Invloed Tijdsinterval bij Tijdreeksanalyse

Inleiding	**2**
Inbouwen Statistische Toetsen	**3**
Vergroten Tijdsinterval	**4**
Discussie + Aanbevelingen	**7**
Bijlage 1 Reactie op Artesia-studie Invloed tijdsinterval bij tijdreeksanalyse	**9**

Opdrachtgever:	Provincie Overijssel
Projectnummer:	20.22.98
Datum:	23 oktober 2020

Auteur(s): **R. Collenteur, F. Schaars**

1 Inleiding

Begin 2020 is het rapport ''Naar betere tijdreeksmodellering met Pastas' van Paul Baggelaar en Eit van der Meulen verschenen met daarin suggesties voor het gebruik van Pastas. Zij hebben een aantal statistische toetsen toegepast op met Pastas gemaakte tijdreeksmodellen afkomstig van een studie die Beekman en Caljé voor de provincie Overijssel hebben uitgevoerd. Het bleek dat de modellen niet aan de toetsen voldeden, wat betekent dat er geen uitspraken met statistische betrouwbaarheidsbanden mogelijk zijn, hetgeen dan ook niet gebeurd is in het rapport van Beekman en Caljé. Deze conclusie werd ook geïllustreerd aan de hand van een aantal synthetische voorbeelden. Dit rapport beschrijft het vervolgonderzoek, dat bestaat uit drie stappen:

1) Inbouwen statistische toetsen uit het rapport van Baggelaar en van der Meulen

De toetsen uit de rapportage van Baggelaar en van der Meulen zijn ingebouwd voor zover die niet in Pastas aanwezig waren. Uitgangspunt was dat we zoveel mogelijk gebruik maken van bestaande Python packages (bijvoorbeeld statsmodel), tenzij de functionaliteit niet toereikend is. Er zijn notebooks gemaakt die de werking van de toetsen illustreren.

2) Vergroten tijdsinterval

Er is onderzocht wat het vergroten van het tijdsinterval tussen metingen betekent voor het behalen van de verschillende toetsen. Eerst zijn de resultaten van Baggelaar en van der Meulen gereproduceerd. Vervolgens is het tijdsinterval vergroot en is geanalyseerd of de toetsen nu wel goed doorlopen worden. De resultaten zijn in een notebook gepresenteerd en gedeeld met de opdrachtgevers en Baggelaar en van der Meulen.

3) Workshop

Baggelaar en van der Meulen hebben gereageerd op de notebooks en de concept rapportage. Deze reactie is bijgevoegd als bijlage 1 van dit rapport. Vervolgens heeft er een overleg plaatsgevonden waarbij de discussiepunten zijn besproken. De resultaten van deze discussie zijn weergegeven in hoofdstuk 4. Ook zijn aanbevelingen gedaan voor nader onderzoek.

2 Inbouwen Statistische Toetsen

De “stats” package van Pastas is aangepast en uitgebreid als onderdeel van dit project. Tabel 1 geeft een overzicht van de nieuwe en bestaande functies in de stats package van Pastas die als onderdeel van dit project zijn toegevoegd dan wel verbeterd. Belangrijk daarbij is dat de input parameters en de return parameters van alle functies zoveel mogelijk gelijk zijn getrokken wat het gebruik en de interpretatie van de functies makkelijker maakt. Alle methoden zijn nu volledig gedocumenteerd met behulp van “docstrings” die zowel in de code als op de documentatie website beschikbaar zijn (https://Pastas.readthedocs.io/). Het gebruik en de interpretatie van de nieuwe methoden zijn uitgelegd in een nieuwe Notebook die aan Pastas en de documentatie website is toegevoegd (04_diagnostic_checking.ipynb).

Tabel 1. Overzicht van nieuwe en bestaande statistische toetsen in Pastas.

