotters  
14.20 - 14.40 Inleiding Jos von Asmuth  
14.40 - 15.00 Werkpauze  
15.00 - 16.00 Discussie adhv stellingen  
16.00 - 16.15 Conclusies / Vervolg  
16.15 - 16.30 Wvttk / sluiting

### **Inleiding**

Frans van Geer van de Universiteit Utrecht verzorgt de inleiding en zal discussieleider zijn deze middag. Frans verwijst naar de LASSA-traditie. Wellicht kunnen deze discussiemiddagen een ervaring hiervan betekenen. De strekking is: informeel, geen gladde verkoop verhalen, meer discussie dan presentatie en toegankelijkheid voor zowel experts als beginnende gebruikers. Frans noemt nog de beschikbaarheid van de CHO boekjes die via de NHV-site te downloaden zijn: <http://www.nhv.nu/cho-publicaties>.

### **Presentatie Martin Knotters**

#### **“Beoordeling van tijdreeksmodellen, theorie? Een pragmatisch geluid!”**

Martin opent met de volgende constatering: “Theorie is als men alles weet en niets klopt”, “Praktijk is als alles functioneert en niemand weet waarom” en de synthese “Vaak zijn theorie en praktijk verenigd: niets klopt en niemand weet waarom!”. Hij heeft gekozen voor een presentatie over pragmatische modelbeoordeling. Het pragmatisme blijkt overigens een filosofische stroming die focust op het verbinden van praktijk en theorie: “De praktijk is de toetssteen voor de geldigheid van de theorie”. Hij benoemt de volgende categorieën voor beoordeling van tijdreeksmodellen:

- Goodness-of-fit: Past het model bij de waarnemingen?

Criteria: Percentage Verklaarde Variantie (EVP), RMSE, etc.

Martin hoopt dat de magische grens in Menyanthes van 70% vanmiddag zal worden verklaard.

- Diagnostic checks: Klopt het model in theorie? Of, voldoet het model aan de modelveronderstellingen? Bijvoorbeeld: autocorrelatie moet niet significant afwijken van nul, normaliteit, variantie constant, etc. Vaak gaat het hierbij om de residureeks waarvan iedereen hoopt dat ie wit is...
- Fysische plausibiliteit: Komt het model overeen met hoe ik denk dat de natuur in elkaar zit? Martin geeft hierbij de waarschuwing dat je kennis beperkend kan zijn en dat de data mogelijk niet leiden niet tot nieuwe inzichten als je toetst op fysische plausibiliteit.
- Validatie  
Er zijn verschillende opties:
  - de scientists rule, stelt de vraag of het model klopt met de theorie
  - de engineers rule, stelt de vraag of het model klopt met de praktijk. Bohlin stelt: validatie is een testprocedure, waarvan de uitkomst aangeeft of een model voldoet aan zijn doel
  - Chatfield ziet validatie als model checking, hoe goed kwantificeert het model de onzekerheid (gegeven model en data)? Hij adviseert het model te valideren op een volledig nieuwe reeks (dus ook op een nieuwe locatie), 'true replication'. In de praktijk gaat dit niet en resulteert dit in data splitting, hold-out samples: onvermijdelijk bij validatie van tijdreeksmodellen.
  - De volgende validatiematen worden genoemd: systematische fout (gemiddelde van de fouten), toevallige fout (standaarddeviatie van de fouten), samenvattende maat (RMSE), percentage waarneming buiten betrouwbaarheidsinterval (P).
  - In de presentatie wordt een grafiek getoond (sheet 19) waarin een slecht validatieresultaat zichtbaar is. Vermoedelijk heeft dit te maken met een wisseling naar hoogfrequente divergingen halverwege de kalibratiereeks. Wellicht een thema voor een vervolgdiscussie...
  - Het is van het belang de validatie af te stemmen op het doel van de tijdreeksanalyse, bijvoorbeeld het bepalen van overschrijdingsduurlijnen.

