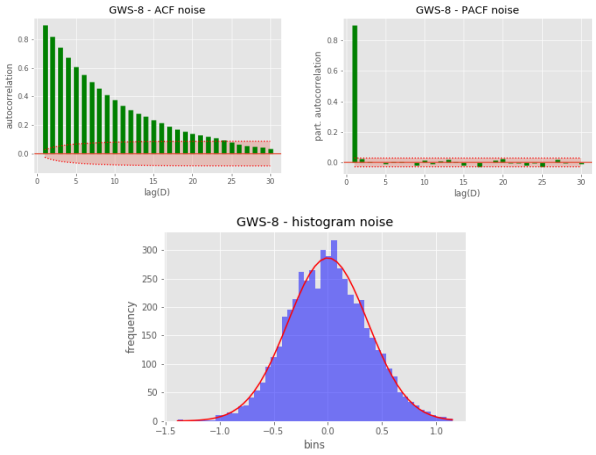
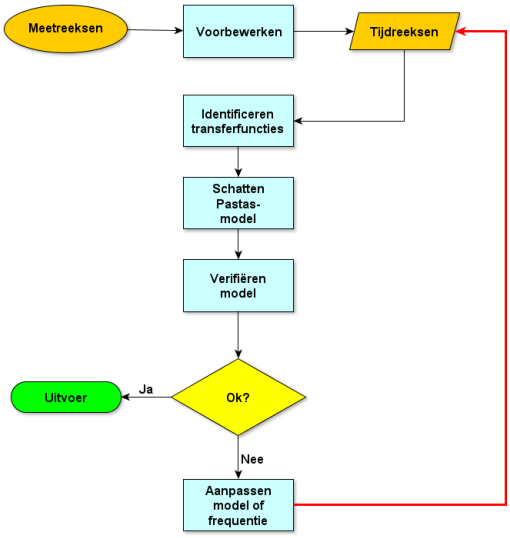
# Naar betere tijdreeks-modellering met Pastas




Modellering en Optimalisatie

# Naar betere tijdreeks-modellering met Pastas

Opdrachtgever:
Provincie Overijssel (ir. Thomas de Meij)

Auteurs:
drs. Paul Baggelaar (*PB Icastat*) en ir. Eit van der Meulen (*AMO*)

28 april 2020

Modellering en Optimalisatie

PB Icastat

## REFERAAT

Baggelaar, P.K. en Van der Meulen, E.C.J., 2020. *Naar betere tijdreeksmodellering met Pastas*.
Rapport van AMO (Hengelo) en PB Icastat (Amstelveen), 70 blz., 24 fig., 16 tab., 11 ref.

Op verzoek van de Provincie Overijssel is een studie uitgevoerd naar knelpunten van de huidige uitvoering van het programma Pastas voor tijdreeksmodellering (open source, in Python). Uit het modelleren van een groot aantal Overijsselse grondwaterstandreeksen op dagbasis, met als invoer neerslag en verdamping, blijkt dat Pastas met de huidige structuur van het ruismodel nog niet in staat is de ruis- of innovatiereeks volledig terug te brengen tot witte ruis. Dit belemmert het doen van uitspraken met gekwantificeerde betrouwbaarheden. Uit simulatie-exercities blijkt dat Pastas goed in staat is een kunstmatig aangebrachte evenwichtsrelatie terug te schatten als die wordt geschat met dezelfde transferfunctie, ook als gekleurde ruis is toegevoegd. Maar als een relatievorm wordt aangebracht die enigszins afwijkt van die waar Pastas van uitgaat, kan de relatieve onzuiverheid van de geschatte evenwichtsrelatie flink oplopen. Gebaseerd op de onderzoeksbevindingen worden de volgende mogelijkheden gepresenteerd om de werking van Pastas te verbeteren: identificeren transferfuncties, flexibiliseren transferfunctie, flexibiliseren ruismodel / verlagen modelfrequentie, optimaliseren schattingsroutine en verifiëren tijdreeksmodel. Ook wordt een werkprocedure voorgesteld (neergelegd in een stroomschema), die neerkomt op een stapsgewijze, interactieve modellering, die zonodig wordt herhaald tot het model voldoet aan de daaraan gestelde randvoorwaarden.

Trefwoorden: Pastas, Python, tijdreeksmodellering, PIRFICT-methode, Box-Jenkins-methode, betrouwbaarheid, statistische significantie, zuiverheid, randvoorwaarden, witte ruis, correlatiestructuur, autocorrelatiefunctie, partiële autocorrelatiefunctie, ruismodel, transferfunctie, blokrespons, modelidentificatie, modelverificatie.

Voor correspondentie
AMO (Adviesbureau Modellering en Optimalisatie)
Ir. E.C.J. van der Meulen
Deldenerstraat 81
7551 AC Hengelo
Tel. 074 – 785 0124 / 06 248 65 321
Nederland
Email: amo@amo-nl.com
Website: https://www.amo-nl.com/amo/ of https://www.amo-nl.com/nieuws

# INHOUD


**SAMENVATTING** ..... **2**

**TERMINOLOGIE** ..... **5**

**1 INLEIDING** ..... **8**

1.1 Inleiding ..... 8
1.2 Probleemstelling en vraagstelling ..... 8
1.3 Hoofdlijnen van onze aanpak ..... 9
1.4 Opbrengsten van deze studie ..... 9
1.5 Over dit rapport ..... 10

**2 RANDVOORWAARDEN VAN EEN TIJDREEKSMODEL** ..... **11**

2.1 Randvoorwaarden die aan een tijdreeksmodel gesteld kunnen worden ..... 11
2.2 Simulaties ter illustratie belang randvoorwaarden ..... 12
*2.2.1 Opzet simulatie-exercities ..... 12*
*2.2.2 Resultaten van simulatie-exercitie A ..... 14*
*2.2.3 Resultaten van simulatie-exercitie B ..... 18*
2.3 Kernpunten van dit hoofdstuk ..... 21

**3 RANDVOORWAARDEN EN HET PASTAS-MODEL** ..... **22**

3.1 De beschouwde praktijkstudie ..... 22
3.2 Verificatie van Pastas-modellen van praktijkreeksen ..... 24
3.3 Voorbeelden van verificatie van Pastas-modellen van praktijkreeksen ..... 25
3.4 Invloed toepassen ruismodel op verklaarde variantie en evenwichtsrelatie ..... 32
3.5 Kernpunten van dit hoofdstuk ..... 34

**4 PRESTATIES PASTAS BIJ KUNSTMATIGE REEKSEN** ..... **35**

4.1 Simulatie met de geschaalde Gammaverdeling ..... 35
4.2 Simulatie met andersoortige verdeling ..... 41
4.3 Kernpunten van dit hoofdstuk ..... 41

**5 VERBETERMOGELIJKHEDEN PASTAS** ..... **43**

5.1 Identificeren transferfuncties ..... 43
5.2 Flexibiliseren transferfunctie ..... 44
5.3 Flexibiliseren ruismodel / verlagen modelfrequentie ..... 45
5.4 Optimaliseren schattingsroutine ..... 45
5.5 Verifiëren tijdreeksmodel ..... 45
5.6 Voorstel werkprocedure modelleren met Pastas ..... 46

**6 AANBEVELINGEN** ..... **47**

**AANGEHAALDE LITERATUUR** ..... **48**

**BIJLAGE 1: ALGEMENE TOELICHTING OP DE BOX-JENKINS-METHODE** ..... **49**

**BIJLAGE 2: VERKENNING FLEXIBILITEIT GAMMAVERDELING** ..... **55**

**BIJLAGE 3 – BESCHRIJVING PYTHON-MODULE VOOR MODELVERIFICATIE** ..... **63**

**BIJLAGE 4 – REACTIE VAN PASTAS-ONTWIKKELTEAM** ..... **67**

# Samenvatting

Dit rapport beschrijft aanpak en resultaten van een studie naar knelpunten van de huidige uitvoering van Pastas, een programma voor tijdreeksmodellering, dat is bedoeld voor hydrologische toepassingen. Voor die knelpunten worden oplossingen aangedragen en ook wordt er gereedschap voor beschikbaar gemaakt. Het onderzoek is uitgevoerd op verzoek van de Provincie Overijssel.

### Belang van voldoen aan randvoorwaarden bij tijdreeksmodelleren

Om aan de hand van een tijdreeksmodel uitspraken met gekwantificeerde betrouwbaarheden te mogen doen, dient aan bepaalde randvoorwaarden te worden voldaan. Zo dienen de resttermen (hier aangeduid als innovaties) te kwalificeren als witte ruis én afkomstig te zijn uit de normale kansverdeling (of een andere kansverdeling waarvan alle eigenschappen bekend zijn). Er is sprake van witte ruis als ze gemiddeld nul zijn, onafhankelijk van elkaar zijn en dezelfde variantie hebben (homoscedasticiteit). Aanvullende randvoorwaarden, die echter vaak over het hoofd worden gezien en daardoor impliciet worden aangenomen, zijn: i) dat de formulering van het deterministische deel van het tijdreeksmodel overeenstemt met de werkelijke vorm (ook wel aangeduid als het *werkelijke systeem*) en ii) dat de schatter van de evenwichtsrelatie tussen de uitvoerreeks en een invoerreeks zuiver is (dat wil zeggen zonder systematische fout).

Als bij tijdreeksmodellering de resterende innovatiereeks geen witte ruis vormt, kan dit leiden tot sterke vertekeningen van zowel het betrouwbaarheidsinterval als de statistische significantie van een geschatte relatie. Voor wat betreft het betrouwbaarheidsinterval is de empirische dekkingsgraad dan ongelijk aan de nominale dekkingsgraad. En voor wat betreft de statistische significantie is het empirische significantieniveau dan ongelijk aan het nominale significantieniveau. Het risico op statistisch significant detecteren van een niet-bestaande relatie kan dan aanzienlijk worden. Dus als de innovatiereeks geen witte ruis vormt is het doen van uitspraken met gekwantificeerde betrouwbaarheden niet meer te verantwoorden. De zuiverheid van een geschatte relatie is daarentegen onafhankelijk van het wel of niet voldoen aan witte ruis van de innovatiereeks.

### Randvoorwaarden en het Pastas-model

Uit het modelleren van 574 grondwaterstandreeksen op dagbasis, met als invoer neerslag en verdamping, blijkt dat Pastas met de huidige structuur van het ruismodel nog niet in staat is de ruis- of innovatiereeks volledig terug te brengen tot witte ruis. Er kan sprake zijn van een zeer sterke autocorrelatiestructuur van de ruis, die gepaard gaat met een geringe tot matige moving-average-structuur en soms ook nog correlatie tussen waarden die een jaar uiteen liggen. Het AR(1)-ruismodel van Pastas blijkt in staat de mate van correlatie te verminderen, maar in geen van deze gevallen worden de innovaties volledig omgezet tot witte ruis. Bovendien blijkt de ruis- of innovatiereeks zelden te voldoen aan normaliteit.
Verder blijkt uit deze exercitie met een groot aantal praktijkreeksen dat Pastas de variantie van de grondwaterstand in vrijwel alle gevallen (99,5%) beter beschrijft zónder toepassing van het ruismodel dan mét toepassing van het ruismodel, waarbij de verschillen relevant kunnen zijn. Blijkbaar kan toepassing van het huidige ruismodel schade aan de pasvorm opleveren.
Aangezien het een omvangrijke en ruimtelijk goed gespreide verzameling van grondwaterstandreeksen betreft uit de Provincie Overijssel, mogen we aannemen dat deze problemen zich vrijwel overal kunnen voordoen. De conclusie is daarom gerechtvaardigd dat Pastas in zijn huidige vorm



© AMO en PB Icastat 

vaak nog niet in staat zal stellen om bij modelleren van grondwaterstanden op dagbasis uitspraken met gekwantificeerde betrouwbaarheden te doen.

### Prestaties Pastas bij gesimuleerde reeksen

Uit simulatie-exercities blijkt dat Pastas met de Gamma-transferfunctie goed in staat is een op basis van diezelfde functie kunstmatig aangebrachte evenwichtsrelatie terug te schatten, ook als gekleurde ruis is toegevoegd. Verder blijkt bij modelleren mét ruismodel de empirische dekkingsgraad van het 95%-betrouwbaarheidsinterval van de geschatte evenwichtsrelatie min of meer gelijk aan de nominale dekkingsgraad. Dit laatste zal komen doordat de correlatiestructuur van de bij onze simulaties aangebrachte ruis – AR(1) – aansluit op wat het huidige ruismodel van Pastas aankan. Maar in de praktijk zal er bij modelleren van de grondwaterstand op dagbasis veelal sprake zijn van een uitgebreidere correlatiestructuur. Het ligt voor de hand dat Pastas dan niet in staat zal zijn de innovaties te laten voldoen aan de randvoorwaarde van witte ruis, met als gevolg dat de risico's op verkeerde conclusies groter zullen zijn dan geaccepteerd. Uit de simulatie-exercities blijkt dat bij modelleren zónder ruismodel de empirische dekkingsgraad van het 95%-betrouwbaarheidsinterval van de geschatte evenwichtsrelatie laag is (35 à 37%), wat een groot risico geeft op het onterecht detecteren van een statistisch significante relatie.
Als we relatievormen aanbrengen die enigszins afwijken van die van de Gamma-transferfunctie, terwijl Pastas bij het modelschatten wel van die functie uitgaat, kan de relatieve onzuiverheid van de geschatte evenwichtsrelatie flink oplopen. Dit resultaat is uiteraard niet onverwacht, omdat voor het simuleren een andere functie is gebruikt dan waar bij het schatten van het model rekening mee wordt gehouden. Het moet dan ook niet worden gezien als een onvolkomenheid van Pastas, maar van het modelleerproces. Het wijst er op hoe belangrijk het is om bij het schatten van het model zo goed mogelijk rekening te kunnen houden met de werkelijke vorm van de relatie. De volgende mogelijkheden verdienen in dat kader nadere uitwerking: i) uitbreiden van de verzameling voorgedefinieerde transferfuncties, in combinatie met het toevoegen van een identificatiestap aan het modelleerproces en ii) ontwikkelen van flexibelere transferfuncties.

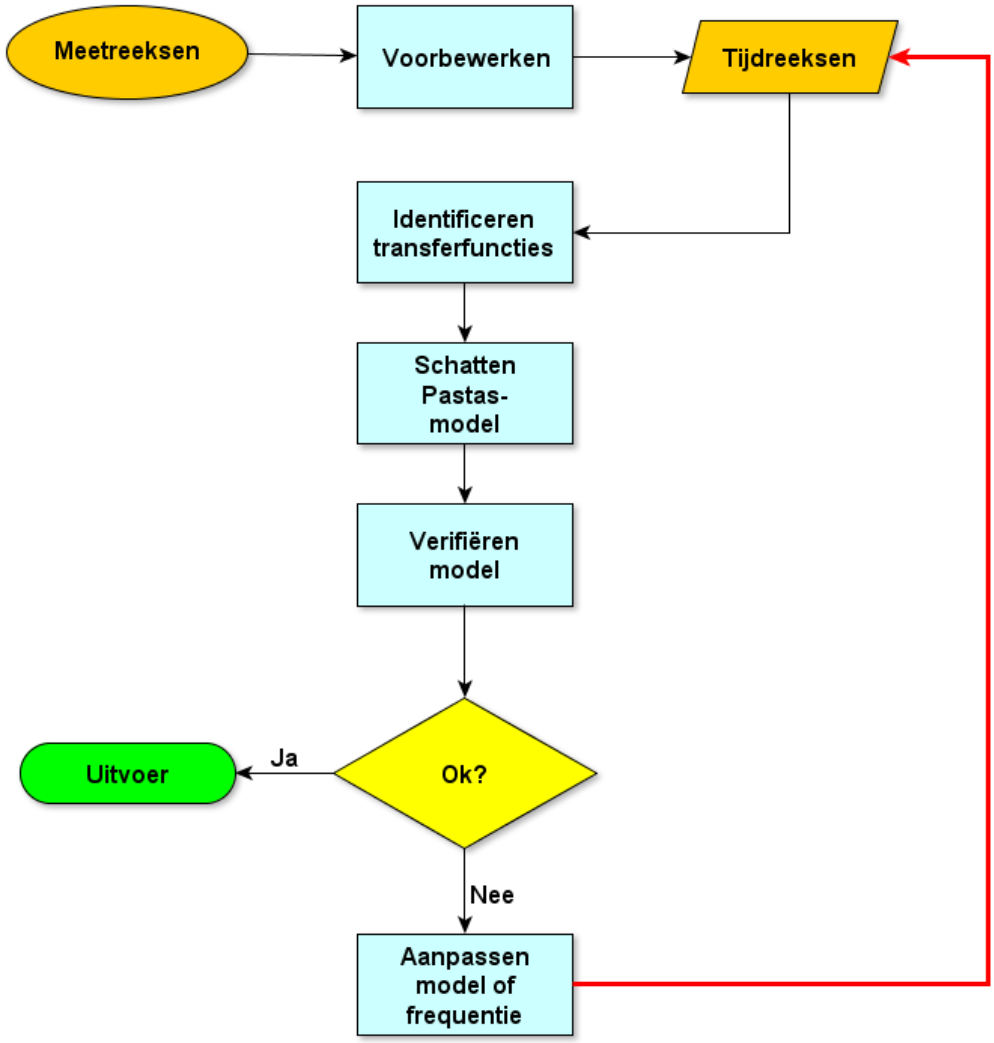
### Verbetermogelijkheden Pastas

Gebaseerd op de onderzoeksbevindingen, zien we verschillende mogelijkheden om de werking van Pastas te verbeteren. Deze betreffen de volgende onderdelen van het modelleerproces: identificeren transferfuncties, flexibiliseren transferfunctie, flexibiliseren ruismodel / verlagen modelfrequentie, optimaliseren schattingsroutine en verifiëren tijdreeksmodel. De details vindt u in hoofdstuk 5.

Het stroomschema op de volgende pagina toont de door ons voorgestelde werkprocedure om met Pastas te modelleren. De procedure komt neer op een stapsgewijze, interactieve modellering, die zonodig wordt herhaald tot het model voldoet aan de daaraan gestelde randvoorwaarden. Deze kan vermoedelijk niet geheel automatisch worden uitgevoerd, aangezien er bij diverse onderdelen interactie nodig is van de modelleur. Het kan dus meer inzet vergen dan tot dusverre gebruikelijk, maar het zal de kwaliteit en zeggingskracht van de modellen verbeteren.


© AMO en PB Icastat

*Stroomschema van de voorgestelde werkprocedure om met Pastas te modelleren.*




De verschillende elementen van de werkprocedure zijn aangegeven met de volgende vormen en kleuren: oranje ellips: invoer | lichtblauwe rechthoek: werkzaamheid | oranje parallellogram: bewerkte invoer | gele ruit: beslismoment | groene ellips: uitvoer


© AMO en PB Icastat

# Terminologie

Autocorrelatie - Het verschijnsel dat opeenvolgende reekswaarden niet onafhankelijk van elkaar zijn. Als een natuurlijk proces met een hoge frequentie wordt waargenomen treedt doorgaans positieve autocorrelatie op, wat inhoudt dat opeenvolgende reekswaarden meer op elkaar lijken dan op verder in de tijd gelegen reekswaarden. Een aantal opeenvolgende reekswaarden zal daardoor meer de kenmerken weerspiegelen van een segment van de kansverdeling waar ze uit afkomstig zijn, dan van de gehele kansverdeling. Als hier geen rekening mee wordt gehouden, zal bijvoorbeeld de standaardafwijking van die kansverdeling worden onderschat.

Autocorrelatiecoëfficiënt - Een maat voor de lineaire samenhang tussen reekswaarden die zijn gescheiden door een bepaald aantal tijdseenheden.

Autocorrelatiefunctie (ACF) - Een grafiek die de autocorrelatiecoëfficiënt weergeeft als functie van het aantal tijdseenheden tussen de waarden.

Betrouwbaarheid bij statistisch toetsen – De kans op het terecht niet verwerpen van de nulhypothese ($H_0$). Deze kans wordt ook wel aangeduid als $1 - \alpha$, waarbij $\alpha$ het vóóraf ingestelde significantieniveau is, zijnde het geaccepteerde risico op het onterecht verwerpen van de nulhypothese.

| **Werkelijkheid** | **Vóóraf geaccepteerde kansen op resultaat bij statistische toets** | |
|---|---|---|
| | **$H_0$ wordt niet verworpen** | **$H_0$ wordt verworpen** |
| **$H_0$ is waar** | $1-\alpha$ (betrouwbaarheid) | $\alpha$ (significantieniveau) |

We onderscheiden:

- Nominale betrouwbaarheid – deze is vóóraf ingesteld (meestal op 95%).
- Empirische betrouwbaarheid – deze wordt in de praktijk gerealiseerd en kan afwijken van de nominale betrouwbaarheid als niet wordt voldaan aan bepaalde randvoorwaarden.

Betrouwbaarheidsinterval van een schatting – Een maat voor de precisie van een schatting. Als wordt voldaan aan de daarvoor geldende randvoorwaarden, zal een 95%-betrouwbaarheidsinterval van een schatting met 95% betrouwbaarheid de werkelijke waarde bevatten. Engels: *confidence interval of an estimate*.

Deterministische deel van een tijdreeksmodel: Het deel dat de relaties met gemeten invoerreeksen (zoals neerslag, verdamping en grondwaterwinning) beschrijft.

Evenwichtsrelatie - De uiteindelijke stationaire verandering van de uitvoerreeks van een systeem als de invoerreeks een constante niveauverandering van één eenheid heeft. Engels: *steady-state gain*.

Homoscedastisciteit: Een reeks vertoont homoscedasticiteit als de variantie van de reekswaarden niet verandert met het niveau. Engels: *homoscedasticity*.

*p*-waarde – Bij statistisch toetsen is dit de kans op een extremere toetsingsgrootheid dan degene die is gerealiseerd, bij geldigheid van de nulhypothese. Als de *p*-waarde kleiner is dan het vooraf


© AMO en PB Icastat

ingestelde significantieniveau ($\alpha$, zoals 5% of 1%), kan de nulhypothese worden verworpen. De betrouwbaarheid van die toetsuitspraak is dan 1-$\alpha$. Als het significantieniveau bijvoorbeeld 5% bedraagt, is de betrouwbaarheid 95%.

Partiële autocorrelatiecoëfficiënt - Een maat voor de lineaire samenhang tussen reekswaarden die zijn gescheiden door een bepaald aantal tijdseenheden, ná correctie voor de lineaire samenhang tussen reekswaarden die zijn gescheiden door minder tijdseenheden.

Partiële autocorrelatiefunctie (PACF) - Een grafiek die de partiële autocorrelatiecoëfficiënt weergeeft als functie van het aantal tijdseenheden tussen de waarden. Een ACF en PACF samen geven een completer beeld van de achterliggende correlatiestructuur dan alleen een ACF en stellen zo in staat het ruismodel beter te identificeren.

Schatter - Een methode om de waarde van een bepaalde parameter te schatten. Engels: *estimator*.

Significantieniveau – Het vóóraf ingestelde – en daarmee nog geaccepteerde - risico op het onterecht verwerpen van de nulhypothese. Deze wordt aangeduid als $\alpha$, of ook wel als het risico van de 1$^e$ soort. We onderscheiden:

- Nominaal significantieniveau - deze is vóóraf ingesteld (meestal op 5%) en geeft aan welk risico wordt geaccepteerd.
- Empirisch significantieniveau - deze wordt in de praktijk gerealiseerd en kan afwijken van het nominale significantieniveau als niet wordt voldaan aan bepaalde randvoorwaarden.

Standaardfout van een schatting – Een maat voor de onzekerheid van een schatting, op te vatten als de standaardafwijking van de schatter. Engels: *standard error of an estimation.*

Tijdreeks - Een chronologisch gerangschikte reeks waarden van een bepaalde variabele, zoals de grondwaterstand, of de neerslag op een bepaalde locatie. Engels: *time series*.

Zuivere schatter - Een schatter zonder systematische fout. De verwachtingswaarde van de schatter is dan gelijk aan de werkelijke waarde van het schattingsdoel, oftewel $\mathrm{E}\left[\hat{\theta}\right] = \theta$, waarin E[.] de verwachtingswaarde, $\hat{\theta}$ de schatting en $\theta$ de werkelijke waarde. Engels: *unbiased estimator*.


© AMO en PB Icastat

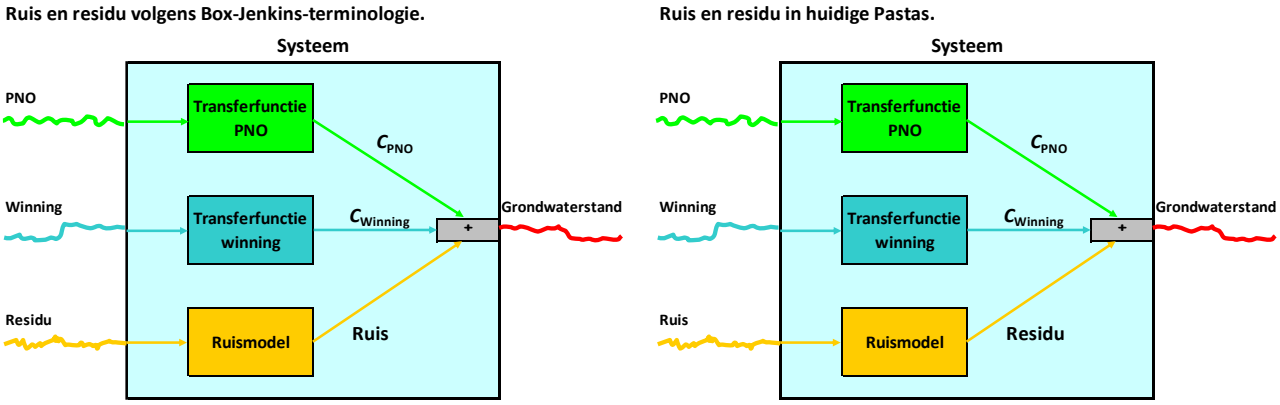
### Eem speciaal woord over de resttermen

Ten slotte nog een speciaal woord over de resttermen, want het is bijna onvermijdelijk dat die verwarring gaan geven, omdat er nu eenmaal verschillende naamgevingen circuleren.

Bij tijdreeksanalyse is er sprake van twee soorten restterm, namelijk:

1. het verschil tussen de meetreeks en het deterministische deel van het tijdreeksmodel. Dit kunnen we ook opvatten als de component van de grondwaterstand die samenhangt met alle invloedsfactoren die niet bij de modellering zijn betrokken;
2. de reeks die resteert als bovenstaande restterm 1 is gefilterd (met de inverse van het ruismodel) met de bedoeling zijn correlatiestructuur te verwijderen. Deze reeks is op te vatten als de invoer van het ruismodel.

Door de grondleggers van het discrete transfer-ruismodel – Box en Jenkins – zijn deze twee resttermen aangeduid als respectievelijk *ruis* en *residu* [Box and Jenkins, 1976]. In het linkerdeel van onderstaande figuur is weergegeven wat hun rol is in het transfer-ruismodel. Het voorbeeld betreft een systeem met als uitvoer de grondwaterstand en als invoer reeksen van het potentieel neerslagoverschot, een grondwaterwinning en het residu. De ruis is hier de uitvoer van het ruismodel. In Pastas zijn deze termen echter omgewisseld en worden ze aangeduid als respectievelijk *residu* en *ruis* (zie het rechterdeel van de figuur). Dit laatste is verwarrend, omdat het taalkundig gezien meer voor de hand ligt dat de ruis uitvoer is van het ruismodel, dan invoer. Mogelijk is deze verwarring ontstaan doordat een belangrijke randvoorwaarde van een tijdreeksmodel is dat de invoer van het ruismodel *witte ruis* vormt. Die term is echter bedoeld als eigenschap van een reeks en niet als een naam.




We hebben ervoor gekozen om in dit rapport het verschil tussen de meetreeks en het deterministische deel van het tijdreeksmodel aan te duiden als *ruis* en de invoer van het ruismodel als *innovatie* (zie onderstaande tabel). Door hier specifiek niet de combinatie *ruis* en *residu* te gebruiken, wordt hopelijk vermeden dat Pastas-gebruikers ervan uitgaan dat beide de betekenis hebben zoals zij die gewend zijn.

| **Dit rapport** | **Box-Jenkins** | **Pastas** |
|---|---|---|
| **Ruis** | Ruis | Residu |
| **Innovatie** | Residu | Ruis |

In de door ons voorgestelde Python-verificatiemodule (zie bijlage 3) kan de gebruiker eenvoudig de terminologie naar wens aanpassen.

In hoofdstuk 2 van dit rapport wordt soms ook de term *modelfout* gebruikt, die dan is bedoeld als een algemene restterm.


© AMO en PB Icastat

# 1 Inleiding

Dit rapport beschrijft aanpak en resultaten van een studie naar knelpunten van de huidige uitvoering van Pastas, een programma voor tijdreeksmodellering, dat is bedoeld voor hydrologische toepassingen. Voor die knelpunten worden oplossingen aangedragen en ook wordt er gereedschap voor beschikbaar gemaakt.

## 1.1 Inleiding

Onder een tijdreeks verstaan we een verzameling chronologisch gerangschikte waarden van een bepaalde variabele, zoals de grondwaterstand, of de neerslag op een bepaalde locatie. Hydrologen krijgen steeds meer de beschikking over hoogfrequente tijdreeksen (op dag-, uur-, of soms zelfs minutenbasis), doordat het meten van de grondwaterstand in Nederland vanaf circa 2004 in toenemende mate geschiedt met loggers. Er is inmiddels gangbare software beschikbaar gekomen die in staat stelt om hydrologische verbanden tussen hoogfrequente tijdreeksen van grondwaterstand, neerslag, verdamping en soms ook grondwaterwinningen te beschrijven en te kwantificeren. Deze software is gebaseerd op de PIRFICT[1]-methode [Von Asmuth et al., 2002 en Von Asmuth, 2012] en wordt bijvoorbeeld toegepast om toekomstige grondwaterstanden te voorspellen, of om de invloed van een winning of een ingreep op de grondwaterstand te kwantificeren. Eerst kwam daarvoor Menyanthes beschikbaar en later ook programma's als HydroSight, Metran en Pastas.

## 1.2 Probleemstelling en vraagstelling

Het is nog onvoldoende duidelijk of deze continue tijdreeksmodellen wel altijd geschikt zijn voor de beoogde toepassingen. Zo is het nog de vraag in hoeverre een model op hoogfrequente basis altijd geschikt is om een traag proces te beschrijven. Verder is het onzeker of het wel altijd geoorloofd is om op basis van deze modellen uitspraken met gekwantificeerde betrouwbaarheid te doen. Een belangrijke randvoorwaarde daarvoor is bijvoorbeeld dat de restwaarden (innovaties) afkomstig zijn uit een normale kansverdeling (of een andere kansverdeling waarvan alle eigenschappen bekend zijn) en onafhankelijk van elkaar zijn. Maar om daaraan te kunnen voldoen zal soms een uitgebreider ruismodel nodig zijn dan nu beschikbaar is, of zal op een laagfrequentere tijdsbasis moeten worden gemodelleerd.

De Provincie Overijssel ziet veel toepassingsmogelijkheden voor tijdreeksanalyses van grondwaterstanden - mede voor zijn eigen beheertaken - en wil daarom dat er gebruiksvriendelijke software beschikbaar is die niet alleen leidt tot verantwoorde modellen, maar ook in staat stelt om op basis van het afgeleide model uitspraken met gekwantificeerde betrouwbaarheden te kunnen doen. De Provincie wil dat de onderzoeksinspanning zich hierbij richt op Pastas, omdat zijn toolbox-structuur met Python-scripts in een open-source omgeving veel flexibiliteit en ontwikkelcapaciteit biedt. Verder is het voor iedereen gratis beschikbaar. Oorspronkelijk werd het programma aangeduid met het acroniem PASTAS, dat staat voor *Python Applied Statistical Timeseries Analysis Software*. Maar

[1] Het acroniem PIRFICT staat voor *Predefined Impulse Response Functions In Continuous Time*.

inmiddels wordt het vaak uitgeschreven als Pastas, dus als een naam.[2] Pastas wordt actief ontwikkeld door de TU Delft, Artesia en inmiddels ook de Universiteit van Graz (Oostenrijk) [Collenteur et al., 2019]. Het ontwikkelteam heeft meegelezen met een conceptversie van ons rapport en op basis daarvan zijn onderdelen aangepast of nader uitgewerkt. In bijlage 4 vindt u de reactie van het Pastas-ontwikkelteam op de eindversie van ons rapport.

### Vraagstelling

In het licht van bovenstaande heeft de Provincie[3] ons verzocht in beeld te brengen welke knelpunten de huidige uitvoering van Pastas heeft voor de beoogde toepassingen, vervolgens aan te geven hoe die kunnen worden aangepakt en daar ook gereedschap voor beschikbaar te maken.

We wijzen er op dat het onderzoek zich vooral richt op knelpunten van de PIRFICT-methode en minder op specifieke eigenschappen van Pastas. De bevindingen zullen daarom vermoedelijk ook van toepassing zijn op andere programma's die de PIRFICT-methode hanteren, zoals Menyanthes, HydroSight en Metran.

## 1.3 Hoofdlijnen van onze aanpak

Voor deze studie hebben we de volgende stappen uitgevoerd.

1. Beschrijven aan welke randvoorwaarden een tijdreeksmodel moet voldoen om uitspraken met gekwantificeerde betrouwbaarheden te mogen doen.
2. Aan de hand van simulaties beschrijven welke vertekeningen er kunnen optreden als een tijdreeksmodel niet voldoet aan die randvoorwaarden, voor achtereenvolgens betrouwbaarheidsinterval, statistische significantie en zuiverheid van de geschatte evenwichtsrelatie tussen een uitvoerreeks en een invoerreeks.
3. Ontwikkelen van een Python-script om te kunnen vaststellen of een Pastas-model voldoet aan de randvoorwaarden om uitspraken met gekwantificeerde betrouwbaarheden te mogen doen.
4. Voor een groot aantal in de praktijk ontwikkelde Pastas-modellen nagaan of wordt voldaan aan de randvoorwaarden om uitspraken met gekwantificeerde betrouwbaarheden te mogen doen.
5. Beoordelen van de mogelijkheden om Pastas-modellen beter te laten voldoen aan die randvoorwaarden en op basis daarvan een werkprocedure ontwikkelen voor het uitvoeren van tijdreeksanalyse met Pastas.

## 1.4 Opbrengsten van deze studie

1. Een digitaal eindrapport (zowel in pdf-formaat als in docx-formaat), met een beschrijving van aanpak, bevindingen en resultaten van deze studie.
2. Een Python-script (ontwikkeld volgens PEP8-richtlijnen voor Python-code) waarmee kan worden geverifieerd of een met Pastas afgeleid tijdreeksmodel voldoet aan bepaalde randvoorwaarden.
3. Aanbevelingen voor het uitvoeren van tijdreeksanalyse met Pastas, vastgelegd in een beslisboom, met toelichting van de stappen.

[2] In dit rapport gebruiken we de laatste schrijfwijze, maar enkele van onze oudere figuren hanteren nog de oude schrijfwijze.

[3] Hier vertegenwoordigd door ir. Thomas de Meij, beleidsmedewerker geohydrologie, Provincie Overijssel.