Functie	**Beschrijving**	**nieuw / bestaand**
ps.stats.plot_diagnostics	Deze functie produceert verschillende visualisaties die gebruikt kunnen worden voor het controleren van de aannames van witte ruis.	Bestaand
ps.stats.plot_acf	Deze functie plot de autocorrelatiefunctie.	Nieuw
ps.stats.diagnostics	Deze functie voert meerdere diagnostische toetsen uit	Nieuw
ps.stats.ljung_box	Statistische toets voor autocorrelatie bij gelijke tijdintervallen.	Bestaand
ps.stats.run_test	Statistische toets voor autocorrelatie bij gelijke en ongelijke tijdsintervallen.	Bestaand
ps.stats.durbin_watson	Statistische toets voor autocorrelatie bij gelijke tijdsintervallen.	Bestaand
ps.stats.stoffer_toloi	Statistische toets voor autocorrelatie bij gelijke en ongelijke tijdsintervallen.[1]	Nieuw
ps.stats.acf	Functie die de autocorrelatie berekend voor gelijke en ongelijke tijdsintervallen.	Bestaand

Alle methoden zijn beschikbaar als losse methoden, en als functie van een Pastas model. De nieuwe methoden zijn zodanig geïmplementeerd dat deze niet alleen toepasbaar zijn op Pastas modellen (bv. ml.stats.diagnostics()), maar ook kunnen worden toegepast op losse tijdreeksen (bv. ps.stats.diagnostics()). Dit heeft als voordeel dat de toetsen ook kunnen worden gebruikt om de residuen of ruis van andere tijdreeksmodellen (bijvoorbeeld afkomstig uit andere software) te toetsen

[1] Eit van der Meulen heeft de eerste versie hiervan via een Pull Request toegevoegd.

op de eigenschappen van witte ruis. De toetsen zijn ingebouwd zonder nieuwe afhankelijkheden aan de Pastas software toe te voegen.

De enige toets die niet is ingebouwd, maar wel was gesuggereerd door Baggelaar en van der Meulen, is de Engle toets voor heteroscedasticiteit. Deze toets is beschikbaar in statsmodel, maar dit package is geen afhankelijkheid van Pastas . De Engle toets gaat echter uit van gelijke tijdsintervallen, wat voor veel Pastas modellen niet opgaat. Het is niet bekend wat de waarde van deze toets is voor tijdreeksen met ongelijke tijdsintervallen. Nader onderzoek is nodig om dit vast te stellen en een goede toets voor heteroscedasticiteit in te bouwen die ook voor ongelijke tijdsintervallen is getest en werkt.

3 Vergroten Tijdsinterval

Een mogelijke oplossing voor het geval wanneer een tijdreeksmodel niet aan de voorwaarden van witte ruis voldoet is het vergroten van het tijdsinterval tussen de gemeten grondwaterstanden. Het tijdreeksmodel wordt daarbij nog steeds op dagbasis doorgerekend (neerslag en verdamping per dag), maar voor de kalibratie van de modelparameters wordt een selectie van de beschikbare metingen gebruikt. De 436 modellen in Overijssel uit het "Beekman" project zijn gekalibreerd op tijdreeksen van grondwaterstanden met een toenemend tijdsinterval tussen de metingen, en vervolgens gecontroleerd met behulp van verschillende statistische toetsen. Twee significantieniveaus zijn daarbij gebruikt: alpha=0.05 en alpha=0.01. Voor meer informatie wordt verwezen naar het bijgevoegde pdf bestand.

4 Discussie + Aanbevelingen

Naar aanleiding van de notebooks en het concept rapport hebben Baggelaar en van der Meulen een reactie geschreven (zie bijlage 1). Deze reactie is puntsgewijs besproken tijdens een online overleg op 28 september 2020.

toetsen op autocorrelatie

Voor toetsing op autocorrelatie kunnen onder andere de Ljung-Box en de Stoffer-Toloi toets gebruikt worden. De eerste is geschikt voor gelijke tijdsintervallen en de tweede voor reeksen met missende waarden. Bij gelijke tijdsintervallen geven deze twee toetsen identieke resultaten. Bij beide toetsen moet er een parameter L worden gekozen, het aantal lags waarvoor op autocorrelatie wordt gecontroleerd. Daarvoor zijn diverse vuistregels (die elkaar niet overlappen) beschikbaar. Deze is nu ingesteld op alle lags tot 1 jaar, en Baggelaar en van der Meulen stellen voor deze te verlagen, omdat de toets anders onderscheidend vermogen verliest en het Pastas ontwikkelteam stemt daarmee in. Voor de visuele inspectie zal de lengte van een jaar worden aangehouden, mogelijk kan in de toekomst visuele inspectie op twee tijdschalen worden ondersteund.