Martin sluit af met de volgende twee conclusies/stellingen:

- Praktische bruikbaarheid van modellen kan niet worden beoordeeld op basis van 'goodness-of-fit', diagnostic checks en fysische plausibiliteit
- Praktische bruikbaarheid van statistische informatie (overschrijdingsrisico's etc.) kan alleen door validatie met onafhankelijke waarnemingen worden getoetst

### **Presentatie Jos van Asmuth**

#### **"Modelbeoordeling vanuit statistiek, fysica en Menyanthes"**

- Jos begint met het benoemen van een aantal mogelijke toepassingen van tijdreeksanalyse: analyse effecten achteraf, voorspellen effecten vooraf, foutendetectie, opvullen / verlengen reeksen, klimaatcorrectie / statistieken, meetnetoptimalisatie, ruimtelijke grondwatermodellen
- Stelling: Kijk naar de toepassing/doel om de aanpak voor methodebeoordeling te bepalen.
- Jos laat de stappen zien die onderdeel zijn van de modelbeoordeling
- Zelfs als aan alle modelveronderstellingen is voldaan en er een statistisch verband is, is nog steeds de vraag of er ook een causaal verband is. Het is daarom van belang ook te controleren met theoretische formules. In het onderzoek van Jos bij het Veerse Meer bleek dat het tijdreeksmodel

een doorwerking met een 3 á 4 keer grotere spreidingslengte liet zien (=maat voor de afstand waarover een effect doorwerkt) dan berekend met de bekende KD en c-waarden volgens de analytische oplossing daarvoor.

- De ruimte voegt een (of meer) dimensies toe aan modelbeoordeling  
Hiermee wordt bedoeld dat het ruimtelijk analyseren van de resultaten van meerdere tijdreeksmodellen (dus meerdere locaties), extra informatie geeft m.b.t. modelbeoordeling. Jos geeft hiervan een aantal voorbeelden, zie hiervoor zijn powerpointpresentatie. Een mogelijkheid is het filteren op berekende standaarddeviaties, ofwel enkel de meest betrouwbare parameters bekijken. Later wordt vanuit de zaal nog gewaarschuwd dat de resultaten op verschillende locaties vaak zijn gecorreleerd.
- Jos laat nog een aantal verbeteringen aan Menyanthes zien waarbij een waarschuwing wordt gegeven wanneer modelparameters niet in orde lijken. Daarnaast is een plot toegevoegd van de residuen tegenover de waarnemingen. Wanneer deze plot nog een patroon toont, is er meestal nog iets niet in orde. In het voorbeeld is sprake van niet-lineaire data en een lineair model. Een niet-lineair model geeft een betere plot. Zo zijn er meer van dergelijke verbeteringen gepland.
- De ondergrens voor de EVP van 70%, blijkt overigens een niet onderbouwde vuistregel van Jos toen hem in het verleden eens gevraagd werd een getal hiervoor te noemen.

#### **Vragen en discussie n.a.v. presentaties en stellingen**

Zie het einde van dit verslag voor een overzicht van de stellingen en de eventuele conclusies.

Op enkele stellingen is tijdens de middag en hieronder nader ingegaan.