## 1.5 Over dit rapport

Na deze inleiding gaat hoofdstuk 2 in op de randvoorwaarden waaraan een tijdreeksmodel moet voldoen om verantwoorde uitspraken met gekwantificeerde betrouwbaarheden te kunnen doen. Tevens wordt met simulaties in beeld gebracht welke vertekeningen er kunnen optreden als dat niet het geval is. Hoofdstuk 3 gaat na hóe en in welke mate een Pastas-model in de praktijk kan afwijken van die randvoorwaarden. Daarna beschrijft hoofdstuk 4 aanpak en resultaten van enkele simulatie-exercities met kunstmatige tijdreeksen, die zijn uitgevoerd om de prestaties van Pastas objectief te kunnen beoordelen. Gebaseerd op de bevindingen van de voorgaande hoofdstukken beschouwt hoofdstuk 5 mogelijkheden om de werking van Pastas te verbeteren en presenteert ook een stroomschema voor het gebruik van Pastas. Ten slotte geeft hoofdstuk 6 onze aanbevelingen naar aanleiding van deze studie. Het hoofddeel van dit rapport sluit af met de alfabetisch gerangschikte lijst van de literatuurverwijzingen.

Verder bevat het rapport vier bijlagen.

- Bijlage 1 bevat een algemene toelichting op tijdreeksanalyse in het discrete domein volgens de Box-Jenkins-methode.
- Bijlage 2 bevat een korte verkenning van de flexibiliteit en daarmee de breedte van het toepassingsgebied van de Gammaverdeling.
- Bijlage 3 beschrijft de in deze studie ontwikkelde Python-module waarmee kan worden geverifieerd of een met Pastas afgeleid tijdreeksmodel voldoet aan de randvoorwaarden om op basis van het afgeleide model uitspraken met gekwantificeerde betrouwbaarheden te kunnen doen.
- Bijlage 4 bevat de reactie van het Pastas-ontwikkelteam op ons rapport. Dit team is samengesteld uit Raoul Collenteur (Universiteit van Graz), Ruben Caljé, Davíd Brakenhoff, Onno Ebbens en Frans Schaars (allen Artesia) en Mark Bakker (TU Delft).

# 2 Randvoorwaarden van een tijdreeksmodel

Dit hoofdstuk gaat in op de randvoorwaarden waaraan een tijdreeksmodel moet voldoen om verantwoorde uitspraken met gekwantificeerde betrouwbaarheden te kunnen doen (§ 2.1). In § 2.2 wordt met simulaties in beeld gebracht welke vertekeningen er kunnen optreden als een tijdreeksmodel niet voldoet aan die randvoorwaarden, voor achtereenvolgens betrouwbaarheidsinterval, statistische significantie en zuiverheid van de geschatte evenwichtsrelatie tussen een uitvoerreeks en een invoerreeks. Het hoofdstuk sluit af met een opsomming van de kernpunten (§ 2.3).

## 2.1 Randvoorwaarden die aan een tijdreeksmodel gesteld kunnen worden

De belangrijkste toepassingen van tijdreeksmodellering zijn: i) het verkrijgen van *inzicht* in een systeem (statistische relaties tussen in- en uitvoer), ii) het *voorspellen* van de uitvoer van een systeem, of iii) het kunnen *beheren* van een systeem, zodanig dat zijn uitvoer aan bepaalde wensen voldoet.

Afhankelijk van de beoogde toepassing kunnen er diverse randvoorwaarden aan een tijdreeksmodel worden gesteld. Als het model bijvoorbeeld dient om resultaten met gekwantificeerde betrouwbaarheid te doen, zoals:

- het voorspellen van toekomstige grondwaterstanden, inclusief 95%-voorspelinterval, of
- het kwantificeren van de invloed van ingrepen in het watersysteem op de grondwaterstand, inclusief 95%-betrouwbaarheidsinterval en eventueel statistische significantie, of
- het kwantificeren van een langjarige trend van de grondwaterstand (eventueel na correctie voor de meteorologische invloed), inclusief 95%-betrouwbaarheidsinterval,

dan dienen de modelfouten[4] te voldoen aan de volgende randvoorwaarden:

1. ze zijn gemiddeld nul;
2. ze zijn onafhankelijk van elkaar;
3. hun kansverdelingen hebben dezelfde variantie (homoscedasticiteit);
4. ze zijn afkomstig uit de normale kansverdeling (of een andere kansverdeling waarvan alle eigenschappen bekend zijn).

Als een reeks voldoet aan de eerste drie voorwaarden spreken we van *witte ruis*. Als deze bijvoorbeeld ook voldoet aan normaliteit spreken we van *normaal verdeelde witte ruis*.
Aanvullende randvoorwaarden, die echter vaak over het hoofd worden gezien en daardoor impliciet worden aangenomen, zijn:

5. de formulering van het deterministische deel van het tijdreeksmodel stemt overeen met de werkelijke vorm (ook wel aangeduid als het *werkelijke systeem*);
6. de schatter van de evenwichtsrelatie tussen de uitvoerreeks en een invoerreeks is zuiver.

Bij modelleren van tijdreeksen met Pastas wordt doorgaans alleen als randvoorwaarde gesteld dat het deterministische deel van het model minstens 70% van de variantie van de tijdreeks beschrijft. Desondanks worden er op basis van dergelijke modellen soms ook uitspraken met gekwantificeerde betrouwbaarheden gedaan, zonder dat er geverifieerd is of aan de daarvoor geldende randvoorwaarden wordt voldaan. In dit hoofdstuk brengen we in beeld wat daarbij de risico's zijn.

[4] De term *modelfout* wordt in dit hoofdstuk gebruikt als een algemene restterm.


© AMO en PB Icastat

## 2.2 Simulaties ter illustratie belang randvoorwaarden

We hebben een aantal simulaties uitgevoerd, om in beeld te brengen welke vertekeningen er kunnen optreden als we uitspraken met gekwantificeerde betrouwbaarheden doen op basis van een tijdreeksmodel dat niet voldoet aan de randvoorwaarden. Dit wordt in beeld gebracht voor uitspraken over betrouwbaarheidsinterval, statistische significantie en zuiverheid van de geschatte evenwichtsrelatie tussen een uitvoerreeks en een invoerreeks. Daarbij beperken we ons tot de randvoorwaarde dat de modelfouten normaal verdeelde witte ruis vormen.

De volgende paragraaf (§ 2.2.1) bevat de technische details van de simulatie-exercities. U kunt er uiteraard ook voor kiezen direct verder te gaan naar de resultaten (§ 2.2.2).

### *2.2.1 Opzet simulatie-exercities*

De simulaties gaan uit van een tijdreeks van de variabele *Z*, met *k* equidistante dagwaarden ($z_1$, $z_2$, …, $z_k$) en van een tijdreeks van de variabele *X*, eveneens met *k* equidistante dagwaarden ($x_1$, $x_2$, …, $z_k$). Verder zijn *Z* en *X* steeds gelijktijdig gemeten. Stel nu dat we objectief willen vaststellen of *Z* en *X* gerelateerd zijn en dat het antwoord belangrijke implicaties kan hebben, bijvoorbeeld op maatschappelijk, beleidsmatig, juridisch of technisch vlak. We zullen dan niet alleen de grootte van de relatie willen schatten, maar tevens zijn 95%-betrouwbaarheidsinterval en eventueel statistische significantie willen vaststellen.

Voor de eenvoud gaan we er voor onze voorbeelden van uit, dat uit theoretische kennis van het betreffende soort systeem volgt dat de vorm van een eventuele relatie tussen *Z* en *X* lineair en instantaan zal zijn, maar zonder na-ijling. Een dergelijke relatie kan worden beschreven met het enkelvoudig lineair regressiemodel, volgens:

$$z_t = \beta_0 + \beta_1 \cdot x_t + e_t \qquad [2.1]$$

waarin $\beta_0$ het intercept, $\beta_1$ de regressieparameter, $e_t$ de realisatie van de modelfout en *t* de (discrete) tijdsindex. De eerste twee termen van het rechterlid van [2.1] vormen het deterministische deel van het regressiemodel en de modelfout vormt het stochastische deel. Aangezien het een relatie tussen tijdreeksen betreft, kunnen we dit regressiemodel opvatten als een eenvoudig tijdreeksmodel. De parameter $\beta_1$ kunnen we dan aanduiden als de evenwichtsrelatie van *Z* en *X*.

Als we niet alleen de grootte van $\beta_1$ willen schatten, maar ook zijn 95%-betrouwbaarheidsinterval en eventueel statistische significantie, dient volgens de statistische theorie de modelfout witte ruis te vormen en afkomstig te zijn uit een kansverdeling waarvan alle eigenschappen bekend zijn (zoals de normale kansverdeling). Om te kunnen vaststellen in welke mate die statistische kenmerken vertekend kunnen zijn als de modelfout correlatie vertoont, voegen we bij onze simulaties geen witte ruis, maar 'gekleurde' ruis toe aan de lineaire relatie tussen deze twee variabelen, volgens:

$$z_t = \beta_0 + \beta_1 \cdot x_t + n_t \qquad [2.2]$$

waarin $n_t$ de realisatie van de ruis. Zie daarvoor onderstaande toelichting.

Modelmatig verdisconteren van gekleurde ruis bij tijdreeksmodellering

Er kunnen twee hoofdvormen van correlatiestructuur van ruis worden onderscheiden, namelijk:

1. autoregressieve structuur – De eenvoudigste daarvan is de 1e-orde-autoregressieve structuur. Deze wordt in een discreet ruismodel verdisconteerd als:

$$n_t = \phi_1 n_{t-1} + a_t \qquad [2.3]$$

waarin $\phi_1$ (*phi*$_1$) de 1e-orde-autoregressieve parameter en $a_t$ de realisatie van de innovatie. Dit wordt aangeduid als AR(1)-model;

2. moving-average-structuur - De eenvoudigste daarvan is de 1e-orde-moving-average-structuur. Deze wordt in een discreet ruismodel verdisconteerd als:

$$n_t = a_t - \theta_1 a_{t-1} \qquad [2.4]$$

waarin $\theta_1$ (*theta*$_1$) de 1e-orde-moving-average-parameter. Dit wordt aangeduid als MA(1)-model.

Om voldoende flexibiliteit te kunnen bieden voor allerlei praktijksituaties moet een ruismodel combinaties van deze correlatiestructuren kunnen beschrijven, niet alleen van de 1e-orde, maar ook van hogere ordes. Dit duidt men wel aan als een ARMA-model (of SARMA-model, als het tevens een seizoensmatige correlatiestructuur kan beschrijven). Er is nog een verdere uitbreiding daarvan, die tevens verschillende soorten differenties kan verdisconteren: het ARIMA-model [Box and Jenkins, 1976]. Zie bijlage 1 voor een verdere toelichting.

Voor onze simulatie is gekleurde ruis gegenereerd volgens een ruismodel dat bestaat uit een combinatie van een AR(1)-model en een MA(1)-model, hier aangeduid als een ARMA(1,1)-ruismodel:

$$n_t = \phi_1 n_{t-1} + a_t - \theta_1 a_{t-1} \qquad [2.5]$$

Dit ruismodel is uitgebreider dan het AR(1)-ruismodel dat is geïmplementeerd in Pastas en Menyanthes. Bij hydrologische tijdreeksmodellering blijkt het AR(1)-ruismodel namelijk zelden afdoende om de innovaties te laten voldoen aan de randvoorwaarde van witte ruis, vooral als het een hoogfrequente tijdreeks betreft (zie ook § 3.2).

We hanteren bij de simulatie 190 uitvoeringen van dit ruismodel, namelijk voor alle combinaties van $\phi_1$ en $\theta_1$, waarbij $\phi_1$ waarden aanneemt tussen 0 en 0,9, met stapgrootte 0,1 en $\theta_1$ waarden aanneemt tussen -0,9 en 0,9, eveneens met stapgrootte 0,1. Voor elk van deze 190 combinaties van $\phi_1$ en $\theta_1$ wordt 5.000 maal een ruisreeks toegevoegd aan het deterministische deel van het tijdreeksmodel, waarbij elke ruisreeks volgens formule [2.5] is gegenereerd met een nieuwe reeks aselecte trekkingen van innovaties ($a$) uit een normale kansverdeling. De standaardafwijking van de innovaties ($\sigma_A$) stellen we daarbij zo in dat bij elk van de 190 ruismodellen de theoretische variantie van de ruis ($\sigma_N$) dezelfde, vooraf beoogde grootte heeft, met behulp van de volgende formule voor hun relatie [Box and Jenkins, 1976]:

$$\sigma_N^2 = \frac{(1 - 2\phi_1\theta_1 + \theta_1^2) \cdot \sigma_A^2}{1 - \phi_1^2} \qquad [2.6]$$

### Verdere specificaties van de opzet

- Er zijn twee simulatie-exercities uitgevoerd (A en B), elk met vaste instellingen voor $\beta_0$ (intercept), $\beta_1$ (regressieparameter) en $\sigma_N$ (standaardafwijking van de ruis). De exercities verschillen alleen in de instelling van $\beta_1$. Dit betreft:
  A. $\beta_0$ = 10, $\beta_1$ = **1** en $\sigma_N$ = 2 (*Z* en *X* zijn hierbij dus gerelateerd via $\beta_1$);
  B. $\beta_0$ = 10, $\beta_1$ = **0** en $\sigma_N$ = 2 (*Z* en *X* zijn hierbij dus níet gerelateerd).
- Elk van de tijdreeksen *Z*, *X* en *N* bevat 3.652 dagwaarden (*k* = 3.652) en beslaat de tienjarige periode 2009 t/m 2018.
- De tijdreeks van de variabele *X* is eenmalig samengesteld (handmatig). Deze omvat elf niveauveranderingen, zodat een relatie tussen *Z* en *X* makkelijker kan worden gedetecteerd (zie figuur 2.1).

- Op elk van de 950.000 (= 190 x 5.000) gesimuleerde tijdreeksen $z_t = \beta_0 + \beta_1 \cdot x_t + n_t$ zijn de volgende statistische bewerkingen uitgevoerd:
  - schatten van de modelparameters $\beta_0$ en $\beta_1$ als $b_0$ en $b_1$, leidend tot het volgende uit de meetwaarden geschatte tijdreeksmodel:

    $$z_t = b_0 + b_1 \cdot x_t + e_t \qquad [2.7]$$

    waarin $e_t$ de modelfout;
  - bepalen of de ingestelde waarde van $\beta_1$ wel of niet binnen het 95%-betrouwbaarheidsinterval van zijn schatting $b_1$ ligt, dat volgt uit:

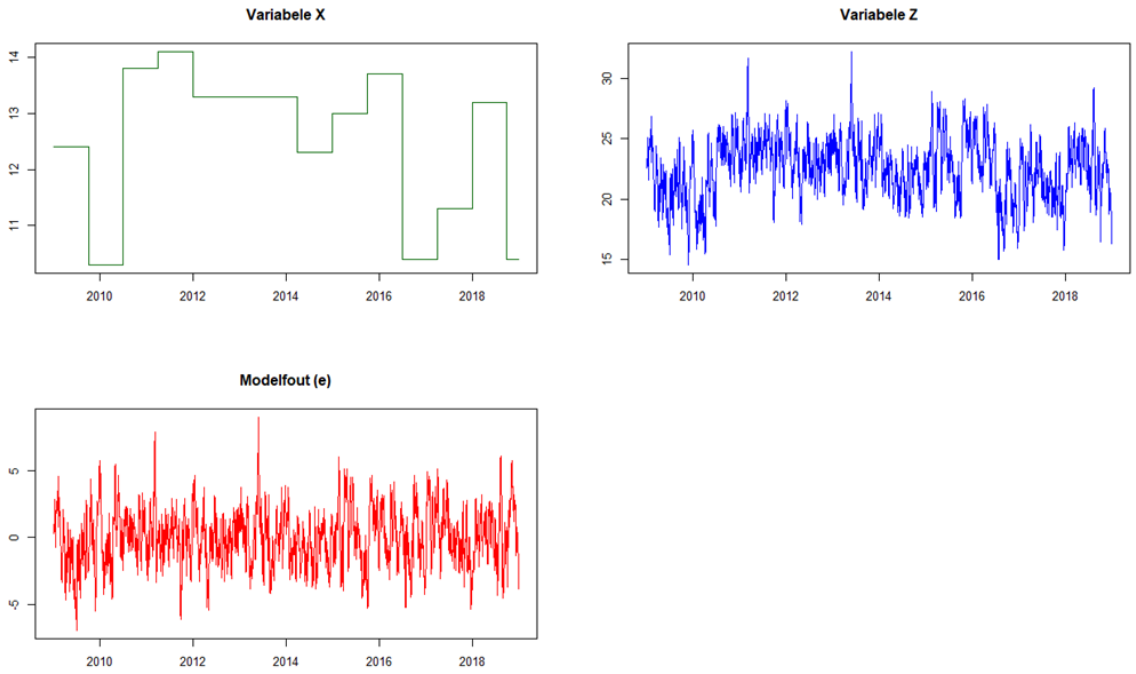
    $$95\%\ \text{b. i.}\,[b_1] \;=\; b_1 \pm t_{97{,}5\%;k-2} \cdot s_{b_1} \qquad [2.8]$$

    waarin $t_{97,5\%;k-2}$ het 97,5-percentiel van de Student-t-verdeling met $k$-2 vrijheidsgraden, $k$ het aantal waarden, $s_{b_1}$ de standaardfout van $b_1$ en de andere symbolen als boven toegelicht.

### 2.2.2 *Resultaten van simulatie-exercitie A*

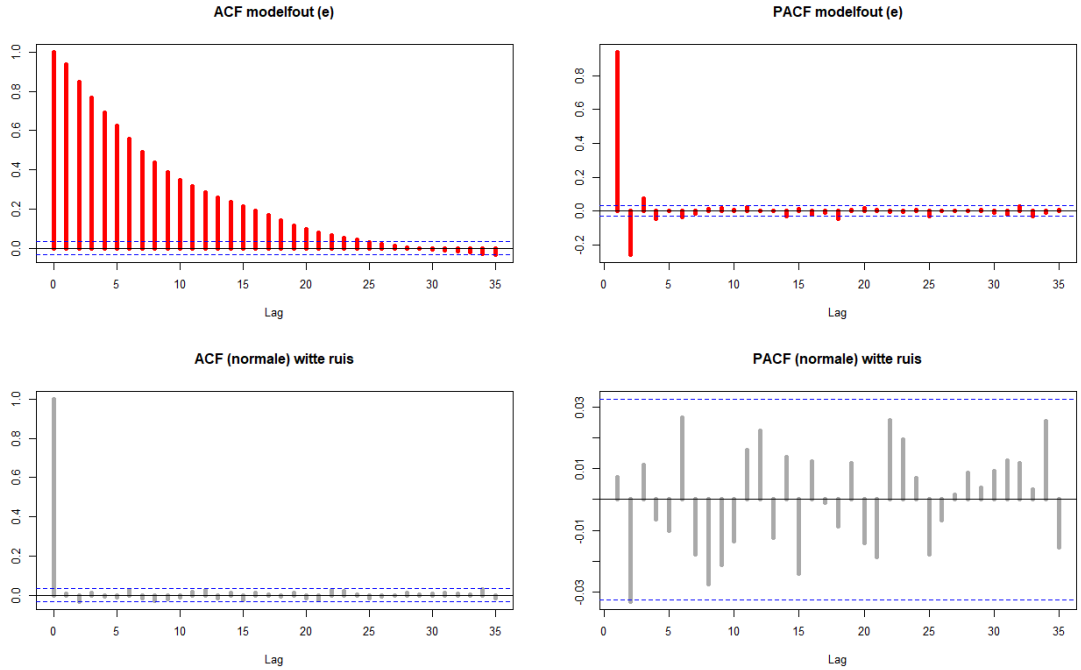
Bij simulatie-exercitie A is een relatie aangebracht tussen variabele *Z* en variabele *X* ($\beta_1$ = 1). Figuur 2.1 toont de tijdreeksen die een rol speelden bij één van de 950.000 voor deze exercitie uitgevoerde simulaties. Het betreft één van de 5.000 simulaties waarbij ruis is gegenereerd volgens het ruismodel met de combinatie $\phi_1$ = 0,9 en $\theta_1$ = -0,3. Aan enkele duidelijke stapveranderingen is te zien dat variabele *Z* gerelateerd is aan variabele *X* (figuur 2.1-rechtsboven). En de hoge $\phi_1$-waarde (0,9) van de toegevoegde ruis uit zich in zichtbare persistentie van de modelfout *e* (figuur 2.1-onder).

*Figuur 2.1: De voor de simulaties gehanteerde tijdreeks van variabele X, evenals voorbeelden van de tijdreeksen van variabele Z en de modelfout (e) voor één van de 5.000 simulaties volgens een ruismodel met $\phi_1$ = 0,9 en $\theta_1$ = -0,3. De simulatie-instellingen zijn: $\beta_0$ = 10, $\beta_1$ = 1 en $\sigma_N$ = 2.*



Om te kunnen beoordelen of de modelfout witte ruis vormt, toont figuur 2.2-boven zijn ACF (autocorrelatiefunctie) en PACF (partiële autocorrelatiefunctie)[5]. Daaruit blijkt dat er duidelijk geen sprake is van witte ruis, maar van gekleurde ruis. De weergegeven combinatie van ACF en PACF wijst – zoals te verwachten - op een ARMA(1,1)-correlatiestructuur, met hoge $\phi_1$-waarde.

*Figuur 2.2: Autocorrelatiefunctie en partiële autocorrelatiefunctie van de in figuur 2.1 getoonde tijdreeks van de modelfout (e), met daarbij ook het 95%-betrouwbaarheidsinterval dat geldt onder de nulhypothese dat het normaal verdeelde witte ruis betreft. Ter vergelijking zijn deze diagnostieken daaronder ook weergegeven voor (normaal verdeelde) witte ruis.*




## Check op dekkingsgraad van betrouwbaarheidsinterval van $b_1$ indien $\beta_1 = 1$

Voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$ is vastgesteld in hoeveel van de 5.000 simulaties de ingestelde waarde van $\beta_1$ binnen het 95%-betrouwbaarheidsinterval van zijn schatting $b_1$ ligt (formule [2.8]). Het zo bepaalde percentage duiden we aan als de empirische dekkingsgraad van het betrouwbaarheidsinterval. Als het model voldoet aan de in § 2.1 vermelde randvoorwaarden zal volgens de statistische theorie de empirische dekkingsgraad gelijk zijn aan de nominale dekkingsgraad (hier 95%). Tabel 2.1 vermeldt de empirische dekkingsgraad van $\beta_1$ voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$, bij een nominale dekkingsgraad van 95%.

[5] Een partiële autocorrelatiefunctie (PACF) is een grafiek die de partiële autocorrelatiecoëfficiënt weergeeft als functie van het aantal tijdseenheden tussen de waarden. Het is de autocorrelatiecoëfficiënt ná correctie voor de lineaire samenhang tussen reekswaarden die zijn gescheiden door minder tijdseenheden. Een ACF en PACF samen geven een completer beeld van de achterliggende ARMA-structuur dan alleen een ACF en stellen zo in staat het ruismodel beter te identificeren.


© AMO en PB Icastat

*Tabel 2.1: Empirische dekkingsgraad van het betrouwbaarheidsinterval van $\beta_1$ voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$ van simulatie-experiment A, bij een nominale dekkingsgraad van 95%.*

| $\beta_1 = 1$ | $\theta_1$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\phi_1$ | -0,9 | -0,8 | -0,7 | -0,6 | -0,5 | -0,4 | -0,3 | -0,2 | -0,1 | 0 | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 | 0,9 |
| 0 | 84% | 84% | 84% | 84% | 85% | 87% | 88% | 90% | 93% | 95% | 97% | 99% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| 0,1 | 82% | 81% | 82% | 82% | 83% | 84% | 85% | 87% | 90% | 92% | 95% | 97% | 99% | 100% | 100% | 100% | 100% | 100% | 100% |
| 0,2 | 78% | 78% | 79% | 80% | 81% | 81% | 82% | 84% | 85% | 88% | 92% | 95% | 97% | 99% | 100% | 100% | 100% | 100% | 100% |
| 0,3 | 76% | 75% | 77% | 76% | 78% | 76% | 79% | 80% | 83% | 85% | 87% | 92% | 95% | 98% | 99% | 100% | 100% | 100% | 100% |
| 0,4 | 72% | 72% | 73% | 73% | 73% | 74% | 74% | 76% | 78% | 81% | 84% | 87% | 91% | 95% | 99% | 100% | 100% | 100% | 100% |
| 0,5 | 68% | 67% | 67% | 68% | 68% | 69% | 69% | 71% | 73% | 74% | 77% | 81% | 85% | 91% | 95% | 99% | 100% | 100% | 100% |
| 0,6 | 62% | 61% | 64% | 61% | 64% | 63% | 63% | 63% | 65% | 66% | 70% | 71% | 77% | 82% | 89% | 95% | 99% | 100% | 100% |
| 0,7 | 56% | 56% | 55% | 55% | 55% | 56% | 57% | 57% | 58% | 58% | 61% | 63% | 67% | 72% | 78% | 87% | 95% | 100% | 100% |
| 0,8 | 46% | 46% | 45% | 47% | 46% | 47% | 47% | 48% | 48% | 48% | 50% | 50% | 55% | 57% | 62% | 71% | 82% | 95% | 100% |
| 0,9 | 36% | 34% | 33% | 35% | 35% | 35% | 35% | 35% | 34% | 36% | 36% | 36% | 37% | 39% | 41% | 46% | 53% | 69% | 95% |

| Empirische dg | Kleur |
|---|---|
| 96% t/m 100% | |
| 94% t/m 96% | |
| 90% t/m 94% | |
| 75% t/m 90% | |
| 50% t/m 75% | |
| ≤50% | |

Op basis van tabel 2.1 komen we tot de volgende conclusies.

1. Als er sprake is van witte ruis ($\phi_1$ = 0 en $\theta_1$ = 0), is de empirische dekkingsgraad van het betrouwbaarheidsinterval van $b_1$ gelijk aan de nominale dekkingsgraad (hier 95%). Dit is in overeenstemming met de statistische theorie.
2. De empirische dekkingsgraad blijkt ook gelijk aan de nominale dekkingsgraad in de negen gevallen dat $\phi_1 = \theta_1$. Dit kan als volgt worden verklaard. Het ARMA(1,1)-model luidt: $n_t = \phi_1 n_{t-1} + a_t - \theta_1 a_{t-1}$. Uit $\phi_1 = \theta_1$ volgt dan: $n_t = \phi_1 n_{t-1} + a_t - \phi_1 a_{t-1}$. Dit kunnen we uitschrijven als: $n_t - a_t = \phi_1 \cdot (n_{t-1} - a_{t-1})$. Het verschil tussen $n_t$ en $a_t$ neemt in het geval dat $\phi_1 = \theta_1$ dus per tijdstap af met een factor gelijk aan $\phi_1$, zodat als $|\phi_1| < 1$ is, $n_t$ uiteindelijk gelijk wordt aan $a_t$ en daarmee witte ruis vormt. Aangezien de ruisreeks is gegenereerd met een voorafgaande 'opwarmperiode' van 100 tijdstappen[6], zal bij $\phi_1 = \theta_1$ elk van de gesimuleerde ruisreeksen over de gehele 3.652 dagen witte ruis vormen.
3. In alle andere gevallen is de empirische dekkingsgraad ongelijk aan de nominale dekkingsgraad. Dit zijn situaties waar het doen van uitspraken met gekwantificeerde betrouwbaarheden niet meer te verantwoorden is.
4. Voor de gevallen dat $\theta_1 < \phi_1$ is de empirische dekkingsgraad kleiner dan de nominale dekkingsgraad. Het 95%-betrouwbaarheidsinterval is dan te smal en het risico op onterecht detecteren van statistische significantie is dan groter dan ingesteld. De afwijking neemt toe met toenemende $\phi_1$ en met kleiner of negatiever wordende $\theta_1$, zij het dat deze vanaf $\theta_1$ ergens tussen -0,7 en -0,8 weer licht afneemt.
5. Voor de gevallen dat $\theta_1 > \phi_1$ is de empirische dekkingsgraad groter dan de nominale dekkingsgraad. Het 95%-betrouwbaarheidsinterval is dan te breed, wat neerkomt op verlies aan onderscheidend vermogen[7]. De afwijking neemt toe met toenemende $\theta_1$ en met kleiner wordende $\phi_1$,

[6] De ruisreeksen zijn gesimuleerd met de functie *arima.sim* van de module *stats* van R, een statistische programmeertaal (versie 3.6.2). R Core Team (2013). R: *A language and environment for statistical computing*.

[7] Het onderscheidend vermogen is de kans om een werkelijk optredend verschil van een bepaalde grootte statistisch significant te detecteren.

maar kan (uiteraard) niet meer bedragen dan 5% (het maximum van de empirische dekkingsgraad is immers 100%).

6. De afwijkingen lopen in deze simulatie op tot maximaal 62%, aangezien de geconstateerde minimale dekkingsgraad 33% bedraagt, namelijk bij de combinatie $\phi_1$ = 0,9 en $\theta_1$ = -0,7.

We kunnen hieruit concluderen dat het niet modelmatig verdisconteren van gekleurde ruis het betrouwbaarheidsinterval van een geschatte relatie sterk kan vertekenen. De toetsrisico's (op onterecht verwerpen respectievelijk onterecht niet verwerpen van de nulhypothese) zijn dan ongelijk aan degene waar de analist rekening mee houdt.

### Zuiverheid van de schatter van $\beta_1$

Tenslotte is ook nagegaan of de schatter van $\beta_1$ zuiver is. Dit is volgens de definitie van zuiverheid het geval als E[$b_1$] = $\beta_1$, dus als de verwachtingswaarde van de schatting gelijk is aan de werkelijke waarde (die bij deze exercitie A is ingesteld op 1).

In onze simulatie-experimenten zijn de modelparameters $\beta_0$ en $\beta_1$ geschat met de methode van de Kleinste-Kwadraten, waarbij de schattingen volgen uit het minimaliseren van de kwadraatsom van de modelfout. Die kwadraatsom - hier aangeduid als $S(b_0,b_1)$ - is namelijk een functie van de schattingen $b_0$ en $b_1$ van de modelparameters, volgens:

$$S(b_0, b_1) \equiv \sum_{t=1}^{k} e_t^2 = \sum_{t=1}^{k} (z_t - b_0 - b_1 \cdot x_t)^2 \qquad [2.9]$$

Het minimum van deze kwadraatsom treedt op voor díe combinatie van $b_0$ en $b_1$ waar de partiële afgeleiden van $S(b_0,b_1)$ naar $b_0$ en $b_1$ nul zijn. Daaruit volgen dan ook de schatters voor beide parameters. De schatter voor $b_1$ is:

$$b_1 = \frac{\sum_{t=1}^{k}(x_t - \bar{x}) \cdot z_t}{\sum_{t=1}^{k}(x_t - \bar{x})^2} \qquad [2.10]$$

Als we door substitutie van $z_t$ deze formule verder uitwerken tot:

$$b_1 = \frac{\sum_{t=1}^{k}(x_t - \bar{x}) \cdot (\beta_0 + \beta_1 \cdot x_t + e_t)}{\sum_{t=1}^{k}(x_t - \bar{x})^2} \qquad [2.11]$$

kan vrij eenvoudig worden afgeleid dat deze schatter zuiver is (E[$b_1$] = $\beta_1$), onder de randvoorwaarde dat de verwachtingswaarde van de modelfout nul is (E($e$) = 0). Blijkbaar maakt het voor de zuiverheid van de schatter niet uit of de reeks van de modelfout wel of geen correlatiestructuur vertoont. Voor de praktijk betekent dit dat de schatting $b_1$ kan worden gepresenteerd als zijnde een zuivere schatting, maar dat gekwantificeerde uitspraken over zijn betrouwbaarheid pas mogelijk zijn als is aangetoond dat de reeks van de modelfout normaal verdeelde witte ruis vormt.

Louter ter illustratie hebben we dit ook voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$ statistisch getoetst met de t-toets. Per combinatie van $\phi_1$ en $\theta_1$ is het gemiddelde en de standaardafwijking bepaald van de 5.000 schattingen van $b_1$. Daarmee kon vastgesteld wat onder geldigheid van de nulhypothese (E[$b_1$] = $\beta_1$) de tweezijdige overschrijdingskans ($p_{2z}$) is van de toetsingsgrootheid $T$ van de t-toets, volgens:

$$p_{2z} = 2 \cdot \text{Kans}\left(t_{4.999} > T = \left|\frac{\bar{b}_1 - \beta_1}{s_{b_1}/\sqrt{5.000}}\right|\right) \qquad [2.12]$$

waarin $t_{4.999}$ een willekeurige trekking uit de Student-t-kansverdeling met 4.999 vrijheidsgraden[8] en $\bar{b}_1$ en $s_{b_1}$ respectievelijk het gemiddelde en de standaardafwijking van de 5.000 schattingen van $b_1$.

[8] Het aantal vrijheidsgraden van de t-toets op een gemiddelde is het aantal waarden (hier 5.000) min 1.



© AMO en PB Icastat 

De resultaten zijn vermeld in tabel 2.2 voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$.