Een mogelijkheid om de orde van het ruismodel te bepalen is om de autocorrelation function (ACF) en de partial autocorrelation function (PACF) visueel te inspecteren. Er is echter geen theorie bekend om een PACF voor ongelijke tijdsintervallen te construeren. Bovendien past visuele inspectie van individuele reeksen niet goed bij de uitgangspunten van Pastas, waarin het mogelijk moet zijn om grote aantallen reeksen op dezelfde objectieve manier te behandelen. Ook is het aantal mogelijke ruismodellen in Pastas tot nu toe beperkt.

De Stoffer-Toloi toets kan weliswaar omgaan met missende waarden, maar voor praktijk-reeksen waarbij het tijdsinterval steeds kleiner wordt, is deze wellicht minder geschikt. Bij keuze van het kleinste tijdsinterval (bijvoorbeeld dag) zullen er zeer veel missende waarden zijn als een deel van de reeks dagelijks gemeten is en een deel van de reeks tweewekelijks. Het is maar de vraag of dan aan de voorwaarden wordt voldaan om een goede toets te doen.

De Durbin-Watson toets is volgens Baggelaar en van der Meulen alleen bedoeld voor de reeks van de restterm van een model dat geen autoregressieve term bevat. Dat is in het geval van Pastas niet het geval, dus deze toets kan beter niet gebruikt worden. Collenteur geeft aan dat de toets alleen voor autocorrelatie met een tijdsverschuiving van 1 tijdsinterval (1 lag) toetst en daarom een goede eerste indicatie kan geven of er sprake is van autocorrelatie.

toetsen op normaliteit

De normaliteitstoetsen zijn bij reeksen met veel meetwaarden (te) kritisch. Visueel worden de histogrammen vaak goed beoordeeld, terwijl de toets de nulhypothese van normaliteit verwerpt. Baggelaar verwacht dat de normaliteitstoetsen met ongelijke tijdsintervallen nog steeds werken, omdat deze uiteindelijk uit dezelfde verdeling zijn getrokken. Afwijkingen van normaliteit hebben volgens Baggelaar een relatief kleine invloed op de bandbreedte van het betrouwbaarheidsinterval.

Toetsen op heteroscedasticiteit

In Pastas wordt er niet getoetst op heteroscedasticiteit met behulp van een objectieve toetsingsgrootheid, maar met behulp van een visuele beoordeling. Heteroscedasticiteit is dat de variantie van de residuen afhangt van het niveau van de grondwaterstand. Vraag is nog of de breedte van het betrouwbaarheidsinterval van de grondwaterstand simulatie dan ook afhangt van de grondwaterstand.

Vergroting tijdsinterval

Constateringen

Het tijdsinterval is vergroot door de reeksen uit te dunnen. Er kan ook gekozen worden voor het middelen van de reekswaarden. Als grondwaterstanden uitgedund worden, maar de stress reeksen (bijv. regen en verdamping) dagelijkse waarden zijn, dan rekent Pastas met een tijdsinterval van één dag, wat middeling minder voor de hand liggend maakt. Voor de Overijssel-dataset komen meer modellen uit de dataset door de diagnostische toetsen naarmate het tijdsinterval tussen grondwaterstandsmetingen groter wordt. Er zijn echter wel bedenkingen bij de geschiktheid van de toetsen voor deze reeksen in combinatie met kleine tijdsintervallen. Verschillende tijdsintervallen kunnen ook leiden tot verschillende geschatte parameter waarden. Deze verschillen zijn alleen relatief in beeld gebracht aan de hand van de parameter recharge_A. Als de fit relatief goed is (EVP > 70%), veranderen de model parameters (Recharge_A) vrijwel niet met een groter tijdsinterval tussen de metingen.