- Stelling 6/7: Het ruismodel is niet/wel van belang voor modelbeoordeling  
Er valt iets voor te zeggen om het ruismodel te negeren. Er is namelijk altijd wel iets mis met het ruismodel zodat je altijd via het ruismodel het model kan afkeuren, mogelijk onterecht. Aan de andere kant geeft het ruismodel wel informatie over de onzekerheid van uitspraken.  
Daarbij kan een goed model kan nog steeds veel ruis bevatten. Je wil echter voorkomen dat er nog correlatie is tussen ruismodel en transfermodel. Het ruismodel wordt bij de beoordeling van tijdreeksmodellen nog wel toegepast, bij de toepassing ervan minder.  
Een andere mogelijkheid voor modelbeoordeling is het model valideren. Daarbij moet je wel data-splitting toepassen en daar zijn vaak te weinig gegevens voor.  
Tot slot moet je een model niet alleen met statistiek beoordelen. Je hebt een eigen conceptueel model in je hoofd. Je moet je ervan bewust zijn dat je juist met dat model ook je conceptuele model toetst en andersom.
- Stelling 4: Modellen moeten vaker worden afgekeurd  
Dat is een centrale vraag van deze middag. Er kunnen ook juist meerdere modellen zijn die een plausibel antwoord geven. Deze zouden tot hetzelfde resultaten moeten leiden. Als dit niet zo is, is er te weinig informatie aanwezig om eenduidig een model af te leiden. Daarnaast is meer aandacht voor modelbeoordeling duidelijk gewenst. Zowel statistiek als fysische plausibiliteit moeten hierbij samengaan. Beiden moeten kloppen. In de praktijk worden zelden meerdere modellen geanalyseerd. Er is vaak druk vanuit opdrachtgevers om juist één enkel model te accepteren. Als dat

het geval is, zou je de resultaten daarvan enkel met veel mitsen en maren moeten presenteren. Opdrachtgevers willen dat liever niet horen, maar de opdrachtgever moet het besluit nemen en niet de adviseur. De adviseur moet niet zelf keuzes gaan maken voor de opdrachtgever. De adviseur moet dus wel de onzekerheden geven. Je moet de opdrachtgever natuurlijk wel door de resultaten, statistiek, etc. heenleiden: De opdrachtgever zal vermoedelijk niet in staat zijn zelf de randvoorwaarden van statistische toetsen te bepalen.

Dat er in ons vakgebied niet vaak met mitsen en maren wordt gewerkt kan komen doordat onzekerheid meestal niet zo'n punt is. Op veel andere beleidsterreinen is dat wel zo, zoals bij medicijnen. Daar wordt wel een bandbreedte opgegeven.

De vraag is hoe je statistische informatie moet presenteren richting opdrachtgevers? Dan moet je weten hoe een opdrachtgever dit leest. Bijvoorbeeld het 95%-betrouwbaarheidsinterval wordt vaak verkeerd begrepen.

- Extra stelling 1: Om een TRA te kunnen doen heb je geen statistische kennis nodig mits de statistische expertise in de software is verwerkt. (je moet natuurlijk wel hydroloog zijn)  
Kunnen we bijvoorbeeld niet voor de paar doelstellingen die er zijn, concreet de juiste toetsen uitwerken? Wat kan er door software gebeuren en wat niet?  
Er ontstaat hier veel discussie over: Hydrologen behoren een basiskennis statistiek te hebben, zeker als ze met statistische pakketten gaan werken.  
De verantwoordelijkheid ligt wel bij de tijdreeksanalisten om de opdrachtgevers de noodzakelijke statistiek uit te leggen. Wellicht kan het in sommige gevallen wel en in andere gevallen niet en is de software wel in staat aan te geven wanneer het wel kan. Met software kan in ieder geval een stap vooruit worden gemaakt. Misschien zal het niet perfect worden, maar kan wel het risico op foute modellen en verkeerde uitspraken verder worden beperkt t.o.v. de huidige situatie. Misschien is het mogelijk eens een aantal praktijkgevallen concreet uit te werken om te zien waar je tegen aanloopt. Voor sommige doelen, zoals opvullen van gaten, zijn de statistische controles minder complex. Soms is de kans op een succesvolle tijdreeksmodellering te koppelen aan het type systeem.

Extra stelling 2 & 3: Tot welke verhoging is een effect van een maatregel normaliter vast te stellen? Bijvoorbeeld tot 10 cm?

Er kan geen algemene uitspraak hierover worden gedaan. Ter indicatie, vaak lukt aantonen van een effect van 10cm nog wel en als je een zeer goed model hebt en een situatie waar er een zeer sterk verband is met neerslag, dan zou 5cm ook nog wel kunnen lukken. Het is globaal mogelijk de doelen of resultaatverwachtingen te koppelen aan het type systeem (zie ook de vorige stelling). Echter, alleen als de statistische proefopzet klopt, kan je concluderen dat wanneer er geen effect wordt gevonden, dat dit betekent dat er inderdaad geen effect is.