*Tabel 2.2: Voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$ de tweezijdige overschrijdingskans van de toetsingsgrootheid van de t-toets op zuiverheid van de schatter van $b_1$, onder de nulhypothese dat deze schatter zuiver is. Als deze kans kleiner is dan 5%, kan de nulhypothese met 95% betrouwbaarheid worden verworpen. Die gevallen zijn rood gekleurd.*

| $\beta_1 = 1$ | $\theta_1$ | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | -0,9 | -0,8 | -0,7 | -0,6 | -0,5 | -0,4 | -0,3 | -0,2 | -0,1 | 0 | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 | 0,9 | Gemid |
| 0 | 41% | 78% | 50% | 15% | 74% | 75% | 90% | 22% | 22% | 37% | 74% | 2% | 68% | 38% | 54% | 77% | 59% | 94% | 88% | 56% |
| 0,1 | 97% | 66% | 84% | 34% | 27% | 59% | 86% | 8% | 37% | 88% | 40% | 14% | 49% | 69% | 83% | 30% | 86% | 55% | 10% | 54% |
| 0,2 | 5% | 2% | 86% | 97% | 17% | 2% | 63% | 87% | 33% | 56% | 73% | 60% | 35% | 80% | 50% | 4% | 8% | 61% | 37% | 45% |
| 0,3 | 91% | 18% | 11% | 72% | 70% | 45% | 47% | 97% | 32% | 67% | 46% | 16% | 87% | 7% | 19% | 40% | 24% | 35% | 54% | 46% |
| 0,4 | 65% | 50% | 21% | 98% | 70% | 68% | 15% | 89% | 75% | 65% | 98% | 39% | 57% | 39% | 57% | 99% | 6% | 90% | 33% | 60% |
| 0,5 | 24% | 99% | 15% | 80% | 78% | 16% | 70% | 74% | 43% | 34% | 98% | 1% | 75% | 17% | 61% | 79% | 26% | 72% | 15% | 51% |
| 0,6 | 70% | 81% | 52% | 2% | 63% | 74% | 82% | 18% | 62% | 65% | 79% | 60% | 75% | 31% | 34% | 42% | 20% | 24% | 11% | 50% |
| 0,7 | 67% | 31% | 22% | 46% | 34% | 9% | 98% | 95% | 99% | 21% | 31% | 50% | 14% | 52% | 85% | 51% | 35% | 95% | 33% | 51% |
| 0,8 | 61% | 11% | 97% | 3% | 16% | 28% | 68% | 96% | 97% | 91% | 51% | 39% | 60% | 21% | 54% | 11% | 38% | 19% | 69% | 49% |
| 0,9 | 7% | 15% | 15% | 28% | 31% | 73% | 80% | 44% | 71% | 79% | 21% | 95% | 76% | 3% | 49% | 9% | 1% | 60% | 22% | 41% |
| Gemid | 53% | 45% | 45% | 47% | 48% | 45% | 70% | 63% | 57% | 60% | 61% | 38% | 60% | 36% | 55% | 44% | 30% | 61% | 37% | 50% |

($\phi_1$)

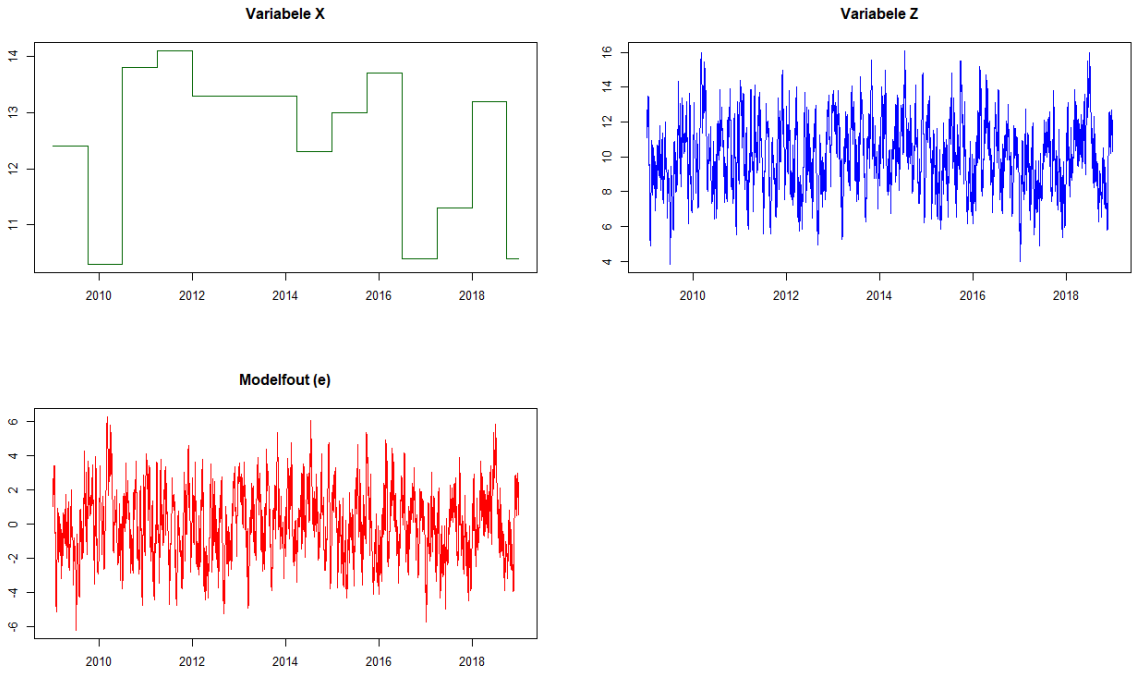
Uit tabel 2.2 blijkt dat de nulhypothese van een zuivere $\beta_1$-schatter voor 9 van de 190 combinaties van $\phi_1$ en $\theta_1$ is verworpen, maar de betreffende gevallen lijken min of meer toevallig verdeeld over het combinatiebereik. Aangezien er is getoetst met 95% betrouwbaarheid is te verwachten dat voor gemiddeld 5% van de 190 combinaties onterecht de nulhypothese zal worden verworpen. Hier is de nulhypothese verworpen voor 9 combinaties, wat neerkomt op 4,7%. Bij de steekproefomvang van 190 loopt het 95%-betrouwbaarheidsinterval van dat percentage van 2,2% tot 8,8%[9] en omvat ook 5%, zodat de nulhypothese van een zuivere $\beta_1$-schatter niet wordt verworpen. Deze simulatie bevestigt daarmee wat al bleek uit de voorgaande theoretische afleiding, namelijk dat de $\beta_1$-schatter zuiver is, ongeacht of de modelfouten wel of geen witte ruis vormen.

### *2.2.3 Resultaten van simulatie-exercitie B*

Bij simulatie-exercitie B is géén relatie aangebracht tussen variabele *Z* en variabele *X* ($\beta_1 = 0$). Figuur 2.3 toont de tijdreeksen die een rol speelden bij één van de 950.000 voor deze exercitie uitgevoerde simulaties. Het betreft één van de 5.000 simulaties waarbij ruis is gegenereerd volgens het ruismodel met de combinatie $\phi_1 = 0{,}9$ en $\theta_1 = -0{,}3$. Er zijn geen visueel waarneembare aanwijzingen dat variabele *Z* gerelateerd is aan variabele *X* (figuur 2.3-boven).

[9] Uitgaande van de eigenschappen van de binomiale kansverdeling.

*Figuur 2.3: De voor de simulaties gehanteerde tijdreeks van variabele X, evenals voorbeelden van de tijdreeksen van variabele Z en de modelfout (e) voor één van de 5.000 simulaties volgens een ruis-model met $\phi_1$ = 0,9 en $\theta_1$ = -0,3. De simulatie-instellingen zijn: $\beta_0$ = 10, $\beta_1$ = 0 en $\sigma_N$ = 2.*




## Check op dekkingsgraad van betrouwbaarheidsinterval van $b_1$ indien $\beta_1$ = 0

Net als bij simulatie-exercitie A is voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$ de empirische dekkingsgraad vastgesteld van het betrouwbaarheidsinterval van $\beta_1$, bij een nominale dekkingsgraad van 95%. De resultaten zijn vermeld in tabel 2.3.

*Tabel 2.3: Empirische dekkingsgraad van het betrouwbaarheidsinterval van $\beta_1$ voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$ van simulatie-experiment B, bij een nominale dekkingsgraad van 95%.*

| $\beta_1$ = 0 | $\theta_1$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | -0,9 | -0,8 | -0,7 | -0,6 | -0,5 | -0,4 | -0,3 | -0,2 | -0,1 | 0 | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 | 0,9 |
| 0 | 84% | 83% | 84% | 84% | 86% | 86% | 89% | 90% | 93% | 95% | 97% | 99% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| 0,1 | 80% | 80% | 82% | 83% | 83% | 85% | 86% | 88% | 90% | 92% | 95% | 97% | 99% | 100% | 100% | 100% | 100% | 100% | 100% |
| 0,2 | 79% | 78% | 78% | 80% | 80% | 81% | 83% | 84% | 86% | 89% | 93% | 95% | 97% | 99% | 100% | 100% | 100% | 100% | 100% |
| 0,3 | 76% | 76% | 75% | 77% | 77% | 78% | 79% | 80% | 83% | 85% | 88% | 92% | 95% | 98% | 100% | 100% | 100% | 100% | 100% |
| 0,4 | 71% | 71% | 72% | 72% | 72% | 74% | 75% | 75% | 77% | 80% | 83% | 87% | 91% | 95% | 98% | 100% | 100% | 100% | 100% |
| 0,5 | 68% | 68% | 68% | 68% | 69% | 69% | 69% | 71% | 72% | 75% | 77% | 80% | 87% | 90% | 95% | 99% | 100% | 100% | 100% |
| 0,6 | 62% | 62% | 61% | 63% | 63% | 62% | 63% | 66% | 65% | 67% | 69% | 72% | 77% | 82% | 89% | 95% | 99% | 100% | 100% |
| 0,7 | 55% | 56% | 55% | 56% | 57% | 56% | 55% | 56% | 57% | 60% | 60% | 64% | 66% | 72% | 77% | 87% | 95% | 100% | 100% |
| 0,8 | 48% | 47% | 46% | 48% | 48% | 48% | 48% | 49% | 48% | 49% | 49% | 52% | 54% | 57% | 61% | 70% | 81% | 95% | 100% |
| 0,9 | 35% | 35% | 34% | 34% | 35% | 35% | 34% | 34% | 34% | 35% | 36% | 36% | 36% | 38% | 42% | 45% | 54% | 70% | 95% |

| Empirische dg | Kleur |
|---|---|
| 96% t/m 100% | |
| 94% t/m 96% | |
| 90% t/m 94% | |
| 75% t/m 90% | |
| 50% t/m 75% | |
| ≤50% | |

Het blijkt dat tabel 2.3 een vrijwel identiek beeld oplevert als de resultaten van simulatie-exercitie A, die zijn vermeld in tabel 2.1. We mogen aannemen dat de kleine verschillen komen door stochastische variatie, aangezien er sprake is van een eindige steekproefgrootte. De conclusie luidt


© AMO en PB Icastat

dan dat de onder tabel 2.1 vermelde bevindingen onafhankelijk zijn van het wel of niet bestaan van een relatie tussen de uitvoerreeks (*Z*) en de invoerreeks (*X*).

### Check op significantieniveau bij toetsen op relatie tussen *Z* en *X* indien $\beta_1 = 0$

Omdat bij deze simulatie-exercitie B de regressiecoëfficiënt $\beta_1$ is ingesteld op nul, is er geen sprake van een relatie tussen *Z* en *X*. Dit stelt in staat na te gaan in welke mate het niet verdisconteren van gekleurde ruis het risico kan beïnvloeden op onterecht concluderen dat er sprake is van een statistisch significante relatie. We zijn daartoe voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$ nagegaan of het empirische risico op zo'n onterechte conclusie - het empirische significantieniveau $\alpha_{emp}$ - overeenstemt met het nominale risico daarop - het nominale significantieniveau $\alpha_{nom}$ (dat we hier hebben ingesteld op 5%). Het empirisch significantieniveau is voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$ bepaald als 1 - de empirische dekkingsgraad van het 95%-betrouwbaarheidsinterval van $\beta_1$ (zoals hiervoor bepaald en vermeld in tabel 2.3). Die empirische dekkingsgraad is immers het percentage van de 5.000 simulaties waarbij de ingestelde waarde van $\beta_1$ (= 0) binnen het 95%-betrouwbaarheidsinterval van zijn schatting $b_1$ ligt (formule [2.8]). Zijn complement is dan het percentage van de 5.000 simulaties waarbij de ingestelde waarde van $\beta_1$ (= 0) buiten dat interval ligt en waarbij $b_1$ dus statistisch significant verschilt van 0, duidend op een statistisch significante relatie tussen *Z* en *X*. De aldus afgeleide empirische significantieniveaus zijn vermeld in tabel 2.4.

*Tabel 2.4: Empirisch significantieniveau bij statistisch toetsen op de nulhypothese dat $\beta_1 = 0$, voor elk van de 190 combinaties van $\phi_1$ en $\theta_1$ van simulatie-experiment B, bij een nominaal significantieniveau van 5%.*

| $\beta_1 = 0$ | $\theta_1$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | -0,9 | -0,8 | -0,7 | -0,6 | -0,5 | -0,4 | -0,3 | -0,2 | -0,1 | 0 | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 | 0,9 |
| 0 | 16% | 17% | 16% | 16% | 14% | 14% | 11% | 10% | 7% | 5% | 3% | 1% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| 0,1 | 20% | 20% | 18% | 17% | 17% | 15% | 14% | 12% | 10% | 8% | 5% | 3% | 1% | 0% | 0% | 0% | 0% | 0% | 0% |
| 0,2 | 21% | 22% | 22% | 20% | 20% | 19% | 17% | 16% | 14% | 11% | 7% | 5% | 3% | 1% | 0% | 0% | 0% | 0% | 0% |
| 0,3 | 24% | 24% | 25% | 23% | 23% | 22% | 21% | 20% | 17% | 15% | 12% | 8% | 5% | 2% | 0% | 0% | 0% | 0% | 0% |
| 0,4 | 29% | 29% | 28% | 28% | 28% | 26% | 25% | 25% | 23% | 20% | 17% | 13% | 9% | 5% | 2% | 0% | 0% | 0% | 0% |
| 0,5 | 32% | 32% | 32% | 32% | 32% | 31% | 31% | 29% | 28% | 25% | 23% | 20% | 13% | 11% | 5% | 1% | 0% | 0% | 0% |
| 0,6 | 38% | 38% | 39% | 37% | 37% | 38% | 37% | 34% | 35% | 33% | 31% | 28% | 23% | 18% | 11% | 5% | 1% | 0% | 0% |
| 0,7 | 45% | 44% | 45% | 44% | 43% | 44% | 45% | 44% | 43% | 40% | 40% | 36% | 34% | 28% | 23% | 13% | 5% | 0% | 0% |
| 0,8 | 52% | 53% | 54% | 52% | 52% | 52% | 52% | 51% | 52% | 51% | 51% | 48% | 46% | 43% | 39% | 30% | 19% | 5% | 0% |
| 0,9 | 65% | 65% | 66% | 66% | 65% | 65% | 66% | 66% | 66% | 65% | 64% | 64% | 64% | 62% | 59% | 55% | 46% | 30% | 5% |

$\phi_1$

| Empirische α | Kleur |
|---|---|
| 0% t/m 4% | |
| 4% t/m 6% | |
| 6% t/m 10% | |
| 10% t/m 25% | |
| 25% t/m 50% | |
| >50% | |

Op basis van tabel 2.4 komen we tot de volgende conclusies.

1. Als er sprake is van witte ruis ($\phi_1 = 0$ en $\theta_1 = 0$), is het empirische significantieniveau gelijk aan het nominale significantieniveau (hier ingesteld op 5%). Dit is in overeenstemming met de statistische theorie.
2. Het empirische significantieniveau blijkt ook gelijk aan het nominale significantieniveau in de negen gevallen dat $\phi_1 = \theta_1$. Dit is reeds verklaard in punt 2 onder tabel 2.1.
3. In alle andere gevallen is het empirische significantieniveau ongelijk aan het nominale significantieniveau. Dit zijn situaties waar het doen van uitspraken met gekwantificeerde betrouwbaarheden niet meer te verantwoorden is.


© AMO en PB Icastat

4. Voor de gevallen dat $\theta_1 < \phi_1$ is het empirische significantieniveau groter dan het nominale significantieniveau. De afwijking neemt toe met toenemende $\phi_1$ en met kleiner of negatiever wordende $\theta_1$, zij het dat deze vanaf $\theta_1$ ergens tussen -0,7 en -0,8 weer licht afneemt.
5. Voor de gevallen dat $\theta_1 > \phi_1$ is het empirische significantieniveau kleiner dan het nominale significantieniveau. De afwijking neemt toe met toenemende $\theta_1$ en met kleiner wordende $\phi_1$, maar kan (uiteraard) niet meer bedragen dan 5% (het minimum van het empirische significantieniveau is immers 0%).
6. De afwijkingen lopen in deze simulatie op tot maximaal 61%, aangezien het geconstateerde maximale significantieniveau 66% bedraagt, zoals bij de combinatie $\phi_1$ = 0,9 en $\theta_1$ = -0,7.

We kunnen hieruit concluderen dat het niet modelmatig verdisconteren van gekleurde ruis de statistische significantie van een geschatte relatie sterk kan vertekenen. Het risico op statistisch significant detecteren van een niet-bestaande relatie kan dan aanzienlijk worden.

## 2.3 Kernpunten van dit hoofdstuk

Als bij tijdreeksmodellering de resterende innovatiereeks geen witte ruis vormt, kan dit leiden tot sterke vertekeningen van zowel het betrouwbaarheidsinterval als de statistische significantie van een geschatte relatie. Voor wat betreft het betrouwbaarheidsinterval is de empirische dekkingsgraad dan ongelijk aan de nominale dekkingsgraad. En voor wat betreft de statistische significantie is het empirische significantieniveau dan ongelijk aan het nominale significantieniveau. Het risico op statistisch significant detecteren van een niet-bestaande relatie kan dan aanzienlijk worden.
Dus als de innovatiereeks geen witte ruis vormt is het doen van uitspraken met gekwantificeerde betrouwbaarheden niet meer te verantwoorden. De zuiverheid van een geschatte relatie is daarentegen onafhankelijk van het wel of niet voldoen aan witte ruis van de innovatiereeks.


© AMO en PB Icastat

# 3 Randvoorwaarden en het Pastas-model

In dit hoofdstuk gaan we na hóe en in welke mate een Pastas-model kan afwijken van de randvoorwaarden die gelden om uitspraken met gekwantificeerde uitspraken te kunnen doen. Daarvoor is materiaal gebruikt van een praktijkstudie die is uitgevoerd met Pastas. Dit stelde ook in staat de mogelijkheden van Pastas te beoordelen en zo verbeterpunten te constateren.

## 3.1 De beschouwde praktijkstudie

Voor deze verkenning zijn we uitgegaan van basisgegevens van de studie *Doorwerking droogte van 2018 in Overijssel – Voorspelling per eind september* [Beekman en Caljé, 2018]. In die studie is aan de hand van tijdreeksmodellering van grondwaterstandreeksen met Pastas onderzocht in welke mate de droogte van de zomer van 2018 zich in Overijssel heeft gemanifesteerd in lage grondwaterstanden en ook of doorwerking van die droge zomer op de voorjaarsgrondwaterstand in 2019 is te verwachten. De studie concludeert dat de droogte van 2018 vrijwel overal heeft geleid tot extreem lage grondwaterstanden, maar dat in de winter in twee deelgebieden (de Kop van Overijssel en het IJsseldal) herstel van het grondwatersysteem is te verwachten. Alleen in de regio Twente bleek structureel sprake van doorwerking van de droge zomer en daar wordt de kans op een normale voorjaarsgrondwaterstand dan ook aanzienlijk lager geschat dan gemiddeld.

De studie omvatte tijdreeksmodellering van de 654 grondwaterstandreeksen van het Provinciaal Meetnet Grondwaterstand van Overijssel, met behulp van Pastas (versie 0.9.7). Elke reeks is gemodelleerd op dagbasis, over de modelperiode 2005 t/m de laatst beschikbare meting in 2018. Als invoerreeks diende de grondwateraanvulling $R$, zoals bepaald uit $R = P - f \cdot E$, waarin $P$ de neerslag, $f$ de verdampingsfactor en $E$ de verdamping. De neerslag is afkomstig van het neerslagstation dat het dichtst bij de betreffende peilbuis ligt en de verdamping is afkomstig van het dichtstbijzijnde weerstation. Voor het modelleren van de dynamische relatie is uitgegaan van de Gamma-transferfunctie. Uiteindelijk zijn alleen díe modellen meegenomen waarbij: i) de reeks minimaal twee jaar met dagwaarden omvat, ii) de reeks langer is dan 99% van de responstijd op neerslag en verdamping en iii) de met het deterministisch deel van het model verklaarde variantie van de grondwaterstand minstens 70% bedraagt. Dit bleek het geval voor 468 van de 654 reeksen (72%).

### Over het ruismodel van Pastas

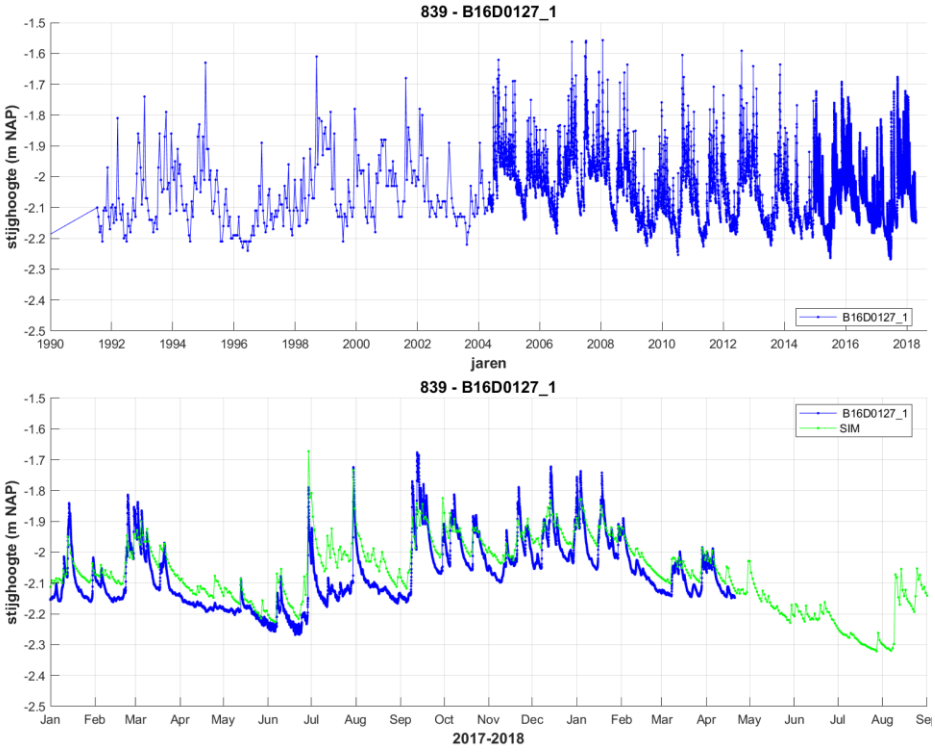
In de bijlage van het betreffende rapport staat beschreven dat Pastas een AR(1)-ruismodel kan verdisconteren (de eenvoudigste vorm van een ruismodel). Maar daar blijkt geen gebruik van te zijn gemaakt, zodat elk tijdreeksmodel alleen een deterministisch deel omvat[10]. Bij modelleren op dagbasis, zoals toegepast in die studie, kan vrijwel nooit zonder ruismodel voldaan worden aan de randvoorwaarde van normaal verdeelde witte ruis (zie § 3.2). Er kunnen dan geen uitspraken met gekwantificeerde betrouwbaarheden worden gedaan, maar dat is bij die studie ook niet gedaan. Een ander nadeel van het weglaten van het ruismodel is echter dat zo de variantie van de grondwaterstand systematisch wordt onderschat. Dit kan bijvoorbeeld worden geïllustreerd aan de hand

[10] De informatie dat bij die studie het ruismodel is weggelaten kregen we van de auteurs, het rapport vermeldt dit niet. Zij gaven als reden dat het percentage verklaarde variantie beduidend groter is als er geen ruismodel wordt gebruikt. Hier zullen we in § 3.4 verder op in gaan.


© AMO en PB Icastat

van  figuur 3.1 (afkomstig uit het betreffende rapport). In de onderste van de twee subplots is duidelijk te zien dat de meetreeks (blauwe lijn) meer variatie vertoont dan het deterministische deel van het tijdreeksmodel (groene lijn). Deze onderschatting van de variantie kan oplopen tot circa 30%, aangezien bij de geaccepteerde tijdreeksmodellen de met het deterministische deel verklaarde variantie van de grondwaterstand minstens 70% bedraagt (criterium iii hierboven). Maar bij het voorspellen van de grondwaterstand en het daaruit afleiden van onder- en overschrijdingskansen is het zaak de spreiding van de grondwaterstand zo goed mogelijk te verdisconteren. Bij niet meenemen van de ruis zullen de betreffende kansen onzuiver worden geschat.

*Figuur 3.1: Voorbeeld van het onderschatten van de variantie van de grondwaterstand bij tijdreeksmodellering op dagbasis zónder toepassing van een ruismodel. Boven: meetreeksplot. Onder: uitvergroting over de periode januari 2017 – medio april 2018 (in blauw) en van het tijdreeksmodel, over de periode januari 2017 t/m september 2018 (in groen). De plots zijn afkomstig uit [Beekman en Caljé, 2018]. De modellijn vertoont duidelijk minder variatie dan de grondwaterstand.*




## Over de voorbewerking van de meetreeksen door Pastas

Het rapport van de beschouwde studie bevat geen detailinformatie over hoe de meetreeksen van grondwaterstand, neerslag en verdamping zijn voorbewerkt voor de tijdreeksmodellering. Vermoedelijk is alleen gebruik gemaakt van de voorbewerkingsmethode die Pastas standaard hanteert. Maar die is niet beschreven in het rapport, zodat we daar geen oordeel over kunnen geven.

Bij de betreffende studie kunnen de details van de voorbewerking extra relevant zijn, omdat de metingen van de invoervariabelen niet parallel lopen. De neerslag is afkomstig van het dichtstbijzijnde neerslagstation en op dagbasis vertegenwoordigt deze de neerslagsom van 8 uur gisteren tot 8 uur vanmorgen. De verdamping is daarentegen afkomstig van het dichtstbijzijnde weerstation en op dagbasis vertegenwoordigt deze de verdampingssom van 0:00 uur tot 24:00 uur vandaag.


© AMO en PB Icastat

## 3.2 Verificatie van Pastas-modellen van praktijkreeksen

Om na te kunnen gaan of een Pastas-model in de praktijk kan voldoen aan bepaalde randvoorwaarden, is uitgegaan van de 654 grondwaterstandsreeksen die zijn gebruikt bij de genoemde studie. We zijn per reeks uitgegaan van de deelreeks die start bij het operationeel worden van de logger. De startdatums liggen daardoor in de periode 1 januari 2005 t/m 3 juli 2018 en de einddatums liggen in de periode 4 september 2017 t/m 20 september 2018. Alleen de reeksen van 5 jaar of langer zijn meegenomen voor onze exercitie, waardoor er 80 reeksen afvielen en 574 reeksen resteerden. Verder hebben we elke reeks equidistant gemaakt op dagbasis, door per reekshiaat voor elke dag de code voor een ontbrekende waarde in te vullen. Dit was nodig omdat het toetsen op autocorrelatie een equidistante ruis- of innovatiereeks vereist. De betreffende toets kan daarbij verdisconteren voor ontbrekende waarden (zie hieronder).

Van elk van de 574 reeksen is een tijdreeksmodel met Pastas geschat, met als invoerreeksen de neerslag en de verdamping. In een eerste sessie zijn alleen modellen geschat zónder ruismodel en in een tweede sessie alleen modellen mét ruismodel. Hierbij zijn alle reeksen gemodelleerd met de standaardinstellingen van Pastas en is dus geen gebruik gemaakt van de mogelijkheid om eventueel optimalere instellingen te hanteren.

### Toetsen op autocorrelatie en normaliteit

Om per tijdreeksmodel te verifiëren of de innovatiereeks kwalificeert als witte ruis zijn de volgende statistische toetsen op autocorrelatie uitgevoerd, met een speciaal daarvoor ontwikkelde verificatiemodule in Python[11]:

- Portmanteau-toets op autocorrelatie – De toetsingsgrootheid is een som van de gekwadrateerde autocorrelatiecoëfficiënten van de innovaties en de toets gaat er van uit dat de innovaties afkomstig zijn uit een normale kansverdeling [Ljung and Box, 1978]. Als de innovatiereeks ontbrekende waarden omvat, is een aangepaste toetsingsgrootheid van deze toets gehanteerd die daar rekening mee houdt [Stoffer and Toloi, 1992]. Zie bijlage 3 voor verdere toelichting.
- Runs-toets op autocorrelatie – Dit is een verdelingsvrije toets[12], met als toetsingsgrootheid het aantal runs. Een run is daarbij een aaneengesloten groep waarden die alle aan dezelfde kant van de reeksmediaan liggen. Deze toets kan ook overweg met ontbrekende waarden van de innovatiereeks. Zie bijlage 3 voor verdere toelichting.

Beide toetsen op autocorrelatie zijn uitgevoerd met 95% betrouwbaarheid.

Verder is getoetst op normaliteit van de innovaties met de D'Agostino-K2-toets [D'Agostino, 1986]. Deze is uitgevoerd met 99% betrouwbaarheid, omdat zelfs een geringe en praktisch niet relevante afwijking van normaliteit al snel statistisch significant kan zijn door het grote aantal reekswaarden (hier meer dan 1.826). Zie ook de kanttekening in het tekstkader hieronder.

**Een bruikbare beoordeling op normaliteit vergt enig pragmatisme**

Bij kleine steekproeven (< 20) hebben toetsen op normaliteit weinig onderscheidend vermogen en kunnen alleen grote afwijkingen van normaliteit detecteren. Bij grote steekproeven (> 250) hebben ze daarentegen

[11] De module is ontwikkeld volgens de PEP8-richtlijnen voor Python-code en zal als product van deze studie beschikbaar worden gesteld voor Pastas-gebruikers.

[12] Een verdelingsvrije toets heeft niet als randvoorwaarde dat de waarden uit een bepaalde specifiek benoemde kansverdeling afkomstig zijn.

veel onderscheidend vermogen en kunnen dan afwijkingen van normaliteit detecteren die praktisch niet relevant zijn. Kleine afwijkingen zijn immers niet relevant omdat toepassingen die uitgaan van de randvoorwaarde van normaliteit in het geval van een grote steekproef vrij robuust zijn tegen afwijkingen van normaliteit. Maar er is geen objectieve methode om vast te stellen tot welke afwijking van normaliteit die robuustheid nog opgaat, zodat dit vaak subjectief moet worden ingeschat, zoals op basis van visuele beoordeling van het histogram van de gegevens.

Bij het modelleren zonder ruismodel bleek de ruisreeks in geen van de 574 gevallen te klasseren als witte ruis, volgens beide toetsen op autocorrelatie. En voor 545 ruisreeksen bleek er geen sprake van normaliteit (94,9% van het aantal reeksen).
Ook bij het modelleren mét ruismodel bleek de innovatiereeks in geen van de 574 gevallen te klasseren als witte ruis, volgens beide toetsen op autocorrelatie. En voor 565 innovatiereeksen bleek er geen sprake van normaliteit (98,4% van het aantal reeksen).

#### Conclusie

Uit deze exercitie blijkt dat Pastas met de huidige structuur van het ruismodel bij op dagbasis modelleren van praktijkreeksen nog niet in staat is de ruis- of innovatiereeks volledig terug te brengen tot witte ruis. Bovendien blijkt de ruis- of innovatiereeks zelden te voldoen aan normaliteit. Aangezien het een omvangrijke en ruimtelijk goed gespreide verzameling van grondwaterstandreeksen betreft uit de Provincie Overijssel, mogen we aannemen dat deze problemen zich vrijwel overal kunnen voordoen. De conclusie is daarom gerechtvaardigd dat Pastas in zijn huidige vorm vaak niet in staat zal stellen om bij modelleren van grondwaterstanden op dagbasis uitspraken met gekwantificeerde betrouwbaarheden te doen.

## 3.3 Voorbeelden van verificatie van Pastas-modellen van praktijkreeksen

Deze paragraaf toont voorbeelden van de verificatie van Pastas-modellen van drie van de 574 grondwaterstandreeksen, zowel zónder als mét ruismodel. Per model zijn de volgende plots weergegeven:

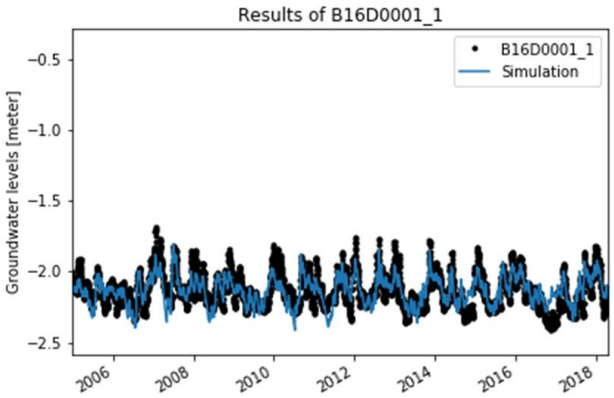
- 1e rij, links: meetreeks en model;
- 1e rij, rechts: ruisreeks en innovatiereeks. Als geen ruismodel is toegepast zet Pastas de innovaties op nul;
- 2e rij, links: ACF (autocorrelatiefunctie) van de ruis indien er geen ruismodel is toegepast, anders is deze van de innovaties;
- 2e rij, rechts: idem, maar over meer tijdsverschuivingen (tot 400 dagen);
- 3e rij, links: PACF (partiële autocorrelatiefunctie) van de ruis indien er geen ruismodel is toegepast, anders is deze van de innovaties;
- 3e rij, rechts: idem, maar over meer tijdsverschuivingen (tot 400 dagen).

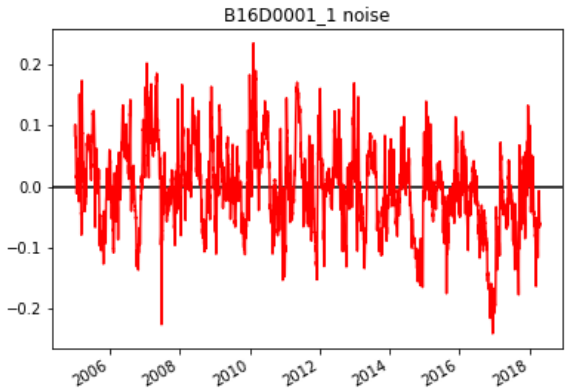
De tabel vermeldt de resultaten en gehanteerde betrouwbaarheden van de volgende statistische toetsen op de ruis of innovaties: D'Agostino-K2-toets op normaliteit, Shapiro-Wilk-toets op normaliteit, Portmanteau-toets op autocorrelatie, Runs-toets op autocorrelatie en Engle-toets op homoscedasticiteit[13]. Ten slotte is aangegeven op welke correlatiestuctuur de combinatie van ACF en PACF van de ruis of innovaties duidt.

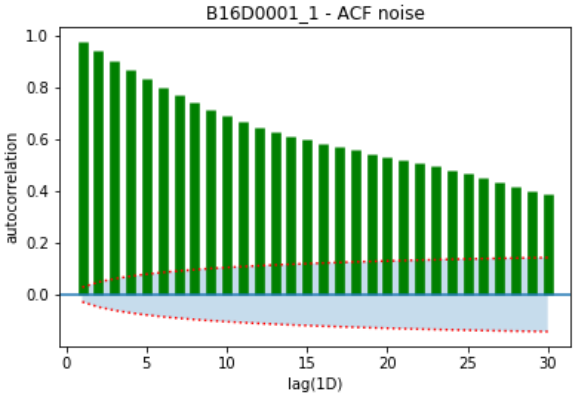
[13] Een reeks vertoont homoscedasticiteit als de variantie van de reekswaarden niet verandert met het niveau.

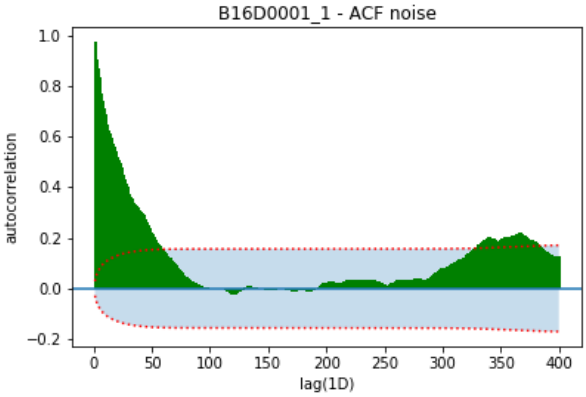

© AMO en PB Icastat

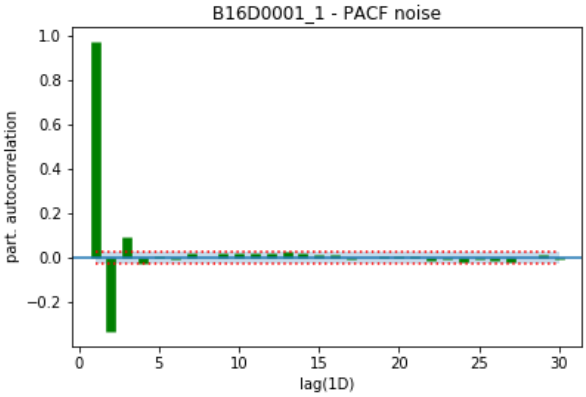
## B16D0001_1 – Zonder ruismodel

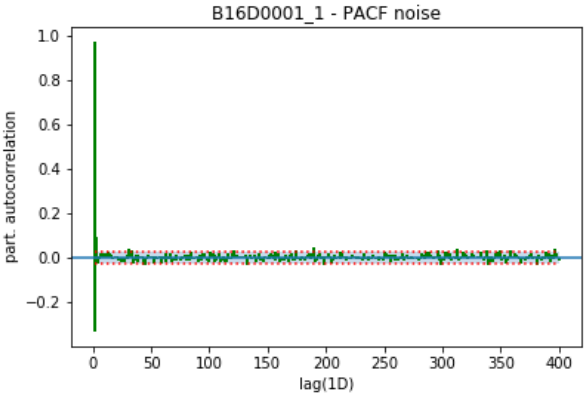


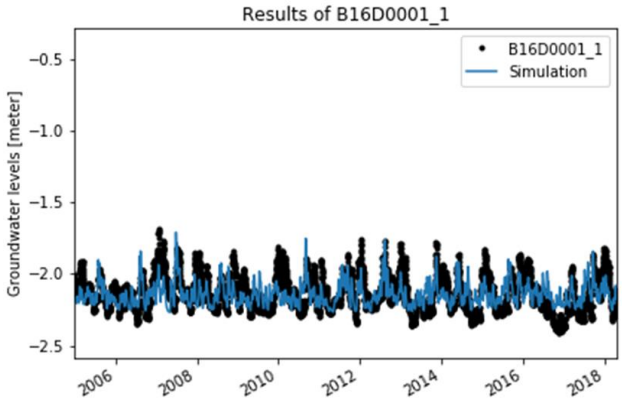

| Test of noise | Test (confidence) | Result | Detail |
|---|---|---|---|
| D'Agostino-K2 | normality (99%) | Yes | p=0.25 |
| Shapiro - Wilk | normality (99%) | No | p=0.00 |
| Portmanteau | no autocorr. (95%) | No | 43088.75>25.00 |
| Runs | no autocorr. (95%) | No | p=0.00 |
| Engle | homosced. (95%) | No | p=0.00 |

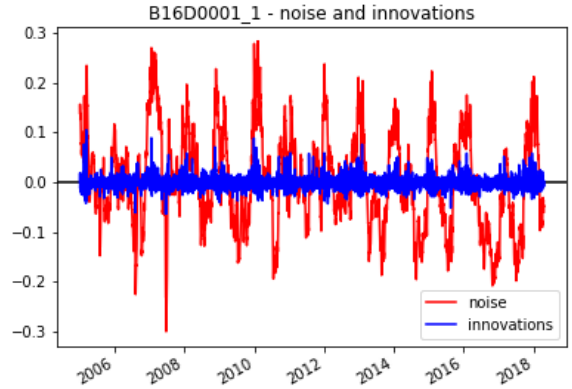
De combinatie van ACF en PACF duidt op een SARMA(1,1)x(1,0)-correlatiestructuur met $\phi_1 \approx 0{,}99$, $\theta_1 \approx -0{,}3$ en $\Phi_1 \approx 0{,}2$ (dit is de 1$^e$-orde autoregressieve seizoensparameter, voor seizoenslengte 365,25 dagen).

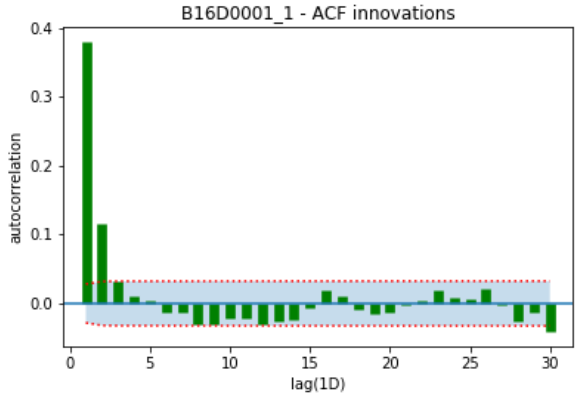


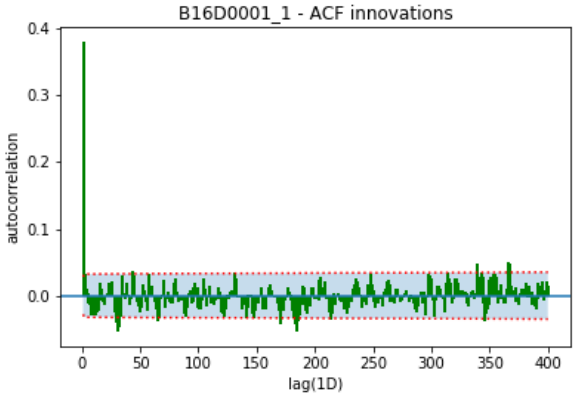
© AMO en PB Icastat 

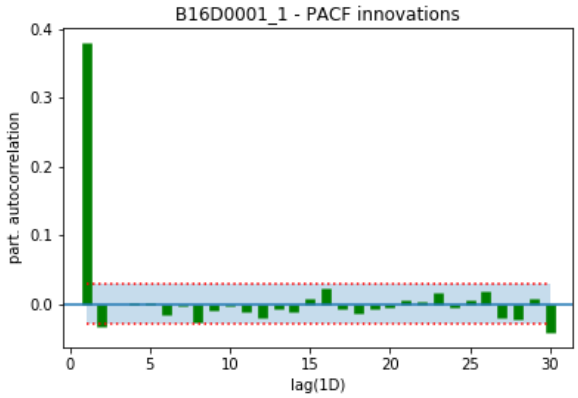
## B16D0001_1 – Met ruismodel

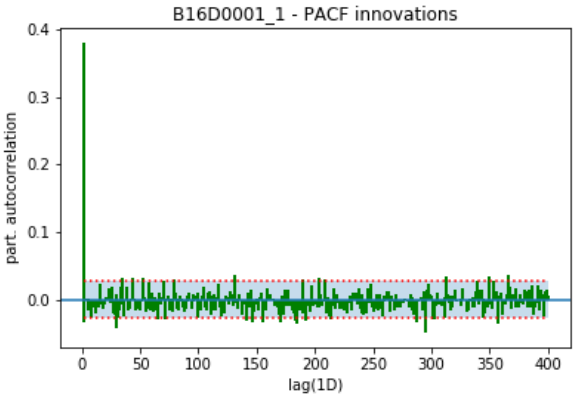


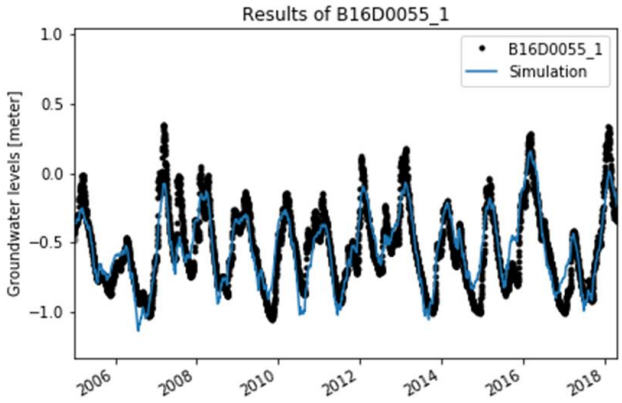



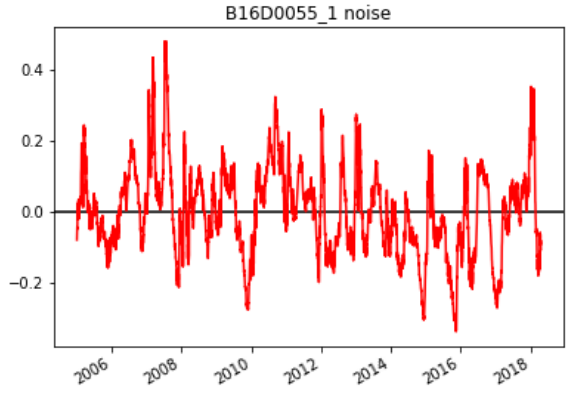

| Test of innovations | Test (confidence) | Result | Detail |
|---|---|---|---|
| D'Agostino-K2 | normality (99%) | No | p=0.00 |
| Shapiro | normality (99%) | No | p=0.00 |
| Portmanteau | no autocorr. (95%) | No | 801.82>23.68 |
| Runs | no autocorr. (95%) | No | p=0.00 |
| Engle | homosced. (95%) | No | p=0.00 |

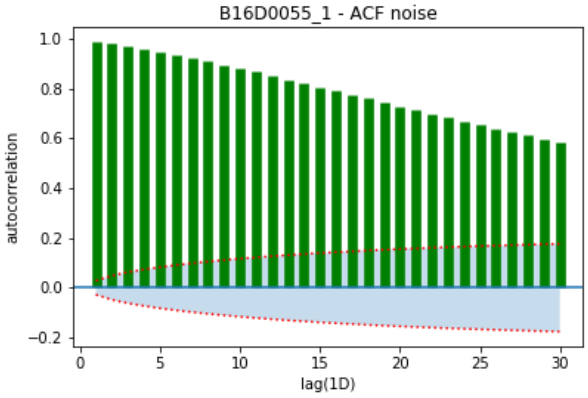
De combinatie van ACF en PACF duidt op een ARMA(1,1)-correlatiestructuur met $\phi_1 \approx 0{,}3$ en $\theta_1 \approx -0{,}1$.

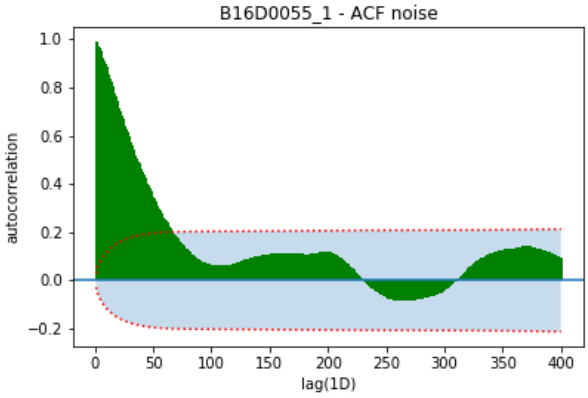

© AMO en PB Icastat

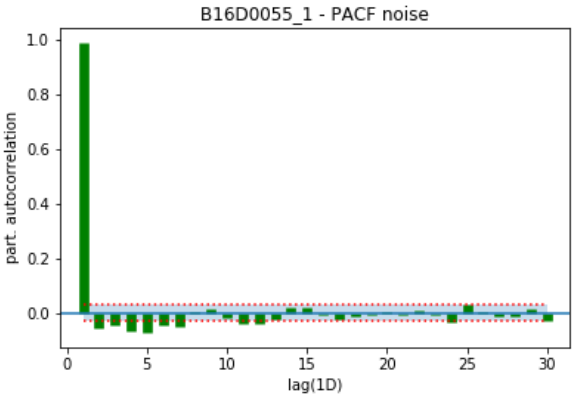
## B16D0055_1 – Zonder ruismodel

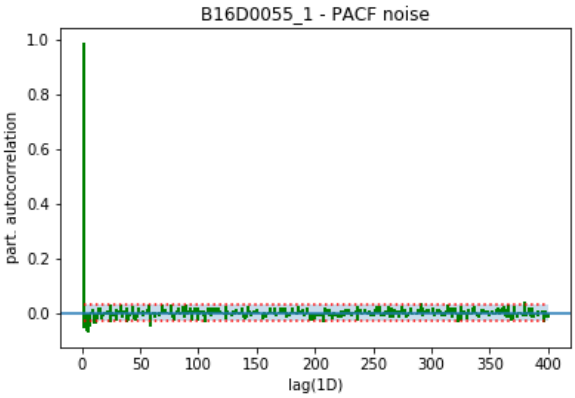


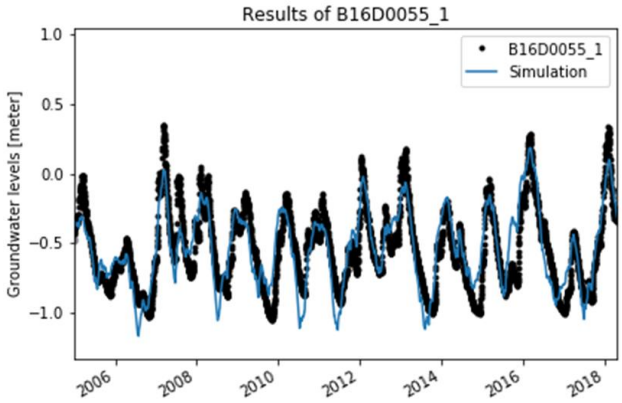



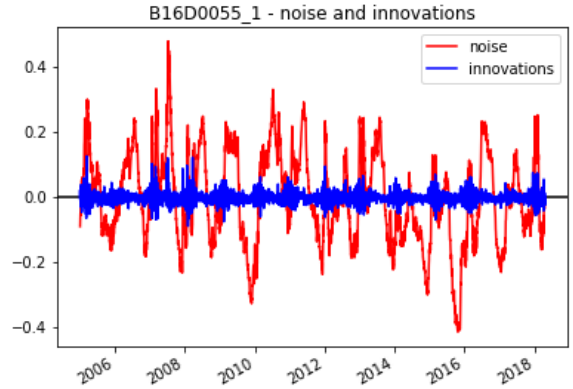

| Test of noise | Test (confidence) | Result | Detail |
|---|---|---|---|
| D'Agostino-K2 | normality (99%) | No | p=0.00 |
| Shapiro | normality (99%) | No | p=0.00 |
| Portmanteau | no autocorr. (95%) | No | 60277.88>25.00 |
| Runs | no autocorr. (95%) | No | p=0.00 |
| Engle | homosced. (95%) | No | p=0.00 |

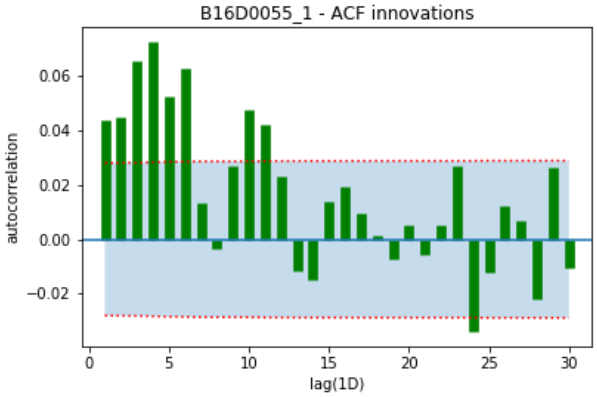
De combinatie van ACF en PACF duidt op een ARMA(1,6)-correlatiestructuur met $\phi_1 \approx 0{,}95$ en $\theta_1$ t/m $\theta_6 \approx -0{,}1$.

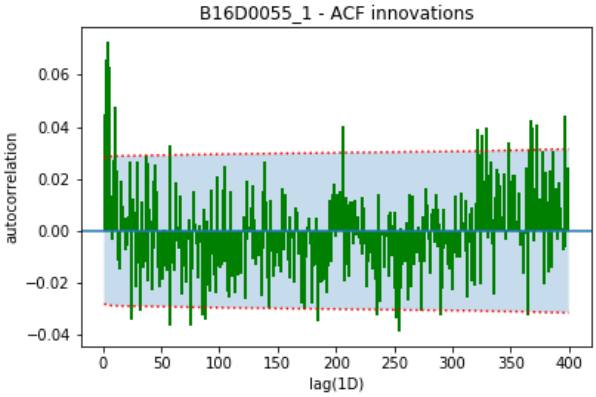

© AMO en PB Icastat

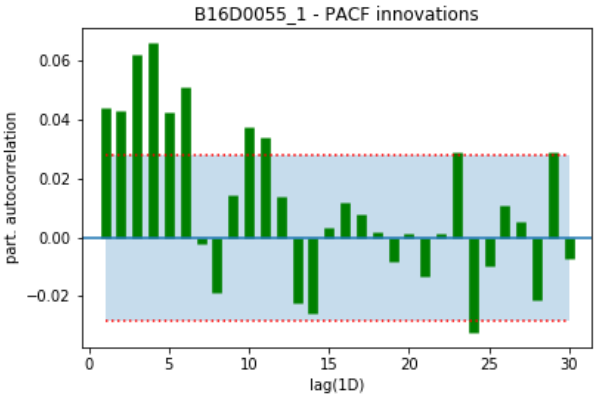
## B16D0055_1 – Met ruismodel

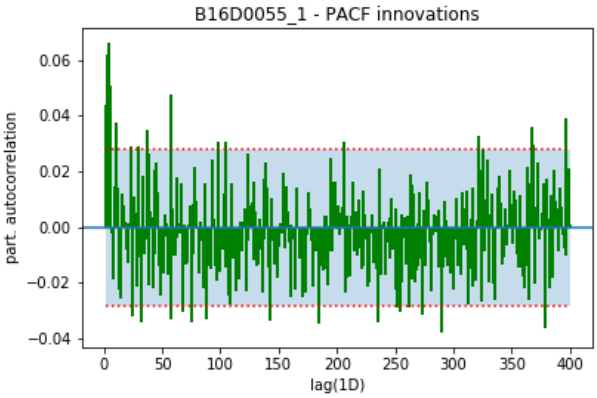


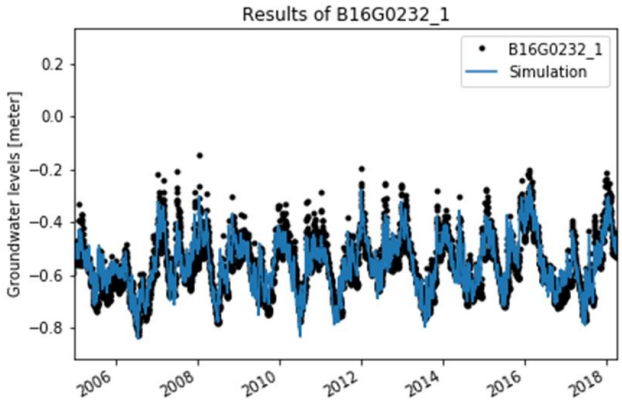

| Test of innovations | Test (confidence) | Result | Detail |
|---|---|---|---|
| D'Agostino-K2 | normality (99%) | No | p=0.00 |
| Shapiro | normality (99%) | No | p=0.00 |
| Portmanteau | no autocorr. (95%) | No | 128.76>23.68 |
| Runs | no autocorr. (95%) | No | p=0.00 |
| Engle | homosced. (95%) | No | p=0.00 |

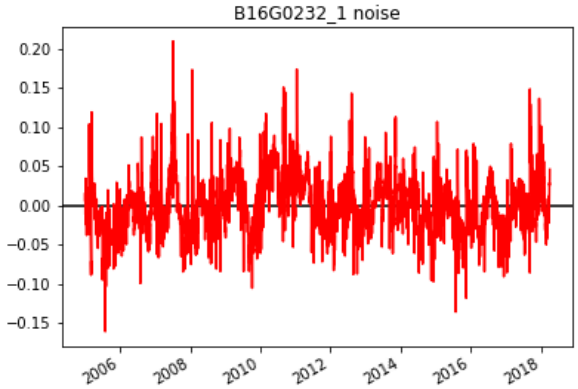
De combinatie van ACF en PACF duidt op een ARMA(0,6)-correlatiestructuur met $\theta_1$ t/m $\theta_6$ ≈ -0,05.

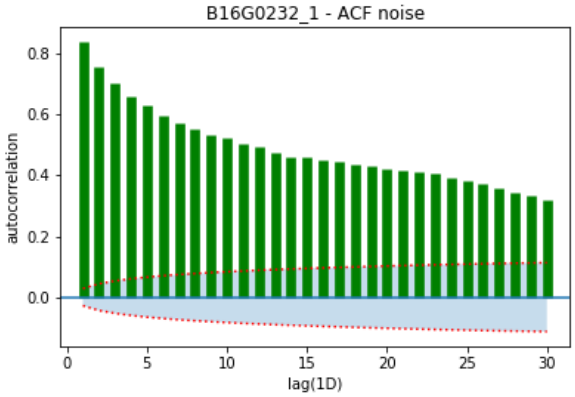


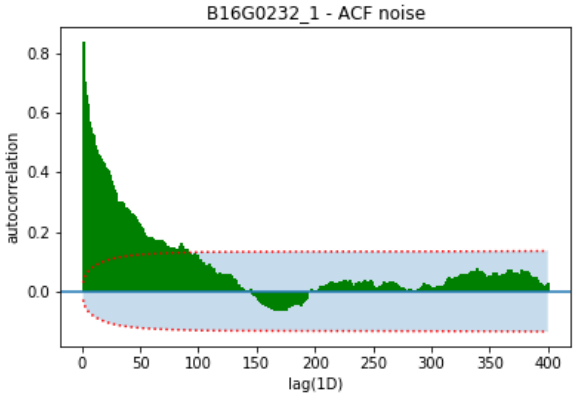
© AMO en PB Icastat 

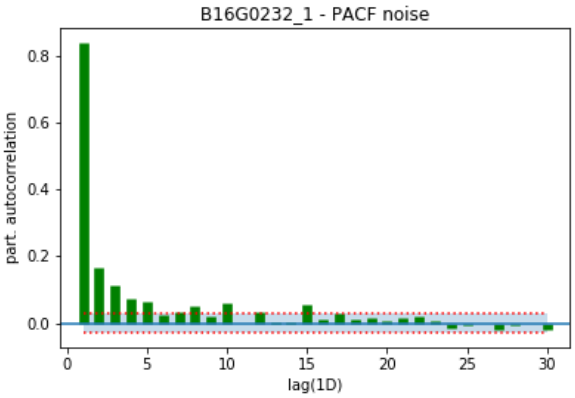
## B16G0232_1 – Zonder ruismodel

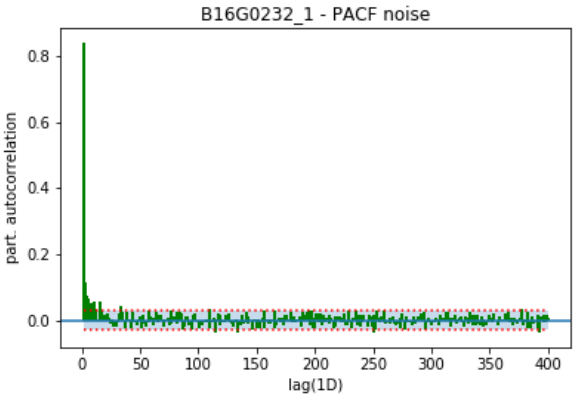


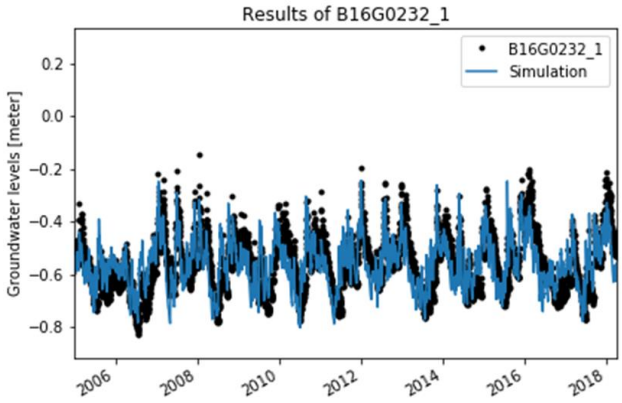



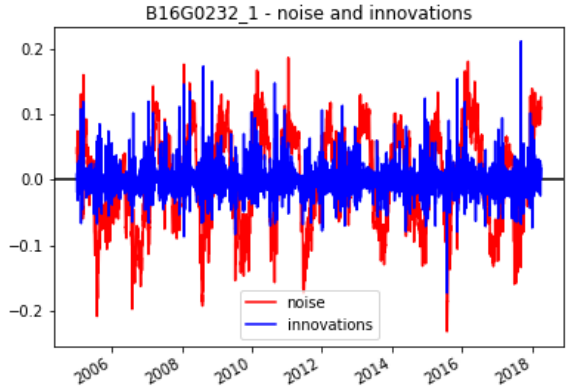

| Test of noise | Test (confidence) | Result | Detail |
|---|---|---|---|
| D'Agostino-K2 | normality (99%) | No | p=0.00 |
| Shapiro | normality (99%) | No | p=0.00 |
| Portmanteau | no autocorr. (95%) | No | 25724.83>25.00 |
| Runs | no autocorr. (95%) | No | p=0.00 |
| Engle | homosced. (95%) | No | p=0.00 |

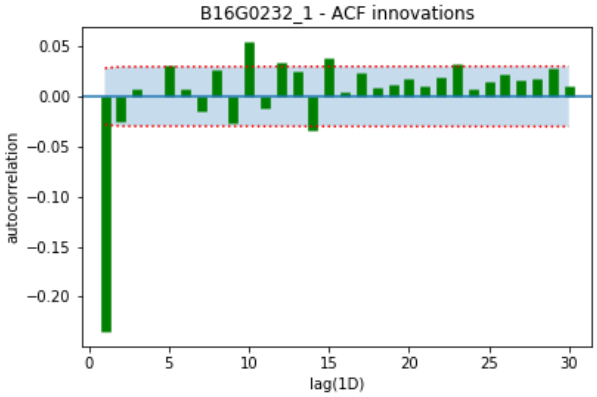
De combinatie van ACF en PACF duidt op een ARMA(1,4)-correlatiestructuur met $\phi_1 \approx 0{,}95$, $\theta_1 \approx 0{,}3$ en $\theta_2$ t/m $\theta_4 \approx 0{,}1$.

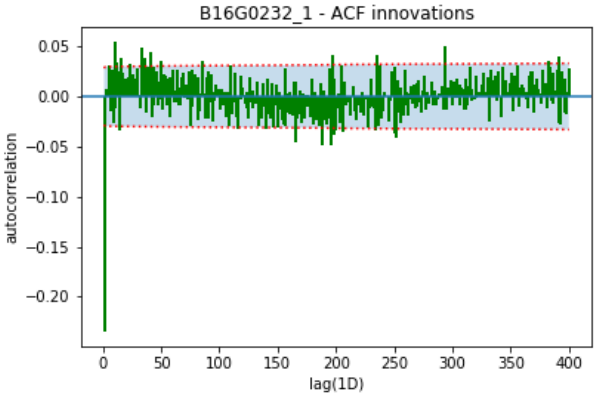

© AMO en PB Icastat

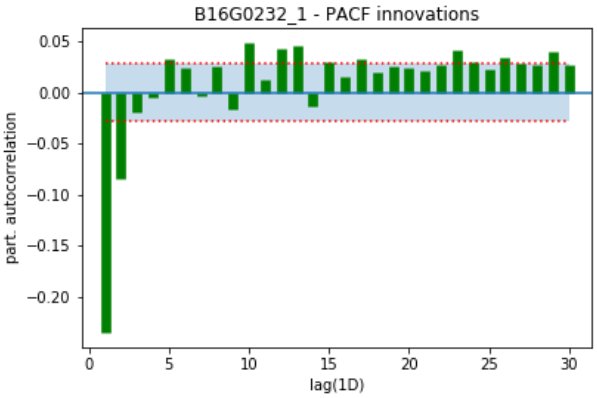
## B16G0232_1 – Met ruismodel

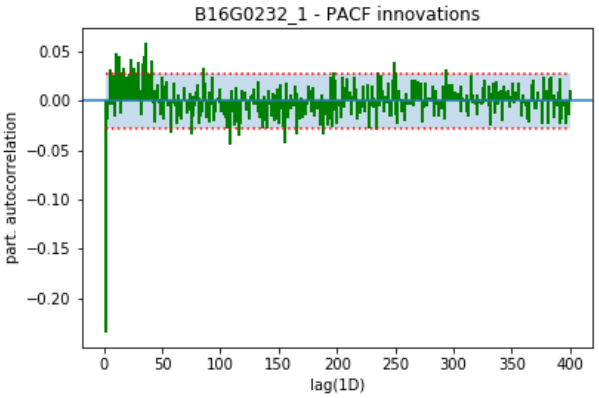



| Test of innovations | Test (confidence) | Result | Detail |
|---|---|---|---|
| D'Agostino-K2 | normality (99%) | No | p=0.00 |
| Shapiro | normality (99%) | No | p=0.00 |
| Portmanteau | no autocorr. (95%) | No | 319.79>23.68 |
| Runs | no autocorr. (95%) | No | p=0.00 |
| Engle | homosced. (95%) | No | p=0.00 |

De combinatie van ACF en PACF duidt op een ARMA(0,1)-correlatiestructuur met $\theta_1 \approx 0{,}3$.


© AMO en PB Icastat

Tabel 3.1 vat voor deze drie grondwaterstandreeksen de correlatiestructuur samen van de restterm van het ontwikkelde tijdreeksmodel op dagbasis. Dit betreft de ruis als er geen ruismodel is toegepast, of de innovaties als er een ruismodel is toegepast.

*Tabel 3.1: Correlatiestructuur van de restterm van het ontwikkelde tijdreeksmodel op dagbasis voor elk van de drie beschouwde grondwaterstandreeksen. Dit betreft de ruis als er geen ruismodel is toegepast (links) en de innovaties als er een ruismodel is toegepast (rechts).*

| Reeks | Model zónder ruismodel<br>correlatiestructuur van de ruis | Model mét ruismodel<br>correlatiestructuur van de innovaties |
|---|---|---|
| **B16D0001_1** | SARMA(1,1)x(1,0): $\phi_1$≈0,99, $\theta_1$≈-0,3 en $\Phi_1$≈0,2 | ARMA(1,1): $\phi_1$≈0,3 en $\theta_1$≈-0,1 |
| **B16D0055_1** | ARMA(1,6): $\phi_1$≈0,95 en $\theta_1$ t/m $\theta_6$≈-0,1 | ARMA(0,6): $\theta_1$ t/m $\theta_6$≈-0,05 |
| **B16G0232_1** | ARMA(1,4): $\phi_1$≈0,95, $\theta_1$≈0,3 en $\theta_2$ t/m $\theta_4$≈0,1 | ARMA(0,1): $\theta_1$≈0,3 |

Uit tabel 3.1 blijkt dat er bij modelleren op dagbasis zonder ruismodel sprake kan zijn van een zeer sterke autocorrelatiestructuur van de ruis ($\phi_1 \geq 0{,}95$), die gepaard gaat met een geringe tot matige moving-average-structuur ($\theta_i \leq 0{,}3$) en soms zelfs ook nog correlatie tussen waarden die een jaar uiteen liggen. Het AR(1)-ruismodel van Pastas blijkt in staat de mate van correlatie te verminderen, maar het kan in geen van deze gevallen de innovaties reduceren tot witte ruis. Dit betreft weliswaar slechts drie voorbeelden, maar uit een steekproefsgewijze verkenning bleek het over een breed front op te treden.

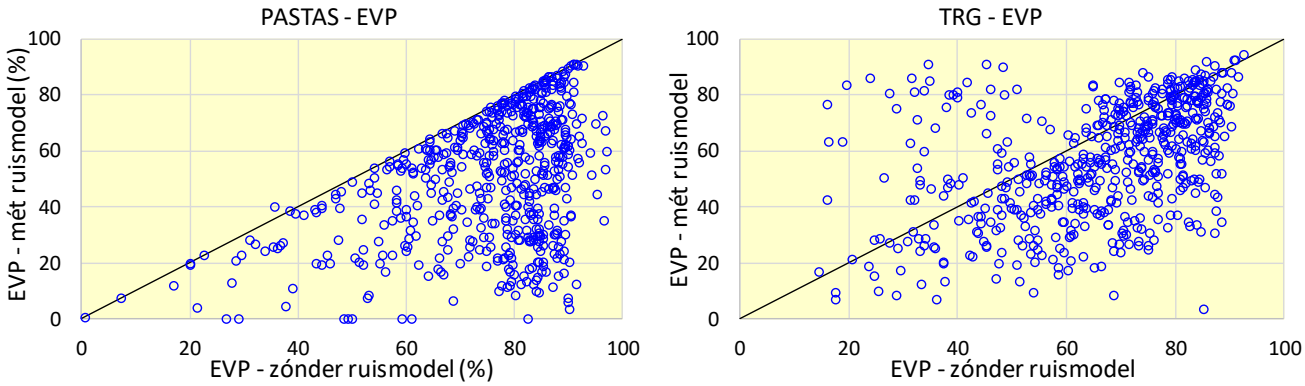
## 3.4 Invloed toepassen ruismodel op verklaarde variantie en evenwichtsrelatie

In figuur 3.2 is voor elk van de 574 met Pastas gemodelleerde reeksen het percentage verklaarde variantie bij modelleren met ruismodel uitgezet tegen dat bij modelleren zonder ruismodel. Ter vergelijking zijn dergelijke modellen ook ontwikkeld met het programma TRG (ook wel aangeduid als *Tijdreeksanalist*), een programma voor discrete tijdreeksmodellering volgens de Box-Jenkins-methode [Van der Meulen en Baggelaar, 2019]. Zie onderstaand tekstkader voor een toelichting op die methode.

Box-Jenkinsmodellen van de grondwaterstandreeksen

Louter ter vergelijking zijn de praktijkreeksen behalve met Pastas, ook gemodelleerd met de Box-Jenkins-methode, met het programma TRG. Deze methode omvat het cyclisch toepassen van drie stappen, namelijk identificeren, schatten en verifiëren en vergt actieve sturing door de modelleur. Normalerwijs dient de cyclus van drie stappen te worden herhaald tot het model voldoet aan de gestelde randvoorwaarden (waaronder die van witte ruis), maar dergelijk maatwerk zou hier een forse tijdsbesteding vergen, gezien het grote aantal reeksen. Daarom is gekozen voor een uniforme modelleeraanpak, waarbij de transferfunctie van grondwaterstand en neerslag is geschat met 365 omega-parameters (zie bijlage 1). Dit biedt TRG maximale vormflexibiliteit om de geschatte transferfunctie aan te laten sluiten op de werkelijke transferfunctie. Verder is het ruismodel eenvoudig geformuleerd, als AR(1)-model.