Voor het opzetten van je meetnet moet je vooraf dus nadenken over hoe groot het risico mag zijn dat je te onrechte concludeert dat er géén effect is of dat er juist een effect is. Dat bepaalt hoe nauwkeurig je moet meten en analyseren.

#### **Afsluiting en suggesties voor vervolg**

- De middag is geslaagd en 80% van de deelnemers wil dit over een paar maanden weer doen



- De discussie is zinvol geweest waarbij er behoefte is om dit breder te trekken naar:
  - oppervlaktewater
  - tijdreeksanalyse in vlakdekkende karteringen.
  - hoogfrequentie reeksen, onzekerheden.
  - statistiek in de hydrologie
- Daarnaast is er de wens een soort handboek op te stellen (in kleiner verband)
- Bram: Er wordt een update gemaakt voor het Grondwaterzakboekje. Statistiek zal hierin worden opgenomen. Dit kan wellicht een soort brugfunctie vervullen tussen de hydrologen en de statistici. Laat het weten als er interesse is hierin mee te denken.
- Een tijdreeksanalyse contest (zoals door de NHV voor de Hupselse beek is gedaan).
- De stellingen nader bediscussiëren op een LinkedIn forum/discussiegroep. Jos zal hiervoor een subgroep starten binnen de NHV groep op LinkedIn.
- Een richtlijn/checklist (per doelstelling) uitwerken voor tijdreeksmodellering. Dit kan uitmonden in een soort good-modelling-practice voor tijdreeksanalyse.
- Nagaan hoe je onzekerheden dient te presenteren richting opdrachtgevers

### **Bijlage: Stellingen**

1. (Be)oordeel niet standaard, maar m.b.t. de geschiktheid voor het doel
2. De ruimte voegt één (of meer) dimensies toe aan modelbeoordeling
3. Meetreeksen zijn meestal niet verkregen volgens regels van statistische proefopzet
4. Modellen moeten vaker worden afgekeurd
5. De berekende betrouwbaarheid geldt gegeven het model en de gebruikte gegevens
6. Het ruismodel is niet van belang voor modelbeoordeling
7. Een modelresiduënanalyse is onontbeerlijk bij een modelbeoordeling.
8. Om de modelbetrouwbaarheid te verhogen is het mogelijk niet verklaarde delen uit de meetreeks te verwijderen
9. Een statistische significante relatie zegt niets over een werkelijk, fysisch verband (tussen verklarende invloed en grondwaterstand)
10. Praktische bruikbaarheid van modellen kan niet worden beoordeeld op basis van 'goodness-of-fit', diagnostic checks en fysische plausibiliteit
11. Praktische bruikbaarheid van statistische informatie (overschrijdingsrisico's etc.) kan alleen door validatie met onafhankelijke waarnemingen worden getoetst
12. Geef meer aandacht aan falsificatie dan aan verificatie van de resultaten, volgens het principe 'houd van de data en wantrouw je modellen'

### Extra vragen/stellingen tijdens de discussiemiddag:

13. Om een TRA te kunnen doen heb je geen statistische kennis nodig mits de statistische expertise in de software is verwerkt. (je moet natuurlijk wel hydroloog zijn)
14. Tot welke verhoging is een effect van een maatregel normaliter vast te stellen? Bijvoorbeeld tot 10 cm?
15. Hoe bepaal je of een peilbuis op de goede plek staat? Wanneer besluit je bijvoorbeeld om een buis niet te plaatsen? Denk aan aanwezigheid drogere sloot, boom of weg
16. Kun je grondwaterstanden voorspellen door modellen te simuleren met een andere reeks aan verklarende variabelen? Zo ja, kan Menyanthes dat?