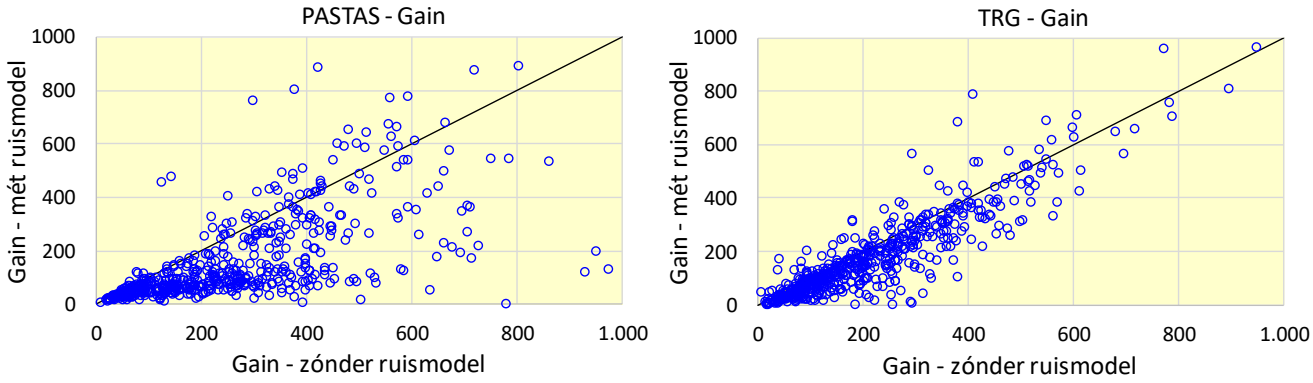
*Figuur 3.2: Spreidingsdiagram waarin voor elk van de 574 gemodelleerde reeksen het percentage verklaarde variantie (EVP) van de grondwaterstand bij modelleren met ruismodel is uitgezet tegen dat bij modelleren zonder ruismodel, zowel voor Pastas (links), als voor TRG (rechts). De doorgetrokken zwarte lijn is de diagonaal.*




Uit figuur 3.2 blijkt dat Pastas de variantie van de grondwaterstand in vrijwel alle gevallen (99,5%) beter beschrijft zónder toepassing van het ruismodel dan mét toepassing van het ruismodel, waarbij de verschillen relevant kunnen zijn. Voor TRG is dat minder vaak, namelijk in 74,4% van de gevallen.

In figuur 3.3 is voor deze 574 reeksen de geschatte evenwichtsrelatie (*gain*) tussen neerslag en grondwaterstand bij modelleren mét ruismodel uitgezet tegen die bij modelleren zónder ruismodel, zowel voor Pastas, als voor TRG.

*Figuur 3.3: Spreidingsdiagram waarin voor elk van de 574 gemodelleerde reeksen de geschatte evenwichtsrelatie (gain) tussen neerslag en grondwaterstand bij modelleren met ruismodel is uitgezet tegen die bij modelleren zonder ruismodel, voor Pastas en voor TRG. De doorgetrokken zwarte lijn is de diagonaal.* *Eenheid van gain ontbreekt.*

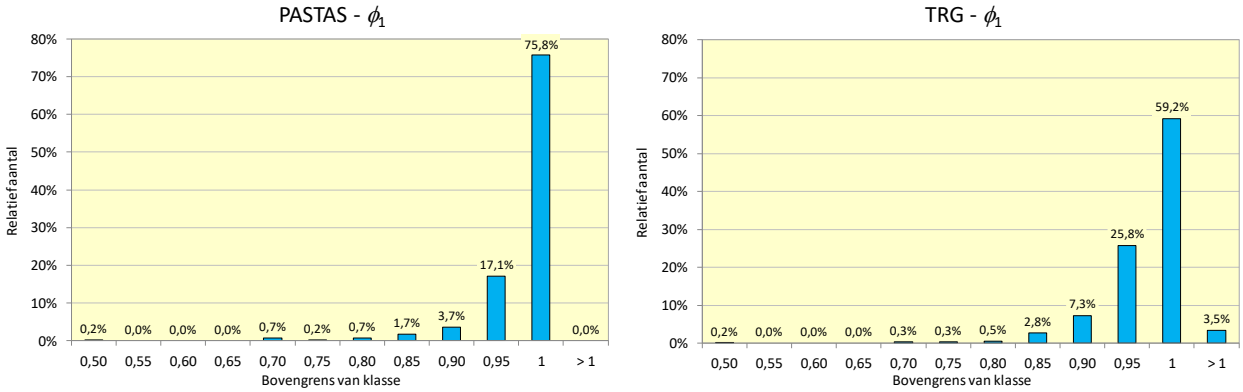



Uit figuur 3.3 blijkt dat Pastas de evenwichtsrelatie in de meeste gevallen (85,7%) groter schat zónder toepassing van het ruismodel dan mét toepassing van het ruismodel. Voor TRG is dat iets minder vaak en treedt het op in 78,6% van de gevallen, maar de verschillen zijn duidelijk kleiner dan bij Pastas. In het volgende hoofdstuk gaan we voor Pastas na wat het weglaten of toevoegen van het ruismodel betekent voor de zuiverheid van de geschatte evenwichtsrelatie.


© AMO en PB Icastat

Figuur 3.4 toont zowel voor Pastas als voor TRG het histogram van de geschatte parameter $\phi_1$ van het gehanteerde AR(1)-ruismodel.

*Figuur 3.4: Histogram van de geschatte parameter van het gehanteerde AR(1)-ruismodel (aangeduid als $\phi_1$), voor Pastas en TRG. De schatting van deze parameter kan met Pastas niet groter dan 1 zijn.*




In veruit de meeste gevallen bleek er sprake van extreem hoge $\phi_1$-waarden van > 0,95, namelijk 75,8% bij Pastas en 62,7% bij TRG (waarvan 3,5% zelfs > 1, duidend op een niet-stationair ruismodel[14]). In de meeste overige gevallen bleek er nog altijd sprake van zeer hoge $\phi_1$-waarden van tussen 0,90 en 0,95. Bij het modelleren van de grondwaterstand op dagbasis moet dus rekening worden gehouden met zeer hoge tot extreme autocorrelatie van de ruis.

## 3.5 Kernpunten van dit hoofdstuk

Uit het modelleren van 574 grondwaterstandreeksen op dagbasis, met als invoer neerslag en verdamping, blijkt dat Pastas met de huidige structuur van het ruismodel nog niet in staat is de ruis- of innovatiereeks volledig terug te brengen tot witte ruis. Er kan sprake zijn van een zeer sterke autocorrelatiestructuur van de ruis, die gepaard gaat met een geringe tot matige moving-average-structuur en soms ook nog correlatie tussen waarden die een jaar uiteen liggen. Het AR(1)-ruismodel van Pastas blijkt in staat de mate van correlatie te verminderen, maar in geen van deze gevallen wordt de innovatiereeks volledig omgezet tot witte ruis. Bovendien blijkt de ruis- of innovatiereeks zelden te voldoen aan normaliteit.

Verder blijkt uit deze exercitie met een groot aantal praktijkreeksen dat Pastas de variantie van de grondwaterstand in vrijwel alle gevallen (99,5%) beter beschrijft zónder toepassing van het ruismodel dan mét toepassing van het ruismodel, waarbij de verschillen relevant kunnen zijn. Blijkbaar kan toepassing van het huidige ruismodel schade aan de pasvorm[15] opleveren.

Aangezien het een omvangrijke en ruimtelijk goed gespreide verzameling van grondwaterstandreeksen betreft uit de Provincie Overijssel, is het aannemelijk dat deze problemen zich vrijwel overal kunnen voordoen. De conclusie is daarom gerechtvaardigd dat Pastas in zijn huidige vorm vaak nog niet in staat zal stellen om bij modelleren van grondwaterstanden op dagbasis uitspraken met gekwantificeerde betrouwbaarheden te doen.

[14] Een ruismodel is niet-stationair als het een reeks beschrijft waarvan statistische eigenschappen (zoals het gemiddelde en/of de standaardafwijking) veranderen met de tijd.

[15] De pasvorm van een model is de mate waarin het model aansluit op de gemodelleerde reeks.



© AMO en PB Icastat 

# 4 Prestaties Pastas bij kunstmatige reeksen

Om de prestaties van Pastas objectiever te kunnen beoordelen, zijn simulatie-exercities uitgevoerd met kunstmatige tijdreeksen. Daarbij zijn grondwaterstandreeksen op dagbasis gesimuleerd en vervolgens gemodelleerd met Pastas, zodat de modelschattingen direct konden worden vergeleken met de werkelijke waarden. Voor deze exercities is versie 0.13.0 van Pastas gebruikt.

## 4.1 Simulatie met de geschaalde Gammaverdeling

Er zijn vier simulatie-exercities uitgevoerd. Bij elk is eerst eenmalig een tijdreeks van de neerslag getransformeerd tot een neerslagcomponent van de grondwaterstand, met een door ons ingestelde blokresponsfunctie, afgeleid uit de geschaalde Gammaverdeling. Vervolgens zijn daar 1.000 grondwaterstandreeksen uit vervaardigd, door 1.000 stochastisch gesimuleerde ruisreeksen aan de component toe te voegen. De verdamping is niet meegenomen als component van de grondwaterstand, louter om de exercities eenvoudig te houden.
De toegevoegde ruis is niet wit, maar gekleurd, in de vorm van een AR(1)-correlatiestructuur. In principe zou Pastas deze eenvoudige correlatiestuctuur zodanig goed moeten kunnen beschrijven dat de innovaties witte ruis vormen.
De voor de simulaties gebruikte reeks van de dagelijkse neerslag is afkomstig van het KNMI en betreft metingen op het weerstation De Bilt.
De gesimuleerde grondwaterstandreeksen lopen van 1 januari 2005 t/m 12 april 2019 en omvatten daarmee 14,3 jaar (5.215 waarden).

Bij elk van de vier exercities is de neerslagcomponent gesimuleerd met behulp van één van de transferfuncties waar Pastas rekening mee kan houden, namelijk de geschaalde Gammaverdeling[16]:

$$f(x) = A\frac{x^{k-1}e^{-x/\theta}}{\theta^k\Gamma(k)}, \text{voor } x \geq 0 \qquad [4.1]$$

waarin *x* de tijd, *k* de vormparameter, *A* en $\theta$ schaalparameters en $\Gamma(.)$ de gammafunctie. Zijn vorm wordt bepaald door de vormparameter *k*. Voor $0 < k < 1$ beschrijft deze een zeer snelle afname (steiler dan een exponentiële afname), voor $k = 1$ een exponentiële afname en voor $k > 1$ een positief scheve verdeling, die geleidelijk minder scheef wordt en uiteindelijk overgaat in een normale kansverdeling. De parameter *A* is op te vatten als de evenwichtsrelatie (gain).

De parameters van [4.1] zijn bij deze vier exercities ingesteld als:

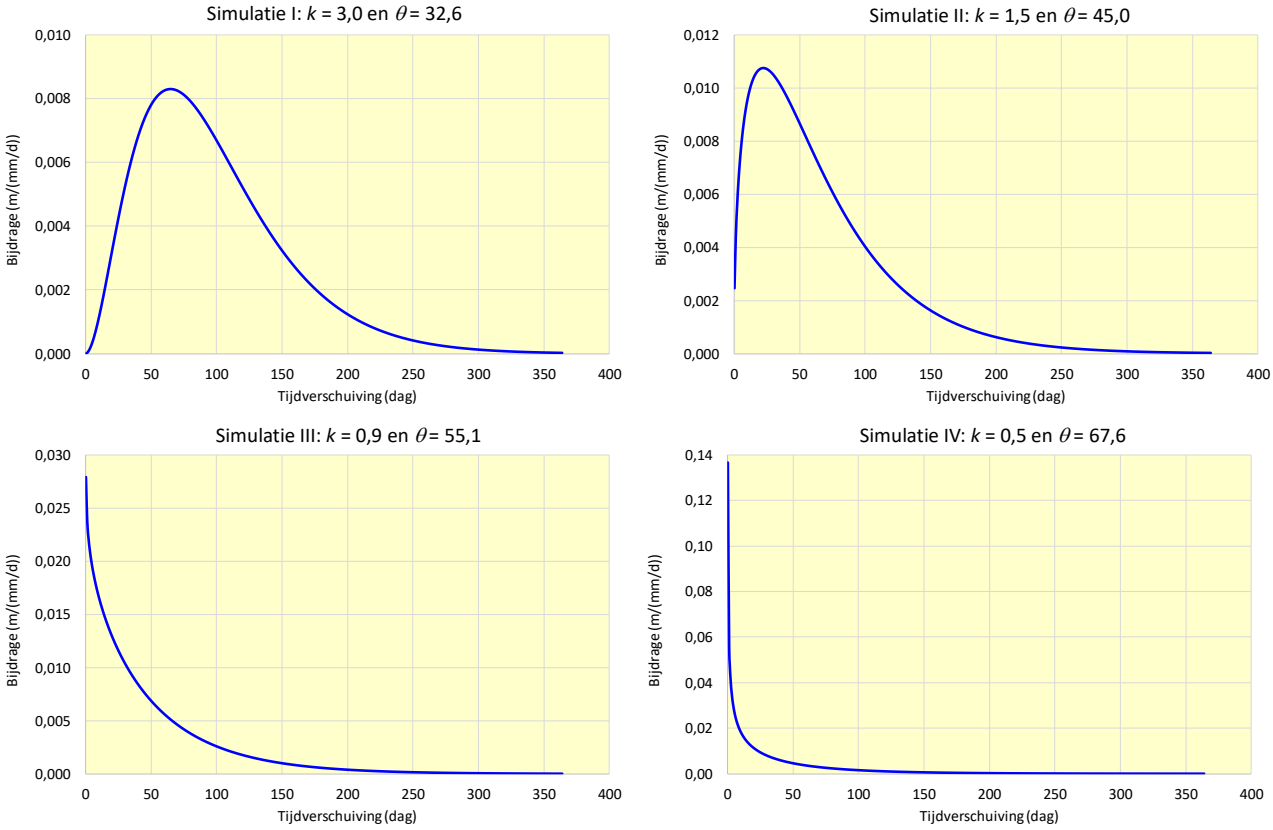
I. $A$ = 1 m/(mm/d), $k$ = 3 en $\theta$ = 32,6;
II. $A$ = 1 m/(mm/d), $k$ = 1,5 en $\theta$ = 45,0;
III. $A$ = 1 m/(mm/d), $k$ = 0,9 en $\theta$ = 55,1;
IV. $A$ = 1 m/(mm/d), $k$ = 0,5 en $\theta$ = 67,6.

[16] Deze is bijvoorbeeld toegepast bij de in § 3.1 besproken Overijsselse praktijkstudie, voor alle geanalyseerde reeksen.


© AMO en PB Icastat

Door de parameter *A* in te stellen op 1 is de evenwichtsrelatie (*gain*) van de grondwaterstand en de neerslag ingesteld op 1 m/(mm/d). De responstijd[17] van de neerslag is ingesteld op 365 dagen. De schaalfactor $\theta$ is afgeleid als de ratio van deze responstijd en de *x*-waarde die 99,9% van de cumulatieve verdeling geeft (de *cutoff*). Elke transferfunctie betreft een blokresponsfunctie die is afgeleid uit de geschaalde Gamma-verdelingsfunctie.[18] Figuur 4.1 toont de resulterende transferfuncties.

*Figuur 4.1: Voor elk van de vier simulaties de transferfunctie waarmee de neerslagreeks is omgezet naar een component van een grondwaterstandreeks. Elk betreft een blokresponsfunctie, zoals afgeleid uit de geschaalde Gamma-verdelingsfunctie.*




Elk van de 1.000 toegevoegde ruisreeksen is met een AR(1)-model (met $\phi_1$ = 0,9) gegenereerd uit een reeks onafhankelijke trekkingen uit een normale kansverdeling met gemiddelde nul, zodanig dat de variantie van de ruis 30% bedraagt van de variantie van de neerslagcomponent. De variantie van de ruis bedraagt daardoor 23,1%% van de variantie van de grondwaterstand.

Van elke gesimuleerde grondwaterstandreeks zijn met Pastas twee tijdreeksmodellen ontwikkeld, namelijk één mét het AR(1)-ruismodel en één zónder ruismodel.

[17] De responstijd van een invloedsfactor is de tijd die het duurt voordat het effect van een kortstondige verandering van die factor op de grondwaterstandstand is uitgewerkt.

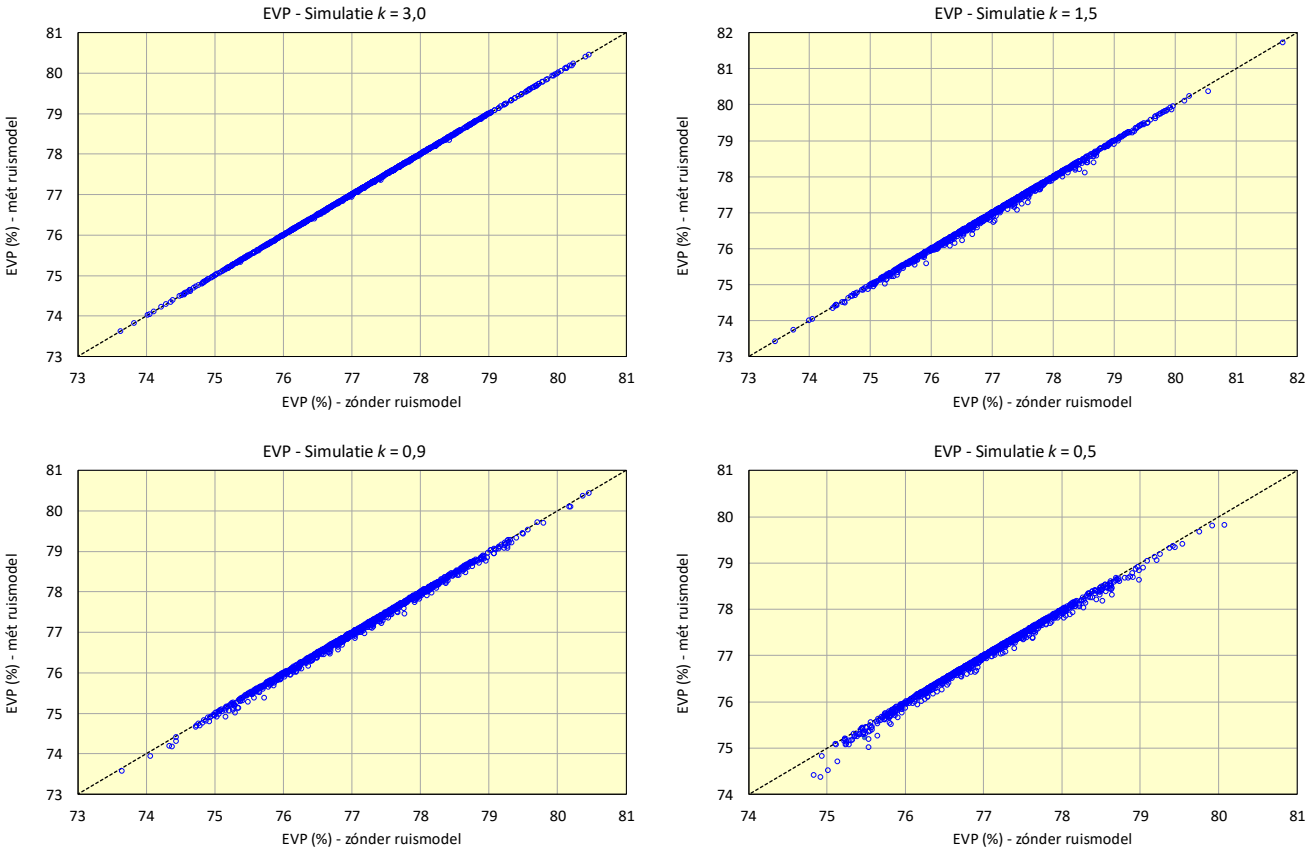
[18] Deze simulatie-aanpak met de blokresponsfunctie is bijvoorbeeld beschreven in: Stijn Klop and Mark Bakker, *Pastas Noise model*, 28 Aug 2019: https://github.com/pastas/pastas/blob/master/examples/notebooks/8_pastas_synthetic.ipynb. De beschrijving omvat ook een Python-script.


© AMO en PB Icastat

## Resultaten simulatie-exercities

De resultaten van de vier exercities zijn weergegeven in figuur 4.2 (verklaarde variantie), figuur 4.3 en figuur 4.4 (geschatte evenwichtsrelatie) en ten slotte samengevat in tabel 4.1.

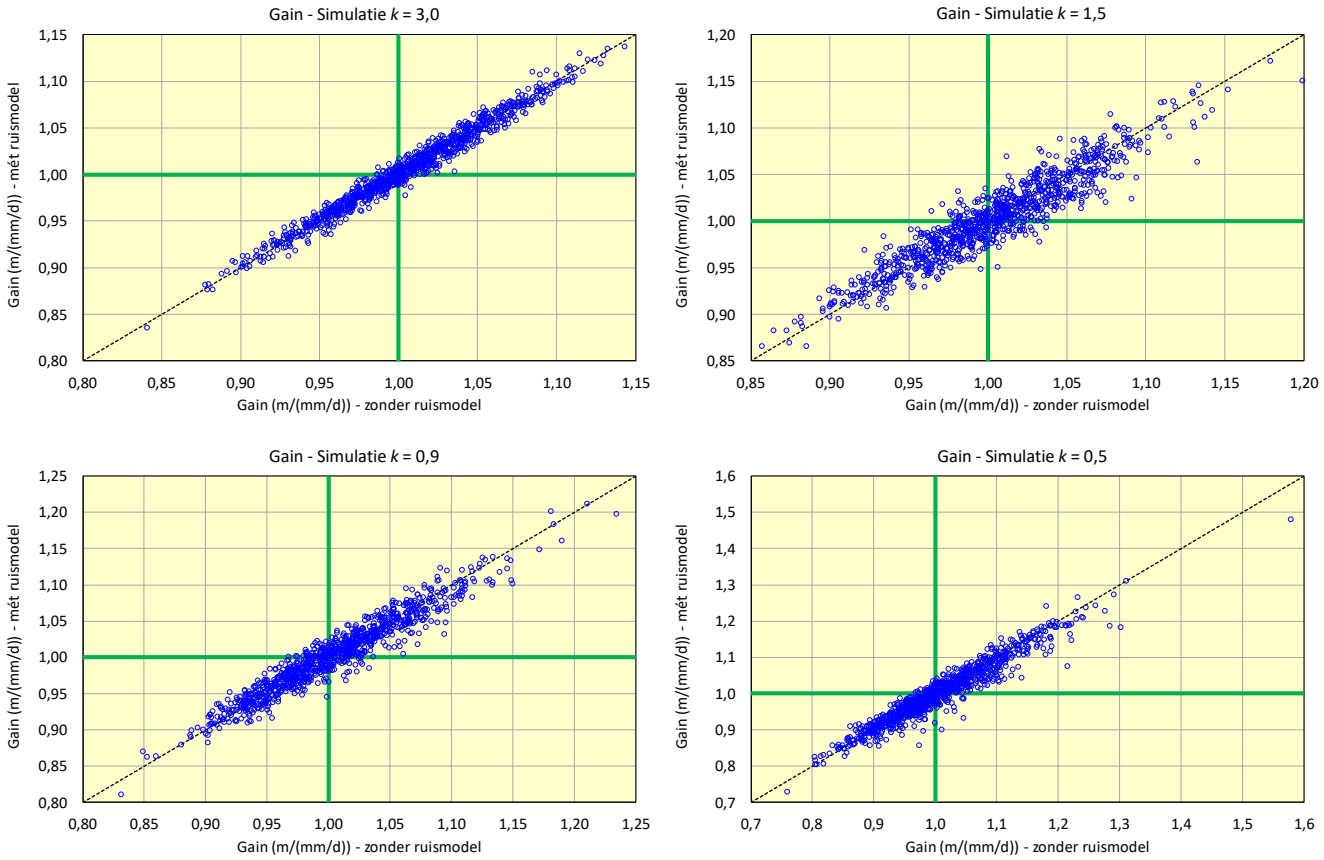
*Figuur 4.2: Resultaten van de vier simulatie-exercities. Verklaarde variantie (EVP) bij modelleren mét ruismodel uitgezet tegen die bij modelleren zónder ruismodel. De gestippelde zwarte lijn is de diagonaal.*




Uit figuur 4.2 blijkt dat bij deze exercities de verklaarde variantie van een grondwaterstandreeks bij modelleren mét ruismodel niet veel afwijkt van die bij modelleren zónder ruismodel. Maar bij modelleren zonder ruismodel wordt de variantie meestal wel enigszins beter verklaard dan bij modellen mét ruismodel (in 95 à 99% van de gevallen, zie daarvoor ook tabel 4.1), waarbij de grootte van dit verschil toeneemt naarmate de vormparameter *k* afneemt. Blijkbaar bewerkstelligt het ruismodel dat het deterministische deel van het tijdreeksmodel de grondwaterstand minder goed beschrijft. Dit verschijnsel zagen we ook al optreden bij de tijdreeksanalyses van de 574 praktijkreeksen van de Provincie Overijssel (zie § 3.4).


© AMO en PB Icastat

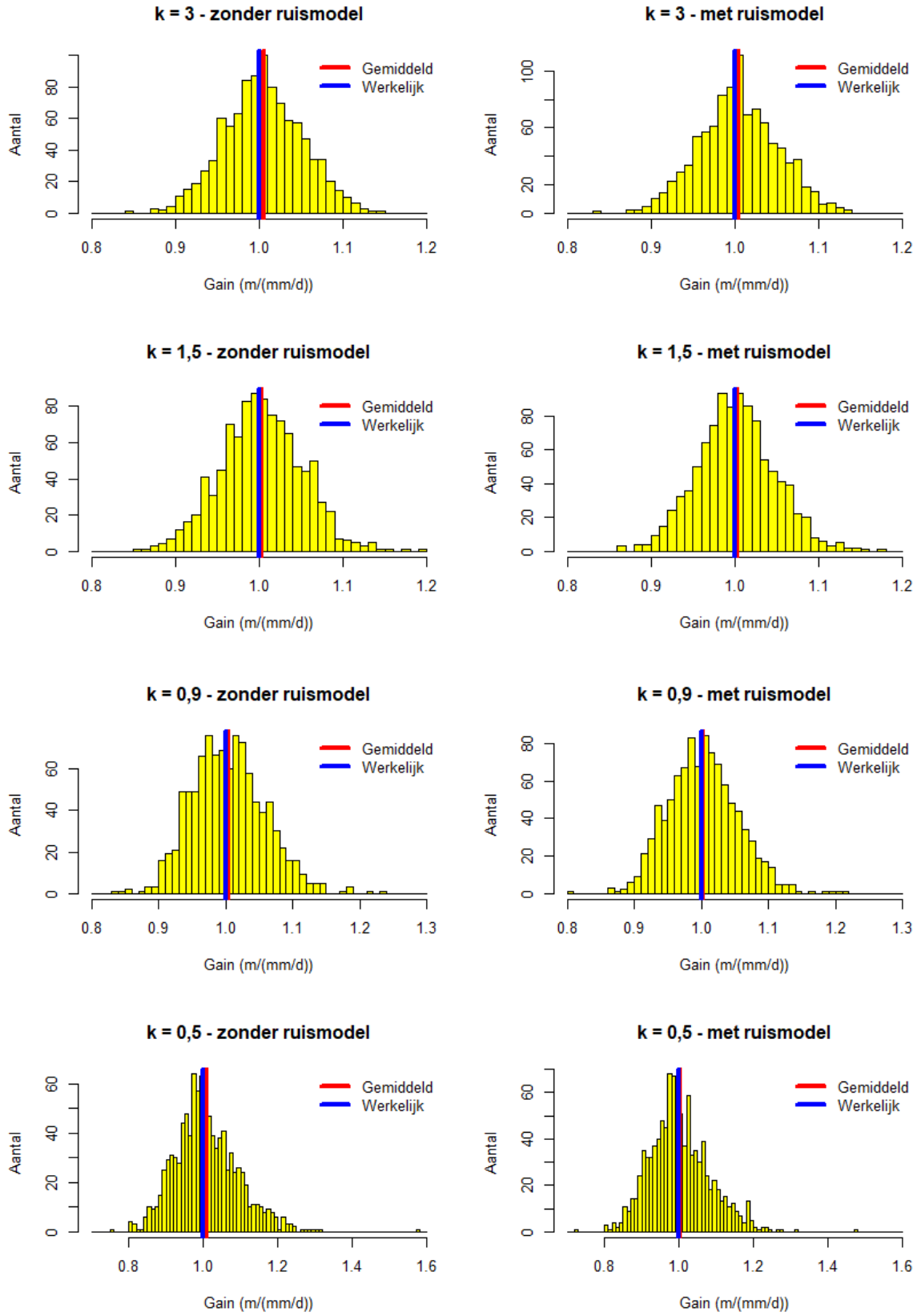
*Figuur 4.3: Resultaten van de vier simulatie-exercities. Geschatte evenwichtsrelatie (gain) bij modelleren mét ruismodel uitgezet tegen die bij modelleren zónder ruismodel. De groene lijnen vertegenwoordigen de werkelijke gain (1 m/(mm/d)). De gestippelde zwarte lijn is de diagonaal.*




Uit bovenstaande figuur 4.3 blijkt dat bij deze simulatie-exercities de geschatte evenwichtsrelaties bij modelleren mét en zónder ruismodel sterk gecorreleerd zijn. Tevens blijken voor beide modelsoorten de geschatte evenwichtsrelaties min of meer symmetrisch ten opzichte van de werkelijke waarde te liggen. Dit laatste blijkt nog duidelijker uit de onderstaande histogrammen van figuur 4.4. Tevens blijkt uit figuur 4.4 dat voor beide modelsoorten het gemiddelde van de 1.000 schattingen zich op of nabij de werkelijke waarde bevindt.


© AMO en PB Icastat

*Figuur 4.4: Resultaten van de vier simulatie-exercities. Histogram van de geschatte evenwichtsrelatie (gain) bij modelleren zónder ruismodel (links) en bij modelleren mét ruismodel (rechts).*






© AMO en PB Icastat 

*Tabel 4.1: Samengevatte resultaten van de vier simulatie-exercities.*

| Kenmerk | *k* = 3,0 Ruismodel | | *k* = 1,5 Ruismodel | | *k* = 0,9 Ruismodel | | *k* = 0,5 Ruismodel | |
|---|---|---|---|---|---|---|---|---|
| | Niet | Wel | Niet | Wel | Niet | Wel | Niet | Wel |
| **Werkelijke gain (m/(mm/d))** | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| **Gemiddelde gain (m/(mm/d))** | 1,005 | 1,005 | 1,003 | 1,002 | 1,005 | 1,003 | 1,009 | 1,002 |
| **Onzuiverheid (%)** | 0,5% | 0,5% | 0,3% | 0,2% | 0,5% | 0,3% | 0,9% | 0,2% |
| **p-waarde T-toets op zuiverheid gain** [1)] | **0,00** | **0,00** | **0,03** | **0,14** | **0,01** | **0,07** | **0,00** | **0,38** |
| **Dekkingsgraad 95%-betrouwb.interval** [2)] | **37,2%** | **94,4%** | **37,1%** | **93,4%** | **35,7%** | **96,1%** | **37,5%** | **95,0%** |
| **Gemiddelde EVP (%)** | 77,18% | 77,17% | 77,08% | 77,04% | 77,08% | 77,03% | 77,03% | 76,98% |
| **EVP_Niet > EVP_Wel (%)** | 95,3% | | 98,8% | | 99,8% | | 99,7% | |

1) Nulhypothese: gain wordt zuiver geschat.
2) Nulhypothese: empirische dekkingsgraad is gelijk aan nominale dekkingsgraad (95%).
Bij beide toetsen is 95% betrouwbaarheid gehanteerd en tweezijdig getoetst.

Nulhypothese niet verworpen
Nulhypothese verworpen

Uit tabel 4.1 blijkt dat bij elk van deze simulatie-exercities de evenwichtsrelatie (*gain*) goed is geschat, waarbij de onzuiverheid beperkt blijft (tussen 0,2% en 0,9%). Bij drie van de modelleringen zónder ruismodel was er weliswaar sprake van een statistisch significante onzuiverheid, maar de mate van onzuiverheid is in deze gevallen praktisch niet relevant te achten.
Bij drie van de modelleringen mét ruismodel blijkt de empirische dekkingsgraad van het 95%-betrouwbaarheidsinterval statistisch niet significant afwijkend van 95%, de nominale dekkingsgraad. Bij de vierde is er weliswaar wél sprake van een statistisch significante afwijking, maar die afwijking is praktisch niet relevant te achten (de empirische dekkingsgraad bedraagt 93,4%).
Bij elk van de modelleringen zonder toepassing van het ruismodel bedraagt de empirische dekkingsgraad van het 95%-betrouwbaarheidsinterval van de evenwichtsrelatie slechts 35 à 37%, in plaats van de ingestelde 95%.[19] Dit duidt op een sterke onderschatting van de standaardfout van de evenwichtsrelatie, met alle risico's van dien op het onterecht detecteren van een statistisch significante relatie. Als er geen sprake is van een relatie kan er desondanks in circa twee-derde van de gevallen een statistisch significante relatie worden gedetecteerd, terwijl de analist er vanuit zal gaan dat het risico daarop slechts zeer beperkt is (het nominale significantieniveau, doorgaans ingesteld op 5%). Deze ongewenste risico's zijn bij modelleren op dagbasis zonder ruismodel ook vrijwel onvermijdelijk, omdat de restterm altijd sterke autocorrelatie zal vertonen, waardoor niet wordt voldaan aan de randvoorwaarde dat deze witte ruis vormt.
De verklaarde variantie (EVP) blijkt bij modelleren zónder ruismodel bijna altijd groter dan bij modelleren mét ruismodel. De verschillen zijn echter niet groot, al nemen ze wel toe naarmate de vormparameter *k* afneemt (de transferfunctie wordt dan steiler).

[19] Dit sluit aan op de bevindingen van hoofdstuk 2 over de invloed van het niet voldoen aan witte ruis op de betrouwbaarheid van uitspraken. Uit de simulatieresultaten met lineaire regressie, vermeld in tabel 2.1, blijkt dat bij een AR(1)-correlatiestructuur en $\phi_1$ = 0,9 de empirische dekkingsgraad van het 95%-betrouwbaarheidsinterval van de geschatte relatie 36% bedraagt.


© AMO en PB Icastat

## 4.2 Simulatie met andersoortige verdeling

Uit deze simulatie-exercities blijkt dat Pastas met de Gamma-transferfunctie goed in staat is een op basis van diezelfde functie kunstmatig aangebrachte evenwichtsrelatie terug te schatten, ook als gekleurde ruis is toegevoegd. Maar in hoeverre lukt dat ook nog als er sprake is van een transferfunctie met een vorm die wel lijkt op die van een Gammaverdeling, maar er niet gelijk aan is?
Met een verkennende exercitie is dit nagegaan. Daarbij zijn simulaties uitgevoerd waarbij de neerslagcomponent is gesimuleerd met een transferfunctie gebaseerd op verschillende uitvoeringen van een links afgeknotte normale verdeling. De opzet en resultaten van deze exercitie zijn beschreven in bijlage 2.
Uit die exercitie blijkt dat er voor vrijwel alle uitvoeringen sprake is van het onvoldoende reproduceren van de gesimuleerde relatievorm, wat voor elk van deze ook tot uiting komt in een gemiddelde van de geschatte evenwichtsrelaties dat statistisch significant afwijkt van de ingestelde waarde. De relatieve onzuiverheid kan daarbij oplopen tot 37% van de werkelijke evenwichtsrelatie. In vrijwel alle gevallen schat Pastas zónder toepassing van een ruismodel de evenwichtsrelatie beter dan mét toepassing van een ruismodel. Blijkbaar kan de toepassing van het huidige ruismodel schade opleveren.
De geconstateerde onzuiverheid is uiteraard niet onverwacht, omdat voor het simuleren een andere functie is gebruikt dan waar bij het schatten van het model rekening mee wordt gehouden. Het moet dan ook niet worden gezien als een onvolkomenheid van Pastas, maar van het modelleerproces. Het wijst er op hoe belangrijk het is om bij het schatten van het model zo goed mogelijk rekening te kunnen houden met de werkelijke vorm van de relatie. De volgende mogelijkheden verdienen in dat kader nadere uitwerking: i) uitbreiden van de verzameling voorgedefinieerde transferfuncties, in combinatie met het toevoegen van een identificatiestap aan het modelleerproces en ii) ontwikkelen van flexibelere transferfuncties (zie ook hoofdstuk 5).

## 4.3 Kernpunten van dit hoofdstuk

Uit simulatie-exercities blijkt dat Pastas met de Gamma-transferfunctie goed in staat is een op basis van diezelfde functie kunstmatig aangebrachte evenwichtsrelatie terug te schatten, ook als gekleurde ruis is toegevoegd. Verder blijkt bij modelleren mét ruismodel de empirische dekkingsgraad van het 95%-betrouwbaarheidsinterval van de geschatte evenwichtsrelatie min of meer gelijk aan de nominale dekkingsgraad. Dit laatste zal komen doordat de correlatiestructuur van de bij onze simulaties aangebrachte ruis – AR(1) – aansluit op wat het huidige ruismodel van Pastas aankan. Maar in de praktijk zal er bij modelleren van de grondwaterstand op dagbasis veelal sprake zijn van een uitgebreidere correlatiestructuur. Het ligt voor de hand dat Pastas dan niet in staat zal zijn de innovaties te laten voldoen aan de randvoorwaarde van witte ruis, met als gevolg dat de risico's op verkeerde conclusies groter zullen zijn dan geaccepteerd. Uit de simulatie-exercities blijkt dat bij modelleren zónder ruismodel de empirische dekkingsgraad van het 95%-betrouwbaarheidsinterval van de geschatte evenwichtsrelatie laag is (35 à 37%), wat een groot risico geeft op het onterecht detecteren van een statistisch significante relatie.
Als we relatievormen aanbrengen die enigszins afwijken van die van de Gamma-transferfunctie, terwijl Pastas bij het modelschatten wel van die functie uitgaat, kan de relatieve onzuiverheid van de geschatte evenwichtsrelatie flink oplopen. Dit resultaat is uiteraard niet onverwacht, omdat voor het simuleren een andere functie is gebruikt dan waar bij het schatten van het model rekening mee wordt gehouden. Het moet dan ook niet worden gezien als een onvolkomenheid van Pastas, maar van het modelleerproces. Het wijst er op hoe belangrijk het is om bij het schatten van het

model zo goed mogelijk rekening te kunnen houden met de werkelijke vorm van de relatie. De volgende mogelijkheden verdienen in dat kader nadere uitwerking: i) uitbreiden van de verzameling voorgedefinieerde transferfuncties, in combinatie met het toevoegen van een identificatiestap aan het modelleerproces en ii) ontwikkelen van flexibelere transferfuncties (zie ook hoofdstuk 5).


© AMO en PB Icastat 

# 5 Verbetermogelijkheden Pastas

Gebaseerd op de bevindingen van de voorgaande hoofdstukken, bepreekt dit hoofdstuk de volgende mogelijkheden om de werking van Pastas te verbeteren: i) identificeren transferfuncties, ii) flexibiliseren transferfunctie, iii) flexibiliseren ruismodel / verlagen modelfrequentie, iv) optimaliseren schattingsroutine, v) verifiëren tijdreeksmodel. Dit hoofdstuk sluit af met het voorstel voor een werkprocedure om met Pastas te modelleren (neergelegd in een stroomschema), die gebruik maakt van de beschouwde mogelijkheden.

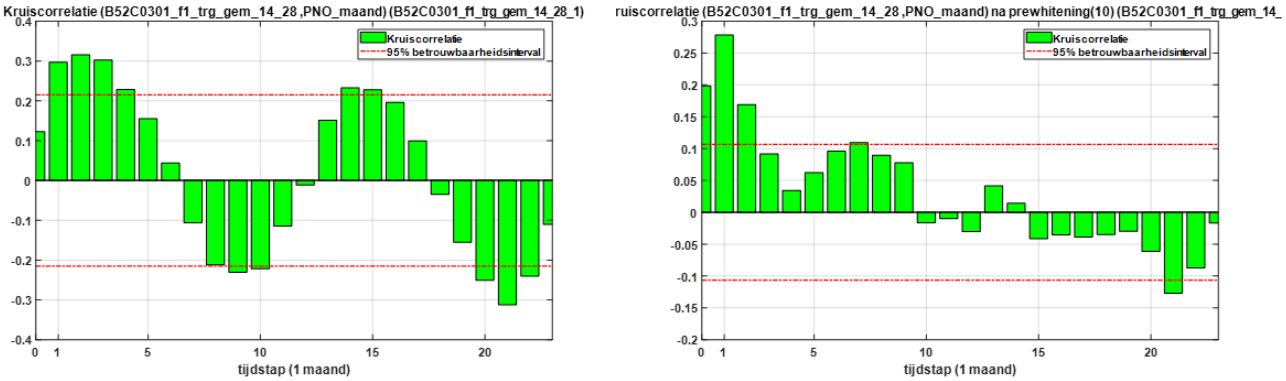
## 5.1 Identificeren transferfuncties

Er zijn twee redenen om voorafgaand aan het modelleren per invoerreeks de vorm van de transferfunctie empirisch te identificeren, namelijk:

1. Er kan dan een transferfunctie voor de modelschatting worden geselecteerd die goed past bij de empirisch geïdentificeerde vorm.
2. Er kan dan na het modelleren worden geverifieerd of de met het model geschatte vorm van de transferfunctie overeenstemt met de empirisch geïdentificeerde vorm. De zuiverheid van een geschatte evenwichtsrelatie wordt namelijk bepaald door de mate van overeenstemming tussen de geschatte vorm van de transferfunctie en de werkelijke transferfunctie.

Het empirisch identificeren van de vorm van een transferfunctie is mogelijk door visuele beoordeling van de kruiscorrelatiefunctie van uitvoerreeks en invoerreeks (zie figuur 5.1). Deze toont het verloop van de kruiscorrelatiecoëfficiënt, een maat voor de lineaire samenhang tussen de waarde van de invoerreeks en de uitvoerreeks *k* tijdseenheden later, als functie van *k* (de tijdsverschuiving, of *lag*). Omdat het beeld van de relatie vertroebeld zal zijn als de invoerreeks autocorrelatie vertoont, worden zowel invoerreeks als uitvoerreeks vóór het berekenen van de kruiscorrelatiefunctie door hetzelfde filter gehaald, namelijk het inverse ARIMA-model voor de invoerreeks (zie bijlage 1). Deze aanpak noemt men ook wel het *witten* (Engels: *prewhitening* [Box and Jenkins, 1976]).
Als er sprake is van meerdere invoerreeksen vergt het beoordelen overigens wel enige ervaring.

*Figuur 5.1: Kruiscorrelatiefunctie van een freatische grondwaterstand en het potentieel neerslagoverschot op maandbasis, zowel vóór (links) als ná het witten (rechts).*




De gewitte kruiscorrelatiefunctie van figuur 5.1-rechts suggereert een responstijd op het potentieel neerslagoverschot van circa 9 maanden en een dynamische relatie gekenmerkt door een snelle toename, gevolgd door een geleidelijke afname. Als het een modellering in het discrete domein


© AMO en PB Icastat

zou betreffen komen hiervoor de volgende modelformuleringen in aanmerking (in Box-Jenkins-notatie, zie bijlage 1 voor een toelichting):

$$Z_t = \omega_0 X_t - \omega_1 X_{t-1} - \ \ldots - \omega_9 X_{t-9} + N_t \qquad [5.1]$$

$$Z_t = \delta_1 Z_{t-1} + \omega_0 X_t - \omega_1 X_{t-1} + N_t \qquad [5.2]$$

waarin *Z* de grondwaterstand, *X* het potentieel neerslagoverschot, $\omega$ een moving-average-parameter, $\delta$ een autoregressieve parameter, *N* de ruis en *t* de tijdsindex (hier van de maand).
De in formule [5.1] beschreven transferfunctie bevat tien moving-average-parameters ($\omega_0$, $\omega_1$, …, $\omega_9$) en biedt daarmee maximale vormflexibiliteit. De in formule [5.2] beschreven transferfunctie bevat twee moving-average-parameters ($\omega_0$ en $\omega_1$), om de responsen in maand 0 en maand 1 te beschrijven en één autoregressieve parameter ($\delta_1$), om de afname na maand 2 te beschrijven. Deze formulering vergt minder parameters dan [5.1], maar biedt minder vormflexibiliteit.

## 5.2 Flexibiliseren transferfunctie

De zuiverheid van een geschatte evenwichtsrelatie wordt sterk bepaald door de mate van overeenstemming tussen de geformuleerde en de werkelijke vorm van de relatie. Mogelijkheden om de capaciteiten van Pastas op dat vlak te vergroten zijn de volgende.

1. Uitbreiden van de verzameling voorgedefinieerde transferfuncties – Deze moet zodanig zijn dat voor een voldoende breed scala aan hydrologische situaties dynamische relaties met invloedsfactoren afdoende kunnen worden beschreven. Maar dit dient dan wel samen te gaan met het toevoegen van een identificatiestap aan het modelleerproces, waarin op basis van de empirische kruiscorrelatiefunctie van invoerreeks en uitvoerreeks de best daarop aansluitende transferfunctie voor de modelschatting wordt geselecteerd (zie ook § 5.1).

2. Hanteren flexibelere transferfuncties. Daarvoor zien wij de volgende opties.
   - Gebruik van polynomen. De PIRFICT-methode gaat – zoals de volledige naam *Predefined Impulse Response Functions In Continuous Time* aangeeft – uit van voorgedefinieerde transferfuncties die hydrologische processen kunnen beschrijven. Als we de restrictie laten vallen dat elke transferfunctie een hydrologische basis moet hebben, komt ook het gebruik van polynomen in aanmerking om transferfuncties te beschrijven en kan de vormflexibiliteit van de transferfunctie worden opgeschroefd door de orde van de polynoom te vergroten. Met bijvoorbeeld één invoerreeks en een polynome transferfunctie heeft het tijdreeksmodel in het continue tijdsdomein de volgende vorm:

     $$Z_t = \int_{u=0}^{rt} \Big( X(t-u) \cdot \big( \beta_0 + \beta_1 (u) + \beta_2 (u)^2 + \ \ldots + \beta_p (u)^p \big) \Big) \, \mathrm{d}u + N_t \qquad [5.3]$$

     waarin *Z* de grondwaterstand, *rt* de responstijd, *X* de invoerreeks, $\beta_i$ *(i = 1…p)* de parameters van de polynoom van orde *p*, *N* de ruis en *t* de tijd. Het afknotten van de transferfunctie bij de responstijd (in plaats van doorlopen tot oneindig) kan hier nodig zijn om negatieve waarden te voorkomen.
   - Gebruik van het soort transferfuncties die worden toegepast bij Box-Jenkins-tijdreeksmodellering. Deze zijn afgeleid van de algemene weergave van een discreet lineair dynamisch systeem met een lineaire differentievergelijking. Met slechts een beperkt aantal parameters kunnen daarmee veel soorten relaties worden beschreven, die bij tijdreeksmodelleringen in allerlei vakgebieden zijn toegepast. Het betreft dan meestal combinaties van moving-


© AMO en PB Icastat 

average-parameters ($\omega$) en een autoregressieve parameter ($\delta$), zoals in het voorbeeld van formule 5.2 hierboven.

## 5.3 Flexibiliseren ruismodel / verlagen modelfrequentie

Een belangrijke verbetering van Pastas kan worden bewerkstelligd als het ruismodel zodanig kan worden uitgebreid dat het flexibiliteit biedt om in de meeste hydrologische situaties tot een tijdreeksmodel te kunnen komen dat voldoet aan de daarvoor geldende randvoorwaarden, ook bij modelleren van een hoogfrequente tijdreeks. Het ruismodel dient daartoe een functionaliteit zoals die van het ARIMA-model te krijgen (zie bijlage 1). Het is echter de vraag of het wel mogelijk is om bij modelleren in het continue domein verder te komen dan het momenteel in Pastas geïmplementeerde AR(1)-ruismodel, vanwege de wiskundige complexiteit. We vonden in de literatuur soms wel wensen op dat gebied, maar geen uitwerkingen.
Een aanzienlijke flexibilisering van het ruismodel is in ieder geval mogelijk in het discrete tijdsdomein, door toepassing van de Box-Jenkins-methode [Box and Jenkins, 1976], waardoor de functionaliteit van het ARIMA-model beschikbaar komt. Daarom verdient het aanbeveling na te gaan of het mogelijk is Pastas uit te breiden tot een hybride model, met de combinatie van transferfuncties die betrekking hebben op het continue tijdsdomein en een ruismodel dat betrekking heeft op het discrete tijdsdomein.

Als de voorkeur wordt gegeven aan alleen modelleren in het continue domein, zijn de opties de tijdreeks uit te dunnen tot een grotere tijdseenheid, of deze te aggregeren tot een grotere tijdseenheid, aangezien dan vaak kan worden volstaan met een minder uitgebreid ruismodel. Dit kan dan echter wel ten koste gaan van het onderscheidend vermogen.

## 5.4 Optimaliseren schattingsroutine

Uit dit onderzoek blijkt dat het ruismodel van Pastas in bepaalde gevallen de prestaties van het deterministische deel van het tijdreeksmodel schaadt. De oorzaak kan liggen in de schattingsroutine, waarbij eerst de ruis wordt bepaald en daaruit vervolgens de innovaties. Dit kan leiden tot sub-optimale schattingen. Het verdient aanbeveling na te gaan hoe dit probleem kan worden opgelost.

## 5.5 Verifiëren tijdreeksmodel

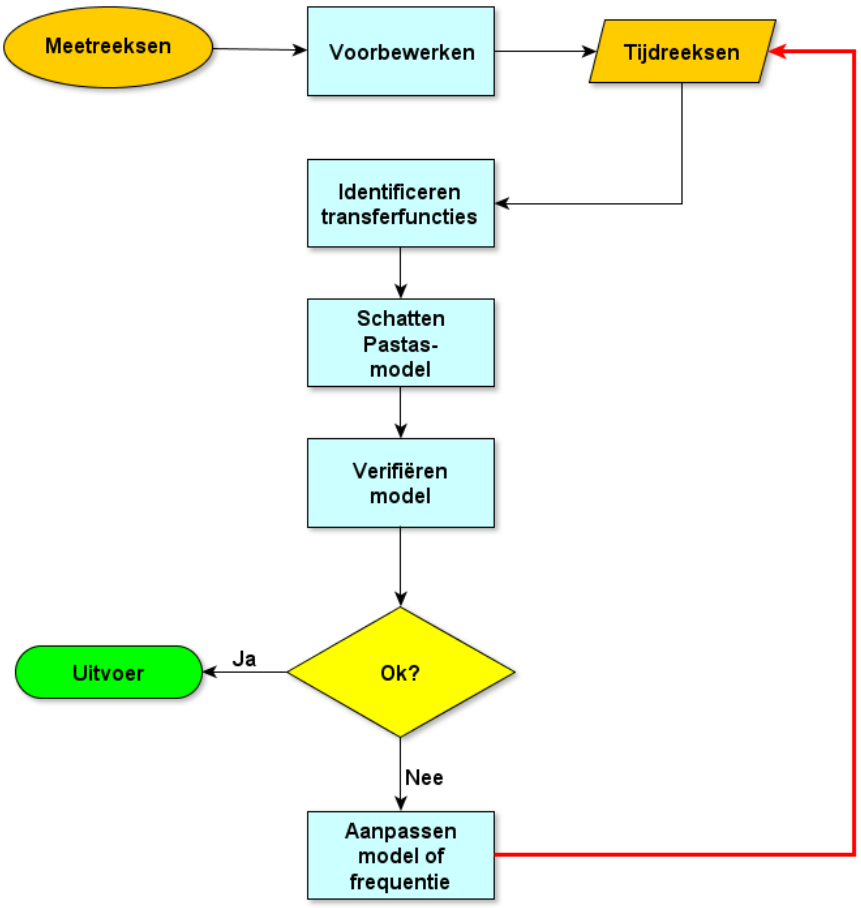
Als de tijdreeksmodellering dient om tot uitspraken met gekwantificeerde betrouwbaarheden te komen, zullen die pas zeggingskracht kunnen krijgen als is aangetoond dat het model voldoet aan de daarvoor vereiste randvoorwaarden (innovaties normaal verdeeld, onafhankelijk van elkaar en zonder relatie met de afzonderlijke invoerreeksen). Er zijn verschillende diagnostieken beschikbaar om dat aan te tonen, zoals ACF, PACF en histogram van de innovaties en verschillende toetsen op autocorrelatie, normaliteit en homoscedasticiteit. Dit is verder uitgewerkt in bijlage 3.
Als het verifiëren leidt tot een onbevredigende diagnose, kan in veel gevallen een beter model ontwikkeld worden aan de hand van de geconstateerde discrepanties (zie het stroomschema van figuur 5.2).

## 5.6 Voorstel werkprocedure modelleren met Pastas

Op basis van de hiervoor besproken mogelijkheden om de werking van Pastas te verbeteren komen we met een voorstel voor een werkprocedure om met Pastas te modelleren. Deze is weergegeven in het stroomschema van figuur 5.2.

De procedure komt neer op een stapsgewijze, interactieve modellering, die zonodig wordt herhaald tot het model voldoet aan de daaraan gestelde randvoorwaarden. Deze kan vermoedelijk niet geheel automatisch worden uitgevoerd, aangezien er bij diverse onderdelen interactie nodig is van de modelleur. Het kan dus meer inzet vergen dan tot dusverre gebruikelijk, maar het zal de kwaliteit en zeggingskracht van de modellen verbeteren.

*Figuur 5.2: Stroomschema van de voorgestelde werkprocedure om met Pastas te modelleren.*




De verschillende elementen van de werkprocedure zijn aangegeven met de volgende vormen en kleuren: oranje ellips: invoer | lichtblauwe rechthoek: werkzaamheid | oranje parallellogram: bewerkte invoer | gele ruit: beslismoment | groene ellips: uitvoer


© AMO en PB Icastat

# 6 Aanbevelingen

1. Ga na of het mogelijk is Pastas uit te breiden, zodanig dat hybride tijdreeksmodellen kunnen worden ontwikkeld, met de combinatie van transferfuncties die betrekking hebben op het continue tijdsdomein (zoals reeds geïmplementeerd in Pastas) en een ruismodel dat betrekking heeft op het discrete tijdsdomein. Dan komt namelijk de functionaliteit van het ARIMA-model beschikbaar, wat flexibiliteit zal bieden om in de meeste hydrologische situaties tot een tijdreeksmodel te kunnen komen dat voldoet aan de daarvoor geldende randvoorwaarden.

2. Gebruik bij modelleren met Pastas in het continue tijdsdomein de stappen van de modelleercyclus (identificeren, schatten en verifiëren), net zoals bij modelleren in het discrete tijdsdomein volgens de Box-Jenkins-methode gebruikelijk is. Er is dan meer inzet van de modelleur nodig dan tot dusverre gebruikelijk was, maar het zal de kwaliteit en zeggingskracht van de modellen verbeteren.

3. Stel in de identificatiefase per invoerreeks empirisch vast welke vorm de transferfunctie heeft en kies vervolgens de daarbij passende functie. Ga daarbij af op de gewitte kruiscorrelatiefunctie van uitvoerreeks en invoerreeks. Als er sprake is van meerdere invoerreeksen vergt dit overigens wel enige ervaring.

4. Stel in de verificatiefase vast of het model voldoet aan de randvoorwaarden om tot uitspraken met gekwantificeerde betrouwbaarheden te kunnen komen (innovaties normaal verdeeld, onafhankelijk van elkaar en zonder relatie met de afzonderlijke invoerreeksen). Er zijn verschillende diagnostieken beschikbaar om dat aan te tonen, zoals ACF, PACF en histogram van de innovaties en verschillende toetsen op autocorrelatie, normaliteit en homoscedasticiteit.

5. Ga na of gebruik kan worden gemaakt van flexibelere transferfuncties. De zuiverheid van een geschatte evenwichtsrelatie wordt namelijk bepaald door de mate van overeenstemming tussen de geschatte vorm van een relatie en de werkelijke vorm.

6. Zoek een oplossing voor het feit dat het huidige ruismodel van Pastas in bepaalde gevallen de prestaties van het tijdreeksmodel schaadt. De oorzaak kan liggen in de schattingsroutine, waarbij eerst de ruis wordt bepaald en daaruit vervolgens de innovaties, wat kan leiden tot suboptimale schattingen.

7. Voorzie Pastas van functionaliteit die de gebruiker in staat stelt om systematisch een aantal stappen te doorlopen waarmee statistisch verantwoorde tijdreeksmodellering kan worden uitgevoerd. En voeg daartoe ook Notebooks met uitgewerkte voorbeelden toe aan de documentatie van Pastas.

Aanbevelingen over verbeteren van het ruismodel en uitbreiden van de verzameling transferfuncties zijn al eerder gedaan [Collenteur, 2018]. Blijkbaar bestaat er dus breder behoefte aan verbeteringen. Maar voor zover wij weten is daar tot dusverre nog niet veel aan gedaan. Hopelijk kan ons rapport daar meer voedingsbodem aan geven.

# Aangehaalde literatuur


- Beekman, W. en Caljé, R. (2018): *Doorwerking droogte van 2018 in Overijssel. Voorspelling per eind september*. Artesia Water Research Unlimited, 2 oktober 2018, Schoonhoven, 65 blz.

- Box, G.E.P. and Jenkins, J.M. (1976): *Time series analysis, forecasting and control*. Holden-Day, San Francisco.

- Collenteur, R.A. (2018): *Over autocorrelatie van tijdreeksmodellen met niet-equidistante tijd-stappen*. Artesia Water Research Unlimited, 10 november 2018, Schoonhoven, 21 blz.

- Collenteur, R.A., Bakker, M., Caljé, R., Klop, S.A., Schaars, F. (2019) *Pastas: open source software for the analysis of groundwater time series*. Groundwater. doi: 10.1111/gwat.12925.

- D'Agostino, R.B. (1986): *Tests for Normal Distribution*. In *Goodness-Of-Fit Techniques*, edited by R.B. D'Agostino and M.A. Stephens, Marcel Dekker, 1986.

- Ljung, G.M and Box, G.E.P (1978): *On a measure of lack of fit in time series models*. Biometrika, 65, blz. 297 – 303.

- McLeod, G. (1983): *Box Jenkins in Practice*. Time Series Library, Gwilym Jenkins & Partners Ltd., Lancaster.

- Stoffer, D.S. and Toloi, C.M.C. (1992): *A note on the Ljung-Box-Pierce portmanteau statistic with missing data*. Statistics & probability Letters 13 (1992), blz. 391 – 396.

- Van der Meulen, E.C.J. en Baggelaar, P.K. (2019): *Tijdreeksanalist (TRGnet en TRG) - Tijdreeks-analyse met uitgebreide Box-Jenkins modellering – Versie 6*. AMO, Hengelo, januari 2019, 114 blz.

- Von Asmuth, J.R., Bierkens, M.F.P en Maas., K. (2002): *Transfer function noise modelling in continuous time using predefined impulse response functions*. Water Resources Research, Volume 38, Issue 12, Pages, 23_1 - 23_12.

- Von Asmuth, J.R. (2012): *Groundwater System Identification through Time Series Analysis*. Ph.D. thesis Delft University of Technology, KWR 2012.001, 221 blz.

# Bijlage 1: Algemene toelichting op de Box-Jenkins-methode

Aan de hand van tijdreeksanalyse volgens de Box-Jenkins-methode [Box and Jenkins, 1976], kan de tijdreeks van een variabele beschreven worden met een discreet lineair dynamisch stochastisch model. In het nu volgende geven we een beknopte uitleg van deze methode. Voor een diepgaande theoretische verhandeling wordt verwezen naar het standaardwerk [Box and Jenkins, 1976]. Voor een praktischer gerichte verhandeling wordt verwezen naar [McLeod, 1983].

Deze bijlage hanteert de klassieke Box-Jenkins-notatie en terminologie. Alleen de Box-Jenkins-term *residu* is hier voor de consistentie vervangen door de term *innovatie*, zoals ook elders gebruikt in dit rapport.

### Uitleg aan de hand van systeemtheorie

Het principe van de Box-Jenkins benadering kan goed worden toegelicht aan de hand van begrippen uit de systeemtheorie. Een systeem zal daartoe worden omschreven als een deel van de werkelijkheid, dat via in- en uitgangssignalen wisselwerkingen vertoont met z'n omgeving. De grondwaterstand in een peilbuis zullen we dan opvatten als het uitgangssignaal (of de uitvoerreeks) van een systeem, dat wordt aangedreven door meerdere ingangssignalen (of invoerreeksen). In het geval van een grondwaterstand zullen deze laatste voornamelijk van meteorologische aard zijn.

Een dergelijk systeem kan mathematisch worden uitgeschreven als:

$$Z(t) = g(G, X, t) \qquad \text{[b1.1]}$$

met $Z$ de grondwaterstand, $g(\cdot)$ één of andere functie, $G$ een verzameling modelparameters, $X$ de verzameling invoervariabelen en $t$ de continue tijd.

Als we de grondwaterstand empirisch willen modelleren verdient het aanbeveling deze te beschouwen als de uitvoerreeks van een discreet lineair dynamisch stochastisch systeem:

- discreet: omdat relevante gegevens over de grondwaterstand en de invoerreeksen meestal slechts beschikbaar zijn in de vorm van discrete tijdreeksen;
- lineair: omdat dit de minste theoretische en praktische problemen oplevert, bovendien is gebleken dat veel systemen redelijk beschreven kunnen worden met lineaire modellen;
- dynamisch: omdat de tijd een essentiële rol speelt in het gedrag van de uitvoerreeks;
- stochastisch: omdat zo'n systeem nooit exact (deterministisch) kan worden beschreven, enerzijds vanwege de benadering die is ingevoerd door de discretisatie en linearisatie, anderzijds vanwege het feit dat het onmogelijk is álle invoerreeksen bij de beschrijving mee te nemen.

Een discreet lineair dynamisch stochastisch systeem kan mathematisch worden uitgeschreven als:

$$Z_t = f(F, X', t) \qquad \text{[b1.2]}$$

met $Z$ de grondwaterstand, $f(\cdot)$ een lineaire functie, $F$ een verzameling modelparameters, $X'$ een deelverzameling van de invoerreeksen, $N$ de ruis en $t$ de discrete tijdsindex. Omdat de algemene vorm van de Box-Jenkins modellen ook kan worden uitgeschreven als bovenstaande formule, is de Box-Jenkins-methode bij uitstek geschikt om zo'n systeem langs empirische weg te beschrijven. En omdat de methode er op is gericht het stochastische element terug te brengen tot een verschijnsel dat minimaal is en dat bekende waarschijnlijkheidswetten volgt, kunnen er uitspraken omtrent het systeem worden gedaan met minimale en kwantificeerbare onzekerheden.

### Statistisch modelleren van een tijdreeks

De klassieke statistische methodologie is zelden geschikt voor een statistische modellering van tijdreeksen, omdat deze gebaseerd is op onafhankelijke waarnemingen, terwijl de waarnemingen in een tijdreeks vaak niet onafhankelijk van elkaar zijn. De belangrijkste bijdragen van Box en Jenkins tot de oplossing van dit probleem bestonden uit de formulering van een familie van statistische modellen om tijdreeksen te beschrijven en uit aanwijzingen tot een verantwoorde modelbouw [Box and Jenkins, 1976]. Voor de modellering van een tijdreeks ontwikkelden ze twee benaderingen, namelijk univariate modellering aan de hand van een ARIMA-model en transfer-ruismodellering aan de hand van een transfer-ruismodel. Beide zullen in het nu volgende kort en voornamelijk intuïtief worden omschreven.

### Univariaat modelleren

Bij de univariate modellering wordt uitsluitend uitgegaan van de te modelleren tijdreeks, die wordt beschouwd als de uitvoerreeks van een discreet lineair dynamisch stochastisch systeem dat wordt aangedreven door een stochastische invoerreeks. Deze laatste is onbekend en ontstaat eigenlijk pas in de modelleerfase als de innovatiereeks. De modellering is er dan ook op gericht die terug te brengen tot een verschijnsel dat bekende waarschijnlijkheidswetten volgt (normaal verdeelde witte ruis) en een minimale variantie heeft. Die variantie mag in dit verband als een soort maat voor de onzekerheid van het model worden beschouwd.
Witte ruis kan worden opgevat als een volledig toevallig signaal, bestaande uit een opeenvolging van onafhankelijke trekkingen uit een kansverdeling, met gemiddelde nul en variantie $\sigma_a^2$.
Het zal duidelijk zijn dat de univariate benadering nauwelijks inzicht levert in het beschouwde systeem. Het wordt wel veel toegepast om voorspellingen van een bepaalde variabele te genereren. Deze zijn dan uitsluitend gebaseerd op een statistische analyse van het gedrag van deze variabele in het verleden. Deze toepassing heeft dan ook vooral opgang gemaakt in de economie, met zijn gecompliceerde en minder inzichtelijke dynamische systemen.

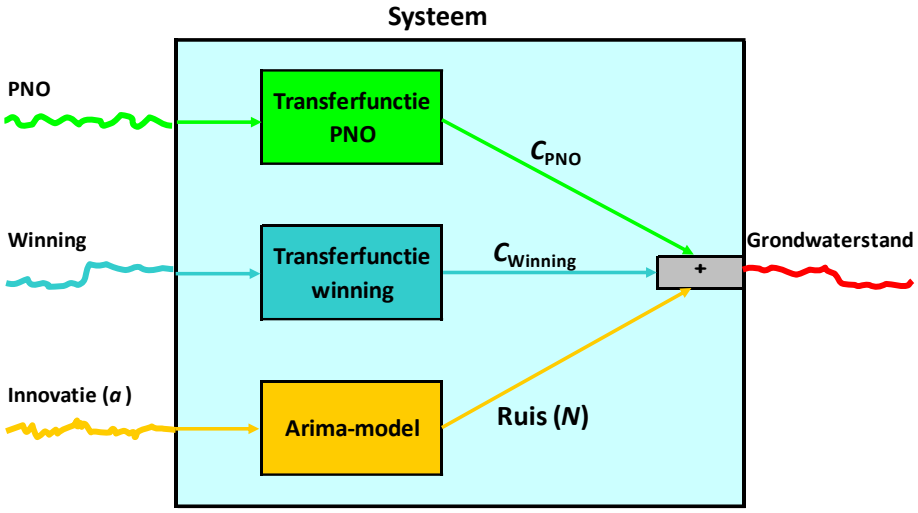
### Transfer-ruismodelleren

Bij de transfer-ruismodellering wordt de te modelleren tijdreeks beschouwd als de uitvoerreeks van een discreet lineair dynamisch stochastisch systeem dat wordt aangedreven door meerdere invoerreeksen, waaronder ook witte ruis. Elke invoerreeks ($X_i$) levert hierbij volgens een afzonderlijke lineaire transferfunctie een component ($C_i$) van de uitvoerreeks ($Y$). De resterende ruis ($N$) vertegenwoordigt de component die wordt geleverd door alle invoerreeksen die niet bij de modellering zijn betrokken. Deze ruis kan op zijn beurt worden weergegeven als de uitvoerreeks van een deelsysteem, dat volgens een ARIMA-model (in feite ook een soort transferfunctie) wordt aangedreven door witte ruis. De verschillende componenten geven door middel van lineaire superpositie de uiteindelijk waargenomen uitvoerreeks, volgens:

$$Z_t = C_{1t} + \cdots + C_{mt} + N_t \qquad \text{[b1.3]}$$

Figuur 1 toont het principe van het transfer-ruismodel, voor het geval een grondwaterstand wordt gemodelleerd als functie van het potentieel neerslagoverschot en een grondwaterwinning.

*Figuur b1.1: Principe van het transfer-ruismodel, hier weergegeven voor een systeem met als uitvoer de grondwaterstand en als invoer het potentieel neerslagoverschot, een grondwaterwinning en innovaties.*




De kracht van het transfer-ruismodel is dat een groot aantal dynamische situaties beschreven kan worden met slechts een gering aantal modelparameters.

### Ontwikkelen van een Box-Jenkins model

Het ontwikkelen van een Box-Jenkins model is - net als bij de meeste modellen - een iteratief proces. Per ronde worden drie fasen doorlopen:

1) identificatie van de vorm het model;
2) schatting van de modelparameters en
3) verificatie van het model.

In het nu volgende beschrijven we de drie fasen afzonderlijk, met als toepassing de ontwikkeling van een transfer-ruismodel.

### Identificeren van de modelvorm

Aan de hand van diagnostieken van de relaties tussen de uitvoerreeks en de invoerreeksen en/of fysisch inzicht in deze relaties, wordt de vorm van het model gepostuleerd. De diagnostieken worden geleverd door de kruiscorrelatiefuncties van de uitvoerreeks en elke afzonderlijke invoerreeks. Zo'n functie geeft de kruiscorrelatiecoëfficiënt, een maat voor de samenhang tussen de waarde van de invoerreeks en de uitvoerreeks *k* tijdseenheden later, als functie van *k* (de tijdsverschuiving, of *lag*). Een kruiscorrelatiefunctie wordt vaak geplot weergegeven, omdat dit de diagnose aanzienlijk vergemakkelijkt. Omdat het beeld van de relatie vertroebeld kan worden door de autocorrelaties binnen de reeksen, worden beide vóór het berekenen van de kruiscorrelatiefunctie door een filter gehaald, namelijk het inverse ARIMA-model voor de invoerreeks. Deze aanpak noemt men ook wel *prewhitening* [Box and Jenkins, 1976].

In deze identificatiefase dient tevens de vorm van het ruismodel te worden geformuleerd. Hiervoor zijn verschillende benaderingen mogelijk. Als er met grote waarschijnlijkheid van kan worden uitgegaan dat alle relevante invoerreeksen in het transfer-ruismodel zijn opgenomen, kan het ruismodel worden geformuleerd als een stationair model. Zo'n model beschrijft een tijdreeks die voldoet aan een bepaald statistisch evenwicht en in feite met een constante spreiding rond een constante waarde


© AMO en PB Icastat

schommelt. Als er daarentegen (vermoedelijk) relevante invoerreeksen ontbreken, kan het ruismodel in eerste instantie worden geformuleerd als het ARIMA-model voor de uitvoerreeks. In alle gevallen zal de juistheid of onjuistheid van de formulering van het ruismodel blijken in de verificatiefase.

### Schatten van de modelparameters

De parameters van het geïdentificeerde model worden geschat volgens een optimalisatieprocedure, waarbij de som van de kwadraten van de innovaties - een functie van de modelparameters - wordt geminimaliseerd. Dit noemt men Kleinste Kwadratenmethode.