### **Conclusies stellingen**

- Stelling 1: Hier zijn de meesten het over eens.
- Stelling 2: Hier zijn de meesten het over eens. Let wel op mogelijke correlatie tussen verschillende locaties.
- Stelling 3: Dit lijkt een gegeven, het is dan niet mogelijk uitspraken over het ontbreken van effecten te doen. Dit vraagt om nadere discussie.
- Stelling 4: Hier lijken de meesten het over eens, een alternatief is het beschrijven van de mitsen, maren en onzekerheden van het model."
- Stelling 5: Dit is een feit. Het is daarom van belang meerdere modellen te bekijken. Vaak is hier echter geen ruimte voor vanuit de opdrachtgever, daarom dien je onzekerheden te vermelden. De manier waarop en mate waarin dit zou moeten gebeuren vraagt nog meer discussie.



- Stelling 6/7: Dit hangt af van het doel.
- Stelling 8: Dit is niet aan de orde gekomen.
- Stelling 9: Hier zijn de meesten het over eens. Het is daarom van belang ook te kijken naar fysische plausibiliteit en wat de fysische informatie en/of formules zeggen over de situatie.
- Stelling 10: Dit vraagt nog meer discussie.
- Stelling 11: Dit vraagt nog meer discussie.
- Stelling 12: Dit vraagt nog meer discussie..
- Stelling 13: De statistici zijn van mening dat dit in het algemeen niet mogelijk. Of het in deelgevallen wel mogelijk is vraagt meer discussie/onderzoek.
- Vraag 14: Op de vraag of effecten met een bepaalde grootte (zeg 10 cm) significant zijn vast te stellen met tijdreeksanalyse bestaat geen eenduidig antwoord, maar alleen ervaringscijfers. Volgens veel aanwezigen zijn effecten van 10 cm (soms ook minder) vaak wel vast te stellen, mits voldaan is aan de regels van statistische proefopzet.
- Vraag 15: Het kiezen van de locatie valt buiten het onderwerp van tijdreeksanalyse. Hiervoor zijn al wel richtlijnen bekend. De onderliggende vraag is echter of het voor een bepaald onderzoek wel nuttig is te investeren in het plaatsen van een buis. Dit zal in de praktijken afhangen van het specifieke geval en de overige beschikbare informatie.
- Stelling 16: Dit is in principe mogelijk . Gezien de toevoeging (Menyanthes) is dit meer een softwarematige wens dan een stelling.

**Bijlage : aanwezigheid**

<b>Naam</b>	<b>Organisatie</b>
Bram Bot	Bot Raadgevend Ingenieur
Jet Lebbink	Gemeente Hilversum
Marc Vissers	Grontmij
Harry Boukes	Adviesburo Harry Boukes
Ir. N.G.J (Nicko) Straathof	Natuurmonumenten 's-Graveland
Johnny van Keulen	Waterschap Aa en Maas
Chris van Rens	Waterschap Aa en Maas
Jos Moorman	Waterschap Aa en Maas
Jelle Buma	Deltares
Daniel van Buren	Provincie Fryslan
Eit vd Meulen	AMO
Lucas Borst	Grontmij
Stefanie Bus	TTE Consultants
Arnoud Soetens	Waterschap Peel en Maasvallei
Henk Krajenbrink	Wareco
Diepak Mohan	MOS Grondmechanica
Mathijs Oudega	Wiertsema & Partners
Rene van den Hoven	Wiertsema & Partners
Christophe Obergfell	KWR
Joanne van der Spek	TU Delft
Jeroen van Uden	Grontmij
Willem Jan Zaadnoordijk	KWR
Frans Schaars	Artesia
Igor Mendizabal	PWN
<b>Aanwezig vanuit de werkgroep:</b>	
Frans van Geer	Universiteit Utrecht
Jos von Asmuth	KWR
Martin Knotters	Alterra
Paul Baggelaar	Icastat
Mark Bakker	TU Delft
Koen van der Hauw	Grontmij
Arjen Roelandse	Oasen