### Verifiëren van het model

In de verificatiefase wordt nagegaan of het model de tijdreeks adequaat beschrijft en of voldaan wordt aan alle vooronderstellingen die aan het model ten grondslag liggen (innovaties normaal verdeeld, onafhankelijk van elkaar en zonder relatie met de afzonderlijke invoerreeksen). Hierbij wordt gebruik gemaakt van diagnostieken van de innovaties, zoals een plot, een histogram en enkele correlatiefuncties, alsmede de schatting van de variantie van de innovaties, de statistische significantie van de modelparameters en de correlatiematrix van de parameterschatters. Bij een onbevredigende diagnose kan een betere modelvorm geformuleerd worden aan de hand van de geconstateerde discrepanties. Het ontwikkelproces vervolgt dan weer met de schatting van de modelparameters.

### Betrouwbaarheidsinterval van een geschatte parameter

De schatting van de parameter van een transfer-ruismodel kan opgevat worden als een waarde van een schatter, een stochastische (of toevals-)variabele met een bepaalde kansverdeling. Een schatting gaat gepaard met een maat voor de spreiding van deze kansverdeling: de standaardafwijking (de wortel uit de variantie). Deze standaardafwijking van een schatter wordt ook wel als standaardfout aangeduid. Als voldaan wordt aan de vooronderstellingen die ten grondslag liggen aan het transfermodel, heeft een schatter een normale kansverdeling waarvan het gemiddelde overeenkomt met de echte waarde van de te schatten parameter. Aan de hand van de schatting en de standaardfout kan dan het interval worden aangegeven waarbinnen de echte waarde van de parameter zal liggen met een bepaalde betrouwbaarheid (meestal wordt hiervoor 95% genomen). Dit noemt men het (95%-)betrouwbaarheidsinterval. Als dit interval niet de waarde 0 bevat, is de modelparameter statistisch significant.

### Formuleren van een transfer-ruismodel

De Box-Jenkins formulering van de dynamische relatie tussen een invoerreeks en een uitvoerreeks van een systeem is afgeleid van de algemene weergave van een discreet lineair dynamisch systeem met een lineaire differentievergelijking. Voor een systeem met één invoerreeks geldt bijvoorbeeld de volgende differentievergelijking:

$$(1 + C_1\nabla + \cdots + C_r\nabla^r)Z_t = g(1 + D_1\nabla + \cdots + D_s\nabla^s)X_t \qquad \text{[b1.4]}$$

met $Z$ de afwijking van de uitvoerreeks ten opzichte van zijn gemiddelde, $X$ de afwijking van de invoerreeks ten opzichte van zijn gemiddelde, $t$ de discrete tijdsindex, $C_1 \ldots C_r$ en $D_1 \ldots D_s$ modelparameters, $\nabla$ de differentie-operator ($\nabla Z_t = Z_t - Z_{t-1}$) en $g$ de evenwichts- of stationaire relatie tussen $Z$ en $X$. Deze evenwichtsrelatie (Engels: *steady-state gain*) is een belangrijke karakteristiek van een dynamische relatie. Het is de uiteindelijke stationaire verandering van de uitvoerreeks als de invoerreeks een constante niveauverandering van één eenheid heeft. Of, in het geval van de modellering van de grondwaterstand aan de hand van een grondwaterwinning, de uiteindelijke stationaire verlaging van de grondwaterstand als de grondwaterwinning een constante niveauverandering van één eenheid heeft (er van uitgaande dat de relatie causaal is en niet louter statistisch).


© AMO en PB Icastat

De oplossing van bovenstaande differentievergelijking kan worden uitgeschreven als:

$$Z_t = \sum_{u=0}^{\infty} v_u X_{t-u} = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \ \ldots \qquad [b1.5]$$
$$= (v_0 + v_1 B + v_2 B^2 + \ \ldots) X_t$$

met $v_0$, $v_1$, $v_2$... de impuls-responsgewichten (die samen de discrete impuls-responsfunctie vormen) en *B* de backshift-operator ($BX_t = X_{t-1}$, let wel: *B* is geen modelparameter, maar slechts een operator die dient om de notatie te vereenvoudigen). Deze impuls-responsfunctie vertegenwoordigt de dynamica van het systeem en beschrijft de reactie van de uitvoerreeks *Z* op een eenheidspuls in de invoerreeks *X*. En als X bestaat uit een opeenvolging van verschillende pulsen, wordt *Z* beschreven als de som van door *X* geschaalde impuls-responsgewichten.

In Box-Jenkins notatie wordt een differentievergelijking met één invoerreeks uitgeschreven als:

$$(1 - \delta_1 B - \ \ldots - \delta_r B^r) Z_t = (\omega_0 - \omega_1 B - \ \ldots - \omega_s B^s) X_t \qquad [b1.5]$$

met $\delta_1 \ldots \delta_r$ de autoregressieve parameters, $\omega_0 \ldots \omega_s$ de moving-average parameters en *B* de backshift-operator (hier gebruikt als substituut voor $\nabla$, aangezien $B = 1 - \nabla$). Dit noemt men de transferfunctie van de orde (*r,s*). In nóg compactere vorm wordt dit ook wel als volgt uitgeschreven:

$$Z_t = \frac{\omega(B)}{\delta(B)} X_t \qquad [b1.7]$$

met $\omega(B)$ de moving-average operator voor *X*, volgens:

$$\omega(B) X_t = \omega_0 X_t - \omega_1 X_{t-1} - \ \ldots - \omega_s X_{t-s} \qquad [b1.6]$$

en $\delta(B)$ de autoregressieve operator, die inwerkt op *Z*, volgens:

$$\delta(B) Z_t = Z_t - \delta_1 Z_{t-1} - \ \ldots - \delta_r Z_{t-r} \qquad [b1.7]$$

Het Box-Jenkins transfer-ruismodel is een model om een tijdreeks te beschrijven als een som van transferfuncties van de relevante invoerreeksen, aangevuld met een ruisterm. In deze zin vormt het een discreet lineair dynamisch stochastisch model. In compacte vorm kan het transfer-ruismodel als volgt worden uitgeschreven:

$$Z_t = \frac{\omega 1(B)}{\delta 1(B)} X_{1t} + \ \ldots + \frac{\omega m(B)}{\delta m(B)} X_{mt} + N_t \qquad [b1.10]$$

met $X_1 \ldots X_m$ de invoerreeksen en *N* de ruis. De tijdreeks van de ruis vertegenwoordigt de invloed van alle invoerreeksen die niet in het model zijn opgenomen. Over het algemeen bevat zo'n reeks nog regelmatigheden die beschreven kunnen worden met een univariaat model (ook wel ARIMA-model genaamd).

### Het univariate model (ARIMA-model)

Een univariaat model beschrijft een variabele als lineaire functie van voorgaande waarden van deze variabele en de innovaties. De algemene vorm van een univariaat model is:

$$\phi(B) \cdot \Phi(B^S) \cdot \left(\nabla^d \nabla_S^D Z_t^\lambda - c\right) = \theta(B) \cdot \Theta(B^S) \cdot a_t \qquad [b1.8]$$

met *Z* de variabele, *a* de innovatie, *t* de tijdsindex, $\nabla$ de differentie-operator ($\nabla Z_t = Z_t - Z_{t-1}$), *d* het aantal differenties, *S* de seizoensperiode (met maandelijkse waarnemingen en een jaarcyclus is deze bijvoorbeeld 12), $\nabla_S$ de seizoensdifferentie-operator ($\nabla_S Z_t = N_t - N_{t-S}$), *D* het aantal seizoensdifferenties, $\lambda$ de transformatieparameter (deze is 1 als *Z* normaal verdeeld is) en *c* een constante.

Verder bevat deze formulering de volgende operatoren:

- $\phi(B)$ de autoregressieve operator, volgens:


© AMO en PB Icastat 

$$\phi(B)Z_t = Z_t - \phi_1 Z_{t-1} - \ldots - \phi_p Z_{t-p} \quad \text{[b1.12]}$$

met $\phi_1 \ldots \phi_p$ de autoregressieve parameters;

- $\Phi(B^S)$ de autoregressieve seizoensoperator, volgens:

$$\Phi(B^S)Z_t = Z_t - \Phi_1 Z_{t-S} - \ldots - \Phi_P Z_{t-PS} \quad \text{[b1.13]}$$

met $\Phi_1 \ldots \Phi_P$ de autoregressieve seizoenparameters;

- $\theta(B)$ de moving-average operator, volgens:

$$\theta(B)a_t = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \quad \text{[b1.9]}$$

met $\theta_1 \ldots \theta_q$ de moving-average parameters;

- en $\Theta(B^S)$ de moving-average seizoensoperator, volgens:

$$\Theta(B^S)a_t = a_t - \Theta_1 a_{t-S} - \ldots - \Theta_Q a_{t-QS} \quad \text{[b1.15]}$$

met $\Theta_1 \ldots \Theta_Q$ de moving-average seizoenparameters.

## Referenties


- Box, G.E.P. and Jenkins, G.M. (1976): *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco.
- McLeod, G. (1983): *Box Jenkins in Practice*. Time Series Library, Gwilym Jenkins & Partners Ltd., Lancaster.

# Bijlage 2: Verkenning flexibiliteit Gammaverdeling

Met een verkennende exercitie is nagegaan in hoeverre de Gammaverdeling flexibiliteit verschaft om ook bij andersoortige transferfuncties bruikbare resultaten op te leveren. Daartoe zijn simulaties uitgevoerd waarbij de neerslagcomponent is gesimuleerd met een transferfunctie gebaseerd op een links afgeknotte normale verdeling. Deze laatste heeft de volgende functie:

$$f(x) = B\frac{e^{-\frac{1}{2}(x-c)/\theta}}{\sqrt{2\pi}}, \text{voor } x \geq 0 \qquad \text{[b2.1]}$$

waarin *B* en $\theta$ schaalfactoren, *x* de tijd en *c* de verwachtingswaarde voor de niet-afgeknotte verdeling.

De simulaties zijn uitgevoerd voor verschillende waarden van *c*. De responstijd van de neerslag is ingesteld op 365 dagen. De schaalfactor $\theta$ is afgeleid als de ratio van deze responstijd en de *x*-waarde die 99,999% van de cumulatieve verdeling geeft (de *cutoff*). Voor de transferfunctie van de neerslag is uitgegaan van de blokresponsfunctie die is afgeleid uit de betreffende verdeling. De schaalfactor *B* is zodanig ingesteld dat de evenwichtsrelatie (*A*, het oppervlak onder de blokresponsfunctie) uitkomt op 100 cm/(mm/dag).

Voor de verkenning zijn voor oplopende waarden van *c* van -2$\theta$ t/m 2$\theta$ (met stapgrootte 0,2$\theta$) reeksen van de neerslagcomponent gesimuleerd, uitgaande van de reeks van de dagsom van de neerslag te De Bilt over de periode 1 januari 1999 t/m 12 april 2019. Van elke zo gesimuleerde neerslagcomponent zijn 200 grondwaterstandreeksen gemaakt, door er 200 ruisreeksen aan toe te voegen. Elk van deze ruisreeksen is met een AR(1)-model (met $\phi_1$ = 0,9) gegenereerd uit een reeks onafhankelijke trekkingen uit een normale kansverdeling met gemiddelde nul, zodanig dat de variantie van de ruis 30% bedraagt van de variantie van de neerslagcomponent. De variantie van de ruis bedraagt daardoor 23,1%% van de variantie van de grondwaterstand.

De modellering van de grondwaterstandreeksen is met Pastas uitgevoerd voor de periode vanaf 1 januari 2005 t/m 12 april 2019, zowel zónder als mét een ruismodel, waarbij is uitgegaan van de Gammaverdeling.

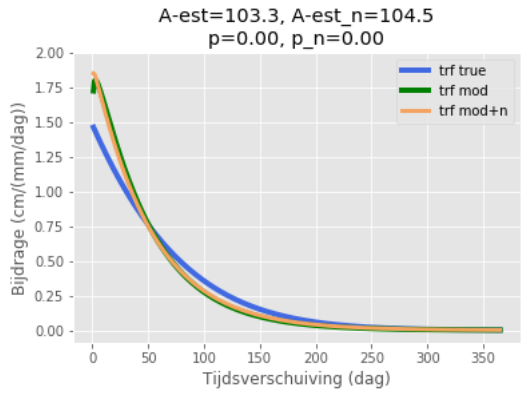
### Bespreking van de resultaten

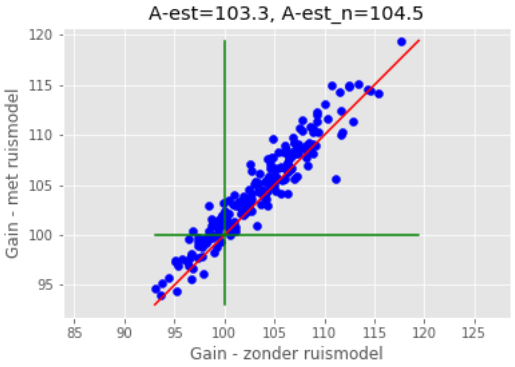
Uit de resultaten (hieronder weergegeven) blijkt dat er voor vrijwel alle uitvoeringen sprake is van het onvoldoende reproduceren van de gesimuleerde relatievorm, wat voor elk van deze ook tot uiting komt in een gemiddelde van de 200 geschatte evenwichtsrelaties dat statistisch significant afwijkt van de ingestelde waarde (100 cm/(mm/d)). De relatieve onzuiverheid kan daarbij oplopen tot 37% van de werkelijke evenwichtsrelatie. In vrijwel alle gevallen schat Pastas zónder toepassing van een ruismodel de evenwichtsrelatie beter dan mét toepassing van een ruismodel. Blijkbaar kan de toepassing van het huidige ruismodel schade opleveren.
De geconstateerde onzuiverheid is uiteraard niet onverwacht, omdat voor het simuleren een andere functie is gebruikt dan waar bij het schatten van het model rekening mee wordt gehouden. Het moet dan ook niet worden gezien als een onvolkomenheid van Pastas, maar van het modelleerproces. Maar het wijst er wel op hoe belangrijk het is om bij het schatten van het model zo goed mogelijk rekening te kunnen houden met de werkelijke vorm van de relatie. De volgende mogelijkheden verdienen in dat kader nadere uitwerking (zie ook hoofdstuk 5):

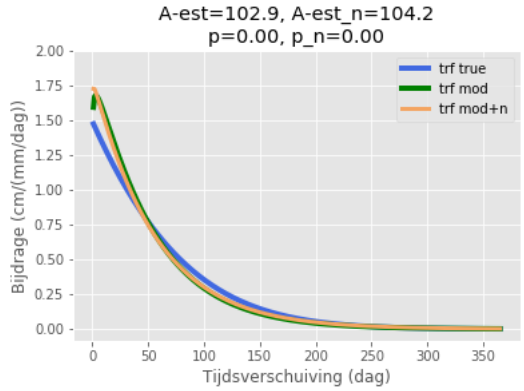
1. Uitbreiden van de verzameling voorgedefinieerde transferfuncties, zodanig dat voor een voldoende breed scala aan hydrologische situaties dynamische relaties met invloedsfactoren afdoende kunnen worden beschreven. Maar dit dient dan wel samen te gaan met het toevoegen van een identificatiestap aan het modelleerproces, waarin op basis van de empirische kruiscorrelatiefunctie van invoerreeks en uitvoerreeks de best daarop aansluitende transferfunctie voor de modelschatting wordt geselecteerd.
2. Ontwikkelen van flexibelere transferfuncties, die in staat zijn allerlei soorten dynamische relaties te modelleren. Als dit wordt losgekoppeld van de restrictie dat elke transferfunctie een hydrologische basis moet hebben, ontstaan er meer mogelijkheden. Zo zijn de transferfuncties die worden toegepast bij Box-Jenkins-tijdreeksmodellering afgeleid van de algemene weergave van een discreet lineair dynamisch systeem met een lineaire differentievergelijking. Met slechts een beperkt aantal parameters kunnen daarmee veel soorten relaties worden beschreven, die bij tijdreeksmodelleringen in allerlei vakgebieden zijn toegepast.

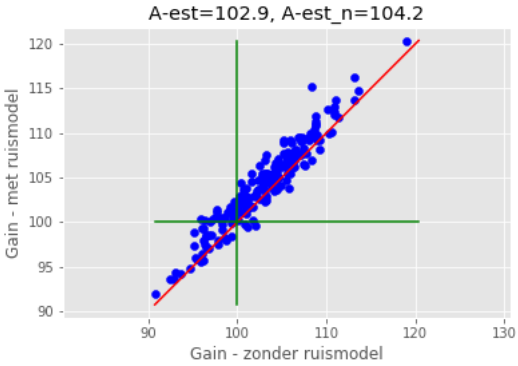
### Resultaten van de exercitie

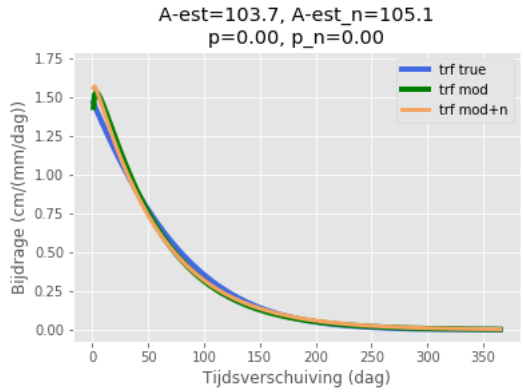
De onderstaande serie plots toont resultaten van het met de Gammaverdelingsfunctie van Pastas schatten van de transferfunctie en de evenwichtsrelatie bij simulatie met een afgeknotte normale verdeling.
Elke linkerplot toont de gehanteerde transferfunctie (*trf true*) en de Pastas-schattingen daarvan, zowel zónder (*trf mod*) als mét toepassing van het ruismodel (*trf mod+n*). Boven elke plot is het gemiddelde van de 200 geschatte evenwichtsrelaties vermeld, zowel zónder (*A-est*) als mét (*A-est_n*) toepassing van het ruismodel. Daaronder is voor beide geschatte gemiddelden vermeld de *p*-waarde bij toetsen op zuiverheid. Als deze kleiner is dan 0,05 is er met 95% betrouwbaarheid sprake van een statistisch significante onzuiverheid bij het schatten van de evenwichtsrelatie.
Elke rechterplot toont het spreidingsdiagram van de twee geschatte evenwichtsrelaties (Y-as mét toepassing van het ruismodel en X-as zónder toepassing van het ruismodel). Voor beide assen is de positie van de werkelijke waarde (100 cm/(mm/d)) weergegeven met een groene lijn. De rode lijn is de diagonaal (X = Y).

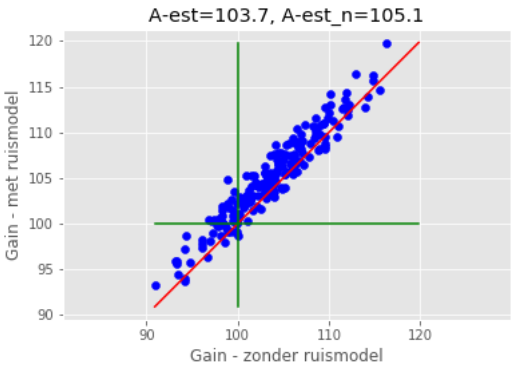


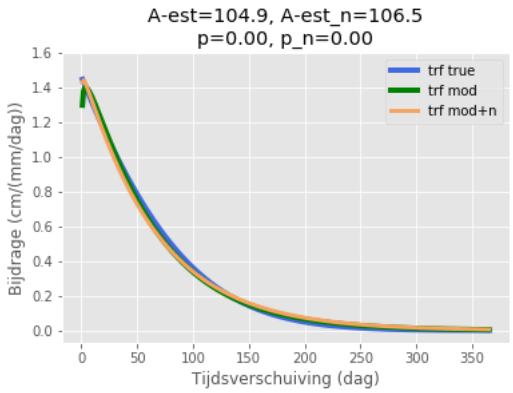



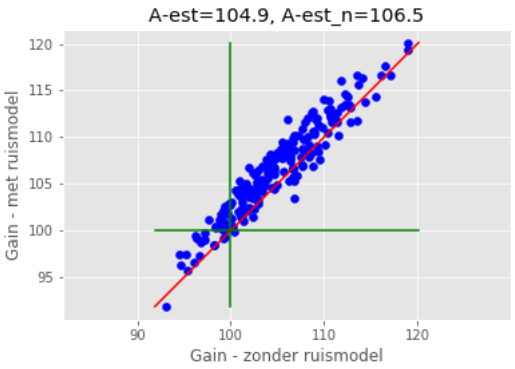



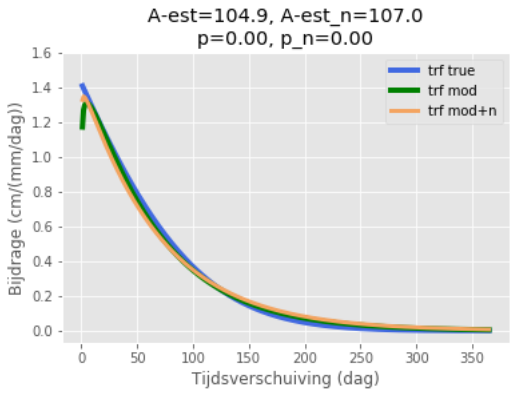



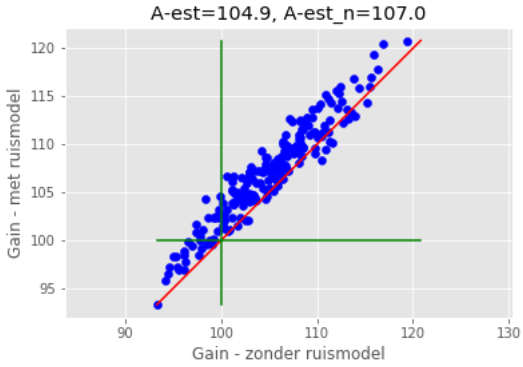



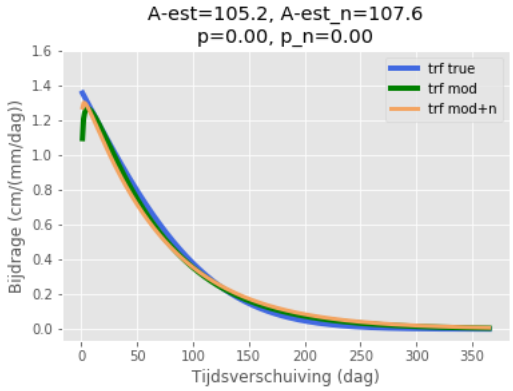



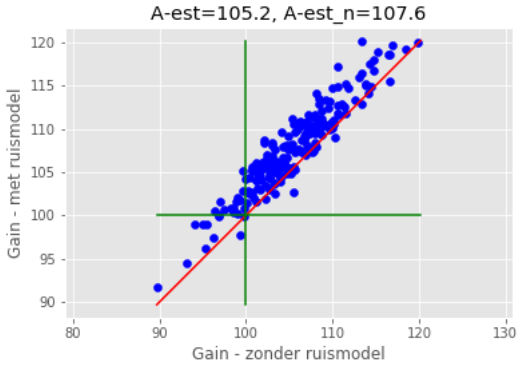


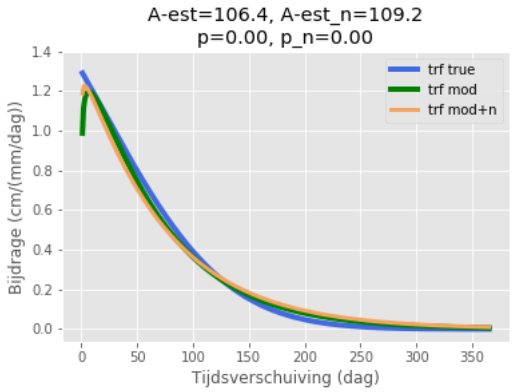
© AMO en PB Icastat

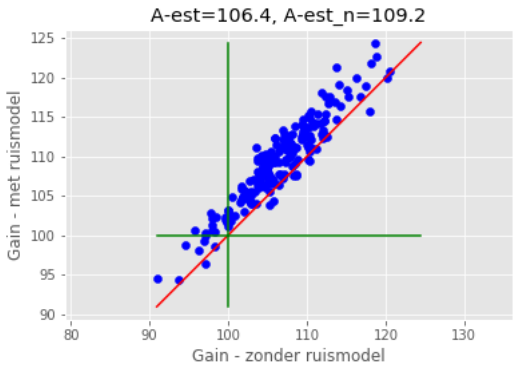


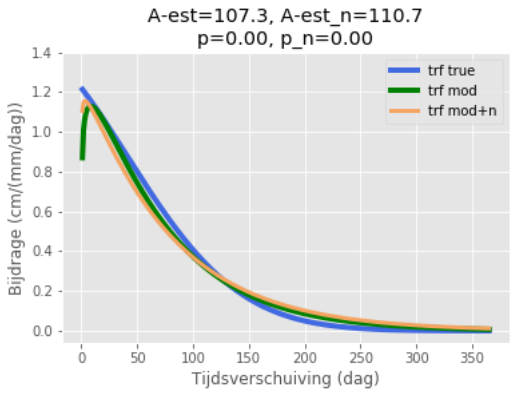



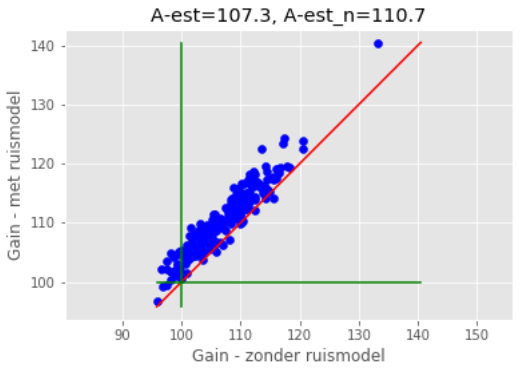



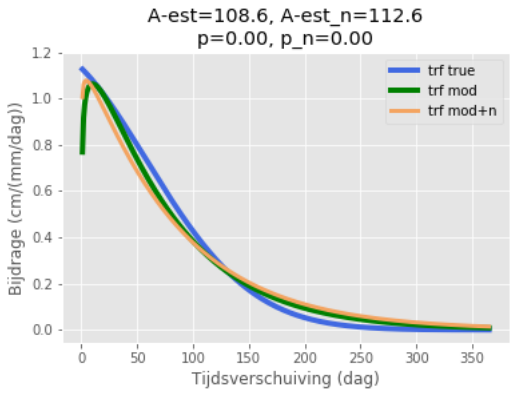



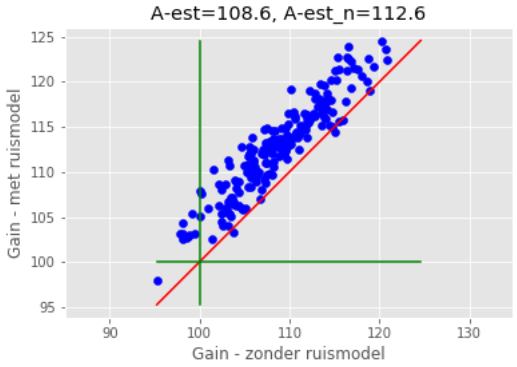



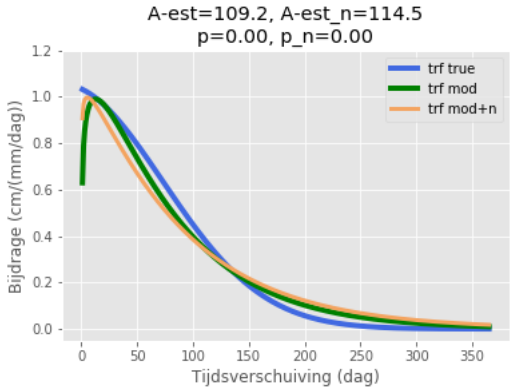



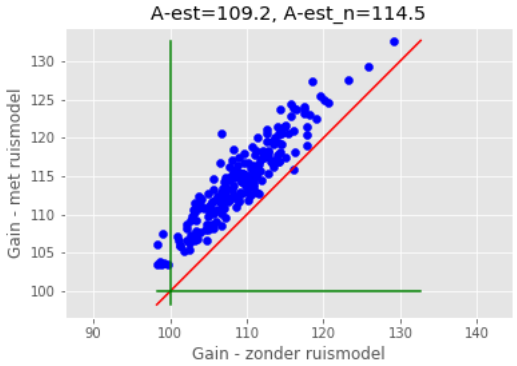


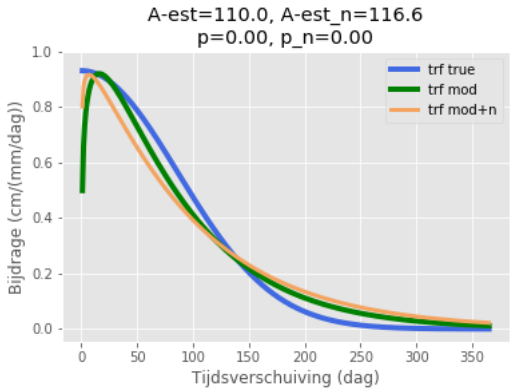

© AMO en PB Icastat 

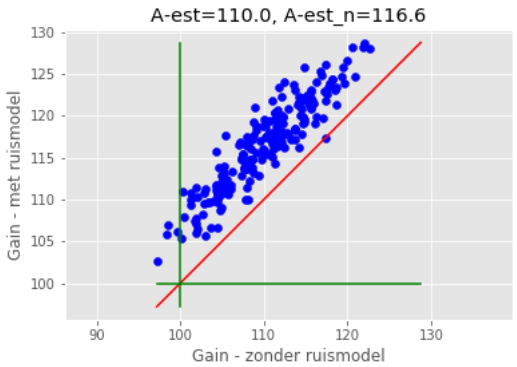


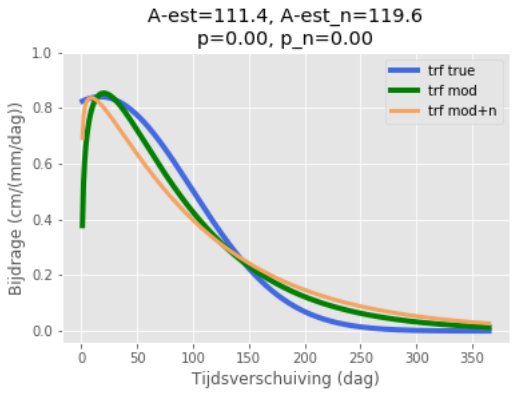



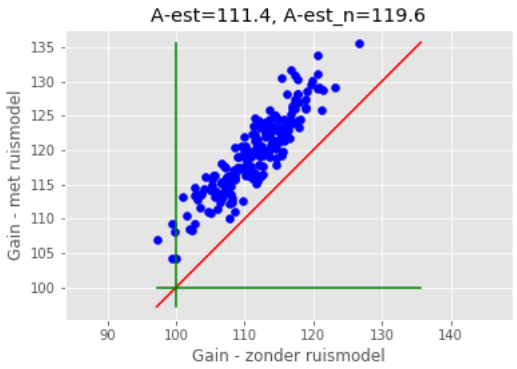



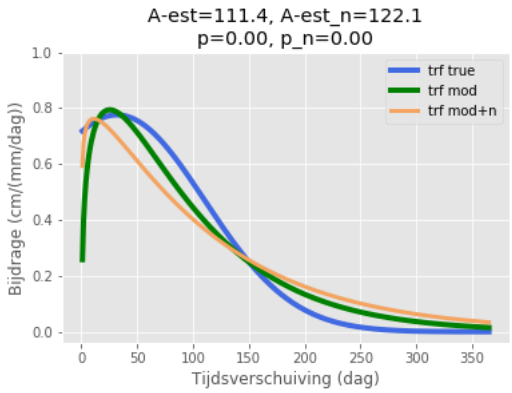



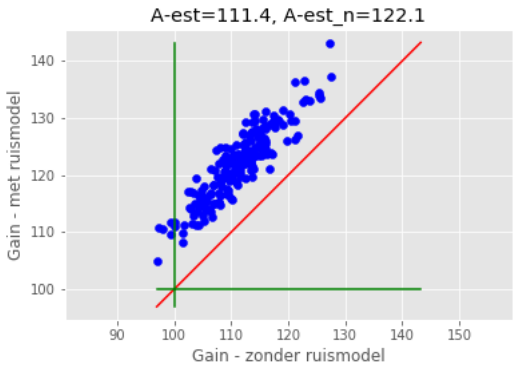



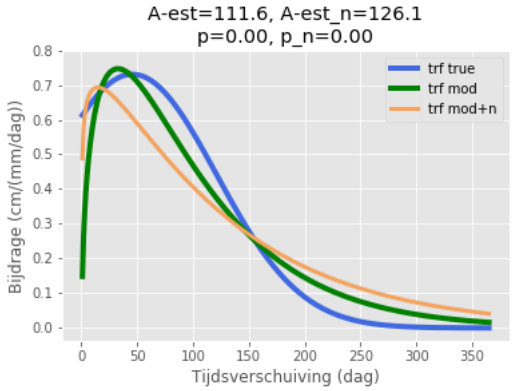



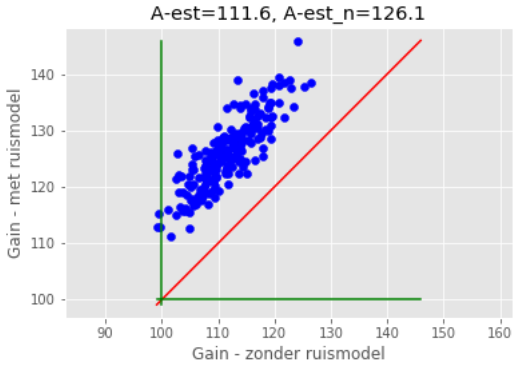


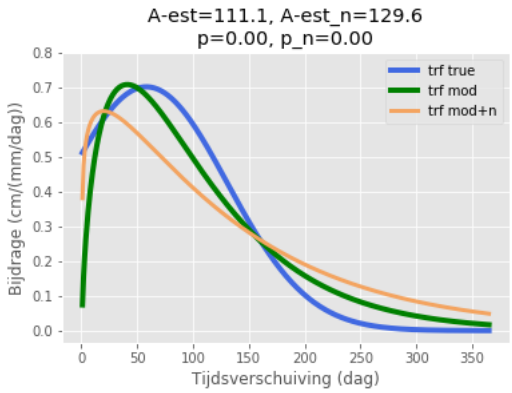
© AMO en PB Icastat

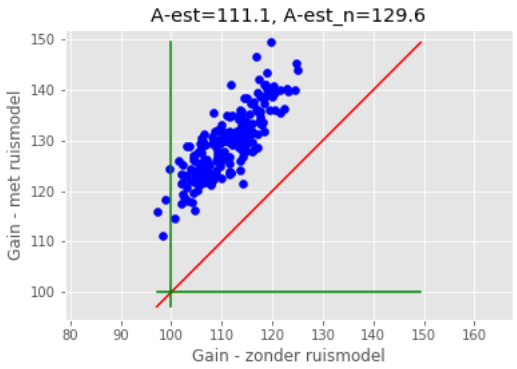


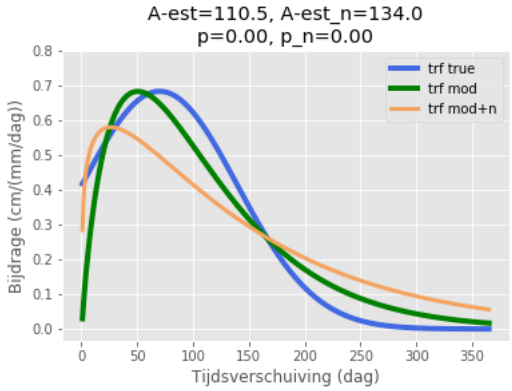



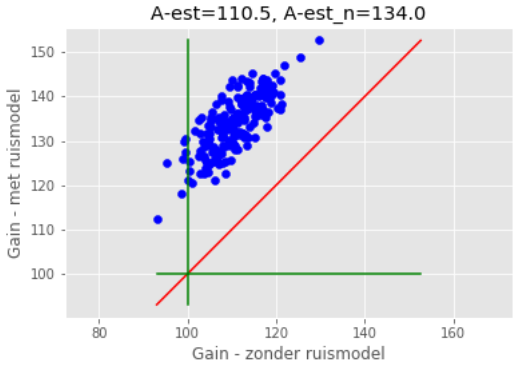



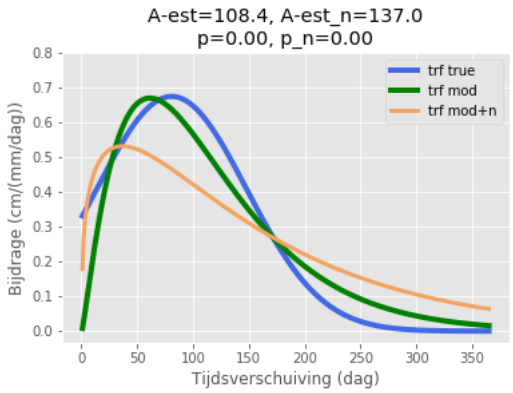



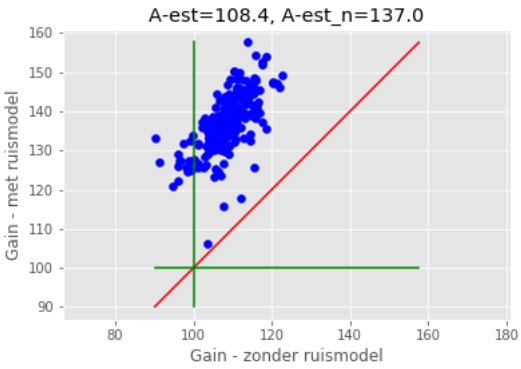



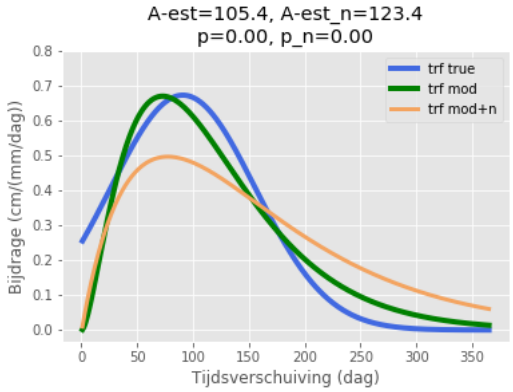



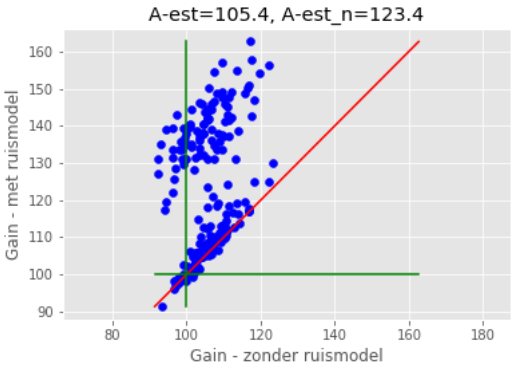


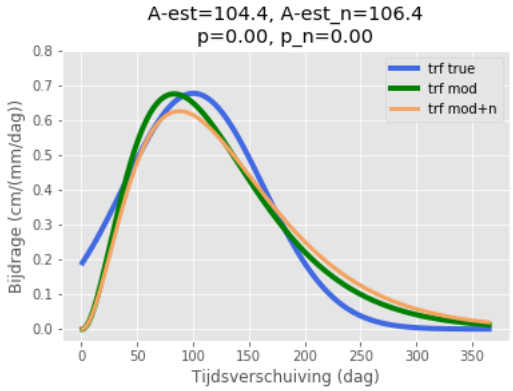

© AMO en PB Icastat 

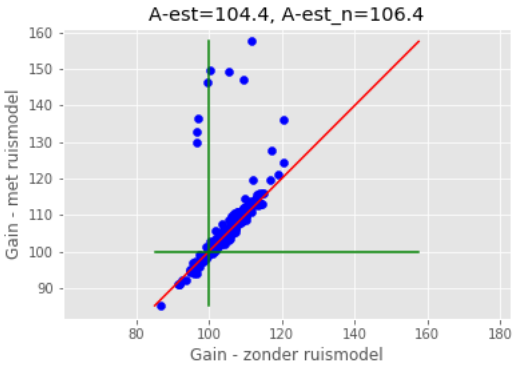


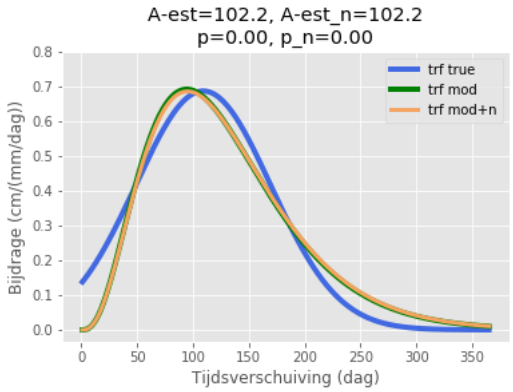



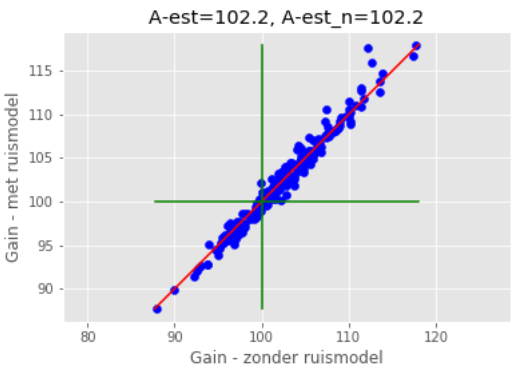



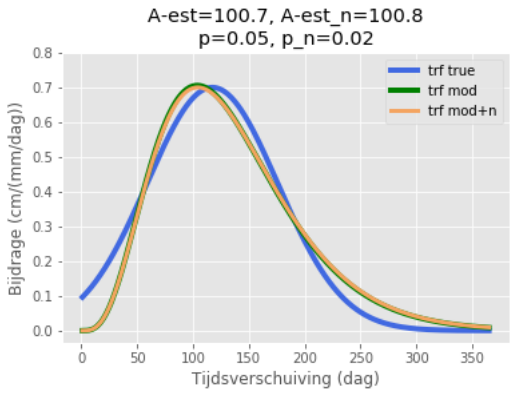



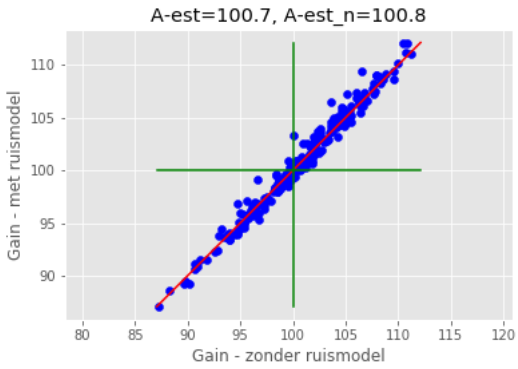





© AMO en PB Icastat






© AMO en PB Icastat 

# Bijlage 3 – Beschrijving Python-module voor modelverificatie

Deze bijlage beschrijft de bij deze studie ontwikkelde Python-module waarmee kan worden geverifieerd of een met Pastas afgeleid tijdreeksmodel voldoet aan bepaalde randvoorwaarden.

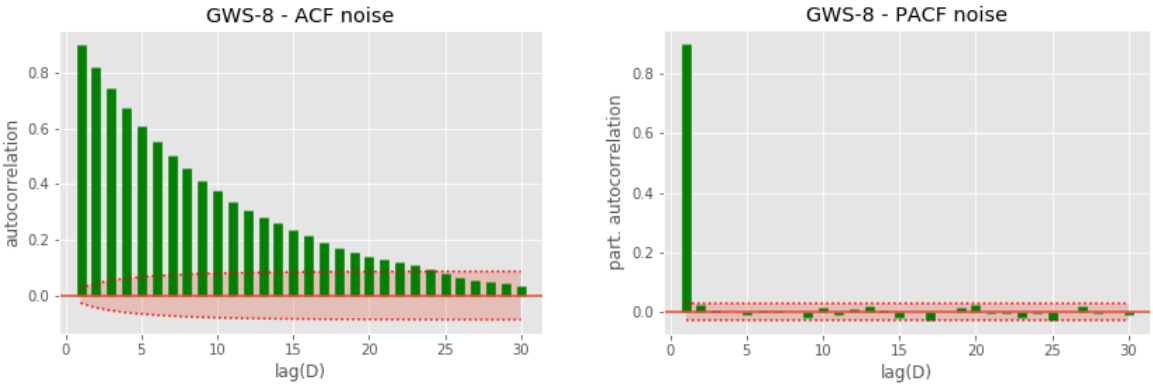
Als het tijdreeksmodel bedoeld is om tot uitspraken met gekwantificeerde betrouwbaarheden te komen (zoals over de grootte van geschatte evenwichtsrelaties met invoerreeksen) zijn de belangrijkste randvoorwaarden dat:

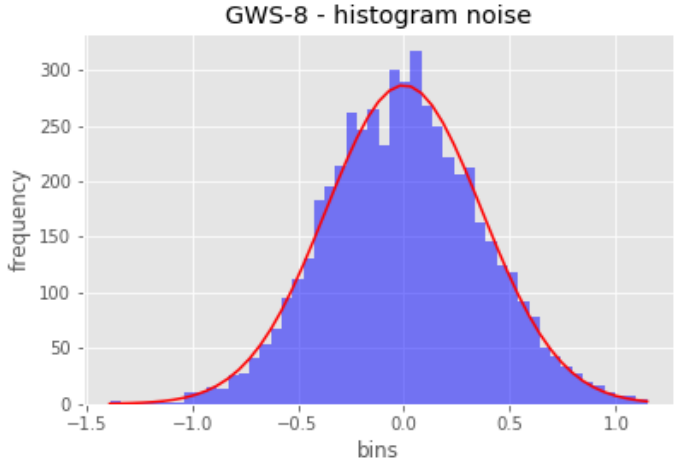
- de vorm van de transferfunctie in het model overeenkomt met de werkelijke vorm en
- de reeks van de restterm (i.c. de ruis als er geen ruismodel is toegepast en de innovaties als er wel een ruismodel is toegepast) normaal verdeelde witte ruis vormt.

Indien dit laatste niet het geval is en er geen verbetering optreedt met toevoegen van invoerreeksen en/of gebruik van andere transferfuncties, dient een uitgebreider ruismodel te worden toegepast, of moet worden gemodelleerd op een laagfrequentere tijdsbasis (zie de voorgestelde werkprocedure in § 5.6).

Onderstaand voorbeelden van de uitvoer van de verificatiemodule. Het betreft een grondwaterstandreeks op dagbasis, die is gesimuleerd uit de neerslag en met toevoeging van AR(1)-ruis met $\phi_1$ = 0,9. Verder betreft het een Pastas-model zónder ruismodel.

*ACF (autocorrelatiefunctie), PACF (partiële autocorrelatiefunctie) en histogram van de ruis van een model zónder ruismodel. In de ACF en de PACF is ook het 95%-betrouwbaarheidsinterval (rood) weergegeven dat geldt onder de nulhypothese van witte ruis. In het histogram is de normale verdeling weergegeven (rood) met het gemiddelde en de standaardafwijking van de ruis.*





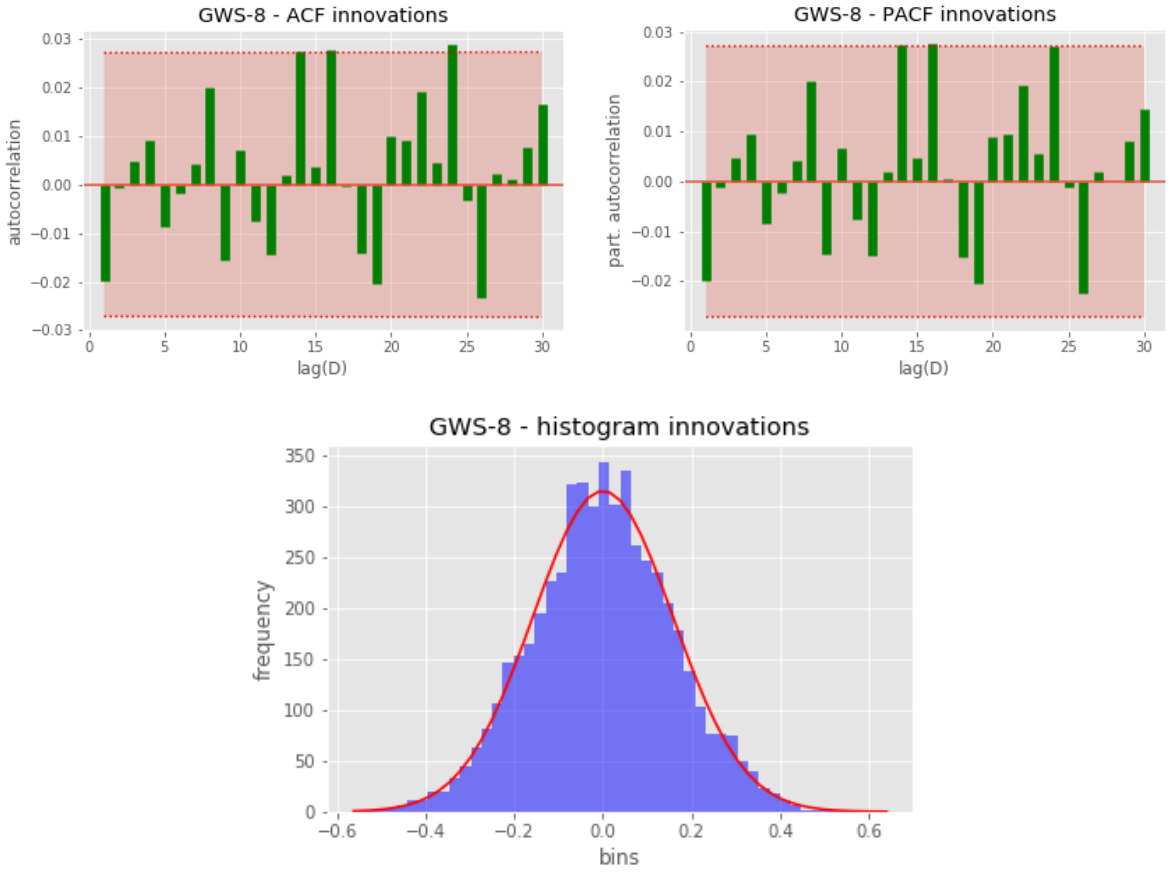

© AMO en PB Icastat 

*Diagnostics of noise of time series model without noise model.*

| Test of noise | Test (confidence) | Result | Detail |
|---|---|---|---|
| D'Agostino-K2 | normality (99%) | Yes | p=0.64 |
| Shapiro-Wilk | normality (99%) | Yes | p=0.02 |
| Portmanteau | no autocorr. (95%) | No | 22702.04>25.00 |
| Runs | no autocorr. (95%) | No | p=0.00 |
| Engle | homosced. (95%) | No | p=0.00 |

Uit de ACF, de PACF en de statistische toetsen blijkt dat er geen sprake is van witte ruis.
Hieronder de verificatie-uitvoer ná het toepassen van het ruismodel van Pastas.

*ACF (autocorrelatiefunctie), PACF (partiële autocorrelatiefunctie) en histogram van de innovaties van een model mét ruismodel.*




*Diagnostics of innovations of time series model with noise model.*

| Test of innovations | Test (confidence) | Result | Detail |
|---|---|---|---|
| D'Agostino-K2 | normality (99%) | Yes | p=0.88 |
| Shapiro-Wilk | normality (99%) | Yes | p=0.48 |
| Portmanteau | no autocorr. (95%) | Yes | 12.22>23.68 |
| Runs | no autocorr. (95%) | Yes | p=0.15 |
| Engle | homosced. (95%) | Yes | p=0.63 |

Uit de ACF, de PACF en de statistische toetsen blijkt dat er nu wél sprake is van witte ruis.


© AMO en PB Icastat

De technische details van de twee toetsen op autocorrelatie van de ruis of de innovaties zijn hieronder beschreven. Voor de beknoptheid worden daarbij zowel ruis als innovaties aangeduid als *modelfout*.

### Portmanteau-toets

De Portmanteau-toets op autocorrelatie baseert zich op een som van de gekwadrateerde autocorrelatiecoëfficiënten van de modelfout. De autocorrelatiecoëfficiënt is een maat voor de lineaire samenhang tussen waarden die zijn gescheiden door een bepaald tijdsinterval. Voor tijdsinterval *k* wordt deze maat geschat als [Box and Jenkins, 1976]:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} e_t \cdot e_{t+k}}{\sum_{t=1}^{n} e_t^{\,2}}$$

met $\hat{\rho}_k$ de geschatte autocorrelatiecoëfficiënt voor tijdsinterval *k*, *e* de modelfout, *t* de tijdsindex en *n* het aantal modelfouten. Als de tijdreeks geen lege tijdseenheden omvat, wordt de toetsingsgrootheid $Q_M$ van de Portmanteau-toets berekend als [Ljung and Box, 1978]:

$$Q_M = n \cdot (n+2) \cdot \sum_{k=1}^{L} \frac{\hat{\rho}_k^2}{n-k}$$

met *L* het aantal tijdsintervallen waarover de autocorrelatie wordt beschouwd (zie onder). Onder de nulhypothese ('géén autocorrelatie') zal $Q_M$ afkomstig zijn uit een $\chi^2$-verdeling met *M* vrijheidsgraden, berekend als $M = L - p$, met *p* het aantal parameters van het ruismodel (in het geval van Pastas is $p$=0 bij modelleren zonder ruismodel en $p$=1 bij modelleren met het AR(1)-ruismodel). We mogen daarom met 95% betrouwbaarheid aannemen dat de modelfout autocorrelatie vertoond als geldt:

$$Q_M > \chi^2_{(95\%,M)}$$

met $\chi^2_{(95\%,M)}$ de waarde van de $\chi^2$-verdeling met een eenzijdige onderschrijdingskans van 95% bij *M* vrijheidsgraden.

Als de tijdreeks daarentegen wél lege tijdeenheden omvat, wordt een aangepaste toetsingsgrootheid van de Portmanteau-toets gehanteerd, die wordt berekend als [Stoffer and Toloi, 1992]:

$$Q_M^* = n^2 \cdot \sum_{k=1}^{L} \frac{C_{a,k} \cdot \hat{\rho}_{e,k}^2}{n-k}$$

met:

$$C_{a,k} = \frac{\sum_{t=1}^{n-k} a_t \cdot a_{t+k}}{n-k}$$

waarbij $a_t$ = 0 als de waarde voor tijdstip *t* ontbreekt, anders is $a_t$ = 1.
En met:

$$\hat{\rho}_{e,k} = \frac{\hat{C}_{z,k} \cdot C_{a,0}}{C_{a,k} \cdot \hat{C}_{z,0}}$$

en:


© AMO en PB Icastat

$$\hat{C}_{z,k} = \frac{\sum_{t=1}^{n-k} z_t \cdot z_{t+k}}{n}$$

waarin $z_t = a_t \cdot e_t$ en $e_t$ de modelfout voor tijdstip $t$.

### Runstoets

De runstoets is een verdelingsvrije toets en vormt een alternatief voor de Portmanteau-toets als de modelfout niet afkomstig is uit een normale kansverdeling. We definiëren een 'run' als een aaneengesloten groep waarden die alle aan dezelfde kant van de mediaan van de reeks liggen. De toetsingsgrootheid van de runstoets is het aantal runs ($r$), met verwachtingswaarde:

$$\mathrm{E}[r] = 1 + \frac{2n_1 \cdot n_2}{n_1 + n_2}$$

en variantie:

$$\mathrm{Var}[r] = \frac{2n_1 \cdot n_2 \cdot (2n_1 \cdot n_2 - n_1 - n_2)}{(n_1 + n_2)^2 \cdot (n_1 + n_2 - 1)}$$

met $n_1$ het aantal waarden groter dan of gelijk aan de mediaan en $n_2$ het aantal waarden kleiner dan de mediaan. Als het aantal waarden groter is dan 50, kan er van worden uitgegaan dat $r$ onder de nulhypothese (*géén autocorrelatie*) een normale kansverdeling volgt. Uit $r$ kan dan, mede met behulp van een continuïteitscorrectie, de gestandaardiseerde toetsingsgrootheid $Tr$ worden berekend, als:

$$T_r = \frac{r - 0{,}5 - \mathrm{E}[r]}{\sqrt{\mathrm{Var}[r]}} \quad \text{als } r - \mathrm{E}[r] > 0$$

$$T_r = \frac{r - \mathrm{E}[r]}{\sqrt{\mathrm{Var}[r]}} \quad \text{als } r - \mathrm{E}[r] = 0$$

$$T_r = \frac{r + 0{,}5 - \mathrm{E}[r]}{\sqrt{\mathrm{Var}[r]}} \quad \text{als } r - \mathrm{E}[r] < 0$$

Onder de nulhypothese zal $T_r$ een standaardnormale kansverdeling volgen. We mogen daarom met 95% betrouwbaarheid aannemen dat de reeks autocorrelatie vertoont als voor de absolute waarde van $T_r$ geldt:

$$|T_r| > U_{(97{,}5\%)}$$

met $U_{(97,5\%)}$ de waarde van de standaardnormale kansverdeling met een eenzijdige onderschrijdingskans van 97,5%.

Als het aantal waarden daarentegen niet groter is dan 50, wordt de tweezijdige overschrijdingskans bepaald van het aantal runs $r$, uitgaande van de binomiale kansverdeling. Als die tweezijdige overschrijdingskans kleiner is dan 5%, mogen we met 95% betrouwbaarheid aannemen dat de reeks autocorrelatie vertoont.



© AMO en PB Icastat 

# Bijlage 4 – Reactie van Pastas-ontwikkelteam

Reactie van het Pastas-ontwikkelteam op het rapport 'Naar betere tijdreeksmodellering met Pastas' van Paul Baggelaar en Eit van der Meulen

Pastas ontwikkelteam: Universiteit van Graz (Raoul Collenteur), Artesia (Ruben Caljé, Davíd Brakenhoff, Onno Ebbens, Frans Schaars), TU Delft (Mark Bakker)

Pastas is een gratis open-source pakket voor tijdreeksanalyse van geohydrologische tijdreeksen. Wij als ontwikkelteam zijn blij met de uitgebreide analyse van de auteurs van het rapport `Naar betere tijdreeksmodellering met Pastas' en verwelkomen de suggesties voor verbetering en uitbreiding van Pastas. Door middel van deze schriftelijke reactie hopen wij een aantal zaken uit het rapport scherp te krijgen en zo een goede inhoudelijke discussie te generen. Deze analyse is mogelijk omdat Pastas gratis en open-source is. Iedereen kan het programma dus gebruiken en in de code kijken wat het programma doet. We zien toevoegingen van iedereen, inclusief de auteurs van het rapport, graag tegemoet, bij voorkeur als een Pull Request op de GitHub website van Pastas. De auteurs trekken een aantal conclusies over de mogelijkheden en toepasbaarheid van Pastas. Een aantal conclusies gelden meer in het algemeen voor tijdreeksmodellen (waaronder Box-Jenkins modellen) terwijl andere conclusies specifiek gelden voor de PIRFICT methode, waar naast Pastas ook Menyanthes, HydroSight en Metran op gebaseerd zijn.

Het rapport bevat de beschrijving van de toepassing van Pastas voor het modelleren van geohydrologische tijdreeksen die gemeten zijn op dagbasis. De auteurs hebben drie analyses verricht:

1. Een analyse waarin gedemonstreerd wordt dat de betrouwbaarheidsintervallen die geschat worden uit data met een gecorreleerde ruis te smal zijn als de de correlatie in de ruis niet meegenomen wordt in de schatting. Vergelijkbare studies zijn eerder gepubliceerd. Bence (1995), bijvoorbeeld, bestudeerde gecorreleerde data (alleen AR(1) correlatie) voor relatief korte reeksen en laat zien dat het moeilijk is voor korte reeksen met hoge correlatie om de AR(1) parameter goed te schatten. Voor de analyse van Hoofdstuk 2, laten de auteurs alleen de resultaten zien zonder ruismodel (tabellen 2.1 t/m 2.4), maar niet de resultaten met ruismodel. Dat zou inzichtelijk zijn, want wij verwachten dat de empirische betrouwbaarheidsintervallen ook niet altijd kloppen als er wel een ruismodel toegevoegd is, zeker niet voor hoge correlatiecoëfficiënten (Bence, 1995), wat op zich ook weer interessant is. De auteurs benadrukken dat de parameters wél zuiver geschat kunnen worden als geen ruismodel toegepast wordt, maar dat de betrouwbaarheidsintervallen niet nauwkeurig geschat kunnen worden. Dat is een belangrijke uitkomst.

2. Een analyse van 574 reeksen die eerder al met Pastas geanalyseerd zijn door Artesia (Beekman en Caljé, 2018). Van de reeksen wordt alleen het deel geanalyseerd dat op dagbasis gemeten is. De auteurs concluderen dat voor modellen met een ruismodel, de innovaties (noise in Pastas) niet aan de standaard tests voldoen dat de innovaties witte ruis zijn en van een normale verdeling komen. Er kunnen voor deze reeksen dus geen betrouwbaarheidsintervallen worden berekend noch kunnen er statistische uitspraken gedaan worden. Dit doen Beekman en Caljé (2018) ook niet in hun rapport. De conclusie dat toepassing van het huidige eenvoudige ruismodel in Pastas niet leidt tot witte ruis (voor de 95% significantie

die de auteurs toepassen in de tests) komt overeen met onze ervaring. Het zou inzicht verschaffen als de analyse op drie punten uitgebreid wordt. Ten eerste, voldoen de innovaties bij een lagere significantie wel aan de statistische voorwaarden? Ten tweede, voldoet het ruismodel wél als wekelijkse stijghoogten of twee-wekelijkse stijghoogten geanalyseerd worden en hoe verhoudt de hieruit berekende betrouwbaarheid zich tot de betrouwbaarheid op basis van dagelijkse metingen? Ten derde, hoe goed zijn de voorspellingsmogelijkheden van het model met en zonder ruismodel? Wat dit laatste betreft, de EVP is maar één van de mogelijkheden om een tijdreeksmodel te beoordelen. Een andere, zeer gebruikelijke manier is om de metingen te verdelen in een kalibratie deel (waarop het model geijkt wordt) en een validatie deel (waarop gekeken wordt hoe goed het model het doet).

De auteurs gaan uitvoerig in op het resultaat dat de verklaarde variantie (EVP) van een Pastas model altijd lager is als er een ruismodel toegepast wordt. Dit is per definitie het geval omdat de vorm van het deterministische deel van het Pastas model altijd hetzelfde is, of er nu wel of geen ruismodel toegepast wordt. De EVP is het hoogst als de som van de kwadraten van de ruis (residuals in Pastas) zo laag mogelijk is. Dit is het geval als de som van de kwadraten van de ruis geminimaliseerd wordt, hetgeen gebeurt als er geen ruismodel toegepast wordt. Als er wel een ruismodel toegepast wordt, dan wordt de som van de kwadraten van de innovaties geminimaliseerd, hetgeen altijd tot een hogere som van kwadraten van de ruis zal leiden. Het interessante is dat bij toepassing van het TRG model, er wel een aantal buizen zijn (~20%) waarbij de EVP hoger is als een ruismodel toegepast wordt (Figuur 3.2). Dat betekent dat het deterministische deel van het model anders is bij toepassing van het ruismodel dan zonder ruismodel, wat bij Pastas niet het geval is. Het is interessant om te begrijpen hoe dat werkt in TRG, wat voor voordelen dit oplevert en of dat eventueel ook in Pastas voordelen op zou leveren.

3. In de derde analyse, onderzoeken de auteurs of Pastas een synthetische reeks, gegenereerd met een Gamma responsfunctie en gecorreleerde data, kan terugvinden. Dit is het geval en soortgelijke vergelijkingen staan ook op de Pastas website. De auteurs benadrukken vervolgens dat synthetische reeksen die met een andere responsfunctie gegenereerd zijn alleen met een Gamma responsfunctie gesimuleerd kunnen worden als de Gamma functie (vrijwel) dezelfde vorm kan aannemen als de andere responsfunctie. Dit is ook eerder onderzocht door bijvoorbeeld Klop (2019), die synthetische reeksen gegeneerd heeft met responsfuncties uit grondwatermodellen, en vervolgens de reeksen met Pastas heeft gesimuleerd. Klop concludeerde dat Gamma niet altijd goed presteert en heeft daartoe twee nieuwe functies aan Pastas toegevoegd die het in bepaalde situaties wel beter doen (de DoubleExponential en de FourParameter functies).

De auteurs eindigen met zes verbetermogelijkheden voor Pastas die uitmonden in de volgende zeven aanbevelingen:

1. Uitbreiding van de mogelijkheden van het ruismodel. De auteurs stellen voor om geavanceerde ARMA ruis modellen te gebruiken met equidistante tijdstappen. Reactie: Uitbreiding van de mogelijkheden van het ruismodel is zeer gewenst door ons. Het liefst gewoon in het continue domein, maar, als dat nog niet mogelijk is, in eerste instantie voor


© AMO en PB Icastat 

equidistante tijdstappen. Daarbij is het vooral ook interessant of (en hoe) de complexheid van het ruismodel afhangt van het tijdsinterval tussen stijghoogtemetingen.

2. Pas de cyclus identificeren-schatten-verifiëren toe bij Pastas modellen. Reactie: Wij denken dat het zeker een goed idee is om de resultaten van tijdreeksmodellen beter te bekijken en een dergelijke cyclus daarbij te gebruiken. We streven er naar om regels te ontwikkelen die kunnen helpen bij het identificeren van de functionele vorm van het tijdreeksmodel, afgestemd op de methode achter Pastas. We willen daarbij benadrukken dat het hier gaat om een veelvoud van mogelijkheden die anders zijn dan de Box-Jenkins methode. Voorbeelden hiervan zijn: 1) aanpassen de responsfunctie, 2) bepalen welke verklarende tijdreeksen (bijv. welke pompstations) meegenomen moeten worden, 3) verbeteren van de bepaling van de grondwateraanvulling uit neerslag en verdamping, en 4) gebruik grondwaterstanden met een lagere frequentie (groter tijdsinterval) dan dagelijks.
3. Stel de vorm van de transferfunctie empirisch vast gebaseerd op de gewitte kruiscorrelatiefunctie. Reactie: Wij begrijpen dat dit een gebruikelijke manier is bij de Box-Jenkins methode, maar de Pirfict methode (waar Pastas op gebaseerd is), is juist ontwikkeld om dit niet te hoeven doen en zo een groot aantal buizen simultaan te analyseren met de garantie dat de toegepaste responsfunctie fysisch realistisch is. Het zou mooi zijn als de empirische methode gevat kan worden in bepaalde regels om de transferfunctie vast te stellen. Zo'n methode zou een waardevolle toevoeging binnen Pastas kunnen zijn.
4. Stel vast of een Pastas model aan de randvoorwaarden voldoet om statistische uitspraken te doen. Reactie: Dit is een belangrijk punt. Pastas heeft al een aantal functies om deze randvoorwaarden te beoordelen, maar moet zeker uitgebreid worden, bijvoorbeeld met een aantal functies die door de auteurs toegepast zijn. Daarnaast dienen dan duidelijk voorbeelden ontwikkeld te worden waarin uitgelegd wordt wat het betekent als er wel of niet aan deze randvoorwaarden voldaan wordt.
5. Onderzoek of er flexibelere transferfuncties mogelijk zijn. Reactie: Dit lijkt ons een goed idee en er zijn al stappen in deze richting gezet door het toevoegen van, bijvoorbeeld, de DoubleExponential en FourParam functies. Ook is het interessant om nog flexibelere, maar wel fysisch realistische, transferfuncties te onderzoeken, bijvoorbeeld door toepassing van splines of KDE functies (Kernel Density Estimation).
6. Zoek een oplossing voor het feit dat het huidige ruismodel van Pastas in bepaalde gevallen de prestaties van het tijdreeksmodel schaadt. Reactie: De prestatie is hierbij gemeten in EVP die, zoals hierboven uitgelegd, per definitie lager is als een ruismodel is toegepast. Wat voor ons interessant is, is waarom het bij TRG niet zo is. Het lijkt ons interessant om te kijken hoe de prestatie bij bijvoorbeeld voorspellingen is en de prestatie aan de hand daarvan te meten.
7. Toevoeging van functionaliteit om de gebruiker stapsgewijs naar een statistisch verantwoord model te leiden. Reactie: Wij vinden het zeker een goed idee om samen met aanbeveling 4 suggesties te doen hoe het model verbeterd kan worden mocht er niet voldaan worden aan de statistische randvoorwaarden. Het maken van Notebooks met uitgewerkte voorbeelden ondersteunen wij graag.

Pastas is opgezet als een ontwikkelplatform voor tijdreeksanalyse met behulp van fysisch plausibele responsfuncties. In de praktijk blijken er nogal wat uitdagingen te zijn bij het toepassen van tijdreeksanalyse voor vraagstukken met echte data. Het doel van Pastas is om in brede kring theorie toe te passen en te ontwikkelen om deze problemen het hoofd te bieden. Het onderzoek van

Baggelaar en van der Meulen levert daar een goede bijdrage aan, en biedt mogelijkheden om Pastas te verbeteren. De meest voor de hand liggende vervolgstappen op de korte termijn zijn:

- Het toevoegen van de statistische toetsen aan de testbank van Pastas
- Het onderzoeken van de invloed van de structuur van het ruismodel in relatie tot de tijdstapgrootte op de statistische toetsen en de voorspellende waarde voor praktische vraagstukken.

Voor beide aspecten geldt dat notebooks ontwikkeld moeten worden waarin geïllustreerd wordt wat het betekent als aan bepaalde toetsen wel of niet voldaan wordt en wat de gevolgen zijn van keuzes m.b.t. de tijdstapgrootte en het ruismodel. We willen graag met de auteurs samenwerken om deze vervolgstappen verder in te vullen.

**Referenties**


- Beekman, W. en Caljé, R. 2018. *Doorwerking droogte van 2018 in Overijssel. Voorspelling per eind september*. Artesia, 2 oktober 2018, Schoonhoven.
- Bence, J.R., 1995. *Analysis of short time series: correcting for autocorrelation*. Ecology, *76*(2), pp.628-639.
- Klop, S., 2019. *From observation well to model area: Estimating groundwater levels spatially using time series analysis*. MSc thesis, TU Delft. https://repository.tudelft.nl/islandora/object/uuid:362f6b0f-b3aa-41c6-b47e-b625c963d8a0