Bij een klein tijdsinterval wordt soms een model gevonden met een slechte fit (gemeten in EVP), terwijl voor dezelfde reeks een betere fit behaald wordt met grotere tijdsintervallen. Mogelijk wordt dit veroorzaakt door de optimalisatie op de ruis, met een te beperkt ruismodel (zonder ruismodel wordt wel een goede fit verkregen). Bij deze modellen wordt vaak niet voldaan aan de statistische toetsen. Dit probleem dient niet verward te worden met het veronderstelde probleem dat is gemeld in Issue #235, wat gaat over het terugvinden van parameters met synthetisch gegenereerde ruis. Hiervan is inmiddels aangetoond dat het probleem werd veroorzaakt door verkeerde instellingen tijdens optimalisatie die afwijken van de Pastas defaults. Baggelaar en van der Meulen denken nog dat het wel hetzelfde probleem is en vermoeden dat de schattingsmethode verbeterd kan worden. Dit is echter geen onderdeel van dit project.

aanbevelingen

1. Het vergroten van het tijdsinterval van de reeksen resulteert vaker in een model waarvan de ruis aan de statistische toetsen voor witte ruis voldoet. Mogelijk gaat bij uitdunning of middeling informatie verloren. Waar de balans ligt hangt af van het doel van de tijdreeksanalyse en de responstijd van het systeem. Het verdient aanbeveling deze relatie te onderzoeken om afhankelijk van de situatie het juiste tijdsinterval te kunnen kiezen. Dit kan bijvoorbeeld gedaan worden door een set synthetische reeksen te ontwikkelen met verschillende relevante eigenschappen, bijvoorbeeld met een grondwatermodel. De vraag is wel hoe op een objectieve manier de ruis gegenereerd moet worden. Het lijkt weinig zinvol om hetzelfde ruismodel te gebruiken bij de synthese als bij de optimalisatie.
2. De problematiek van de slechte fit bij een klein tijdsinterval dient goed beschreven te worden, zodat deze niet verward wordt met andere zaken. Het maken van een GitHub discussie met een voorbeeldreeks ligt voor de hand. Mogelijke oplossingen (ander ruismodel, andere doelfunctie, andere response functie, apart ruis optimaliseren) zouden getoetst moeten worden aan synthetische en praktijk-reeksen.
3. Omdat Pastas zich juist op fysische relaties richt, heeft het ruismodel voor de gebruiker wellicht niet de prioriteit. Wellicht kan in het kader van de STOWA handleiding in een notebook een praktisch voorbeeld uitgewerkt worden om te demonstreren wat het nut en de noodzaak van het ruismodel zijn.

Tabel 2. Overzicht van aanbevelingen naar aanleiding van het overleg.

Aanbevelingen	**Wie?**	**Prio (1-3)**
Diagnostiek		
Pas het aantal lags voor de Portmanteau toetsen aan naar 15-20 lags.	Raoul	1
Gebruik de mediaan voor de berekening van de Runs toets.	Raoul	1
Literatuuronderzoek of toetsen voor heteroscedasticiteit met ongelijke tijdsintervallen beschikbaar zijn.		3
Literatuuronderzoek naar de beschikbaarheid van methoden voor het construeren van een PACF voor tijdreeksen met ongelijke tijdsintervallen.		3
Maak een synthetische dataset met residuen reeksen met ongelijke tijdsintervallen om het volgende uit te zoeken: - Kan de Runs toets de andere toetsen voor autocorrelatie vervangen? - Werken de toetsen voor normaliteit goed met ongelijke tijdsintervallen? - Werken de toetsen voor heteroscedasticiteit goed met ongelijke tijdsintervallen?		2
Zoek uit of de bandbreedte van de autocorrelatiefunctie voor ongelijke tijdsintervallen goed wordt berekend met synthetische tijdreeksen.		2
Vergroten tijdsinterval		
Maak een synthetische test dataset met afhankelijke en onafhankelijk tijdreeksen om te onderzoeken wat het effect is van het vergroten van het tijdsinterval op antwoorden op specifieke vraagstellingen.		1
Wat is het effect van verschillende methoden voor het vergroten van de tijdsintervallen (bv. uitdunnen of aggregeren) ?		2
Maak het voor de gebruiker van Pastas makkelijker om metingen uit te dunnen of te aggregeren		2
Algemeen		
Onderzoek of de schattingsmethode van Pastas kan worden verbeterd	Eit	2
Maak een ruismodel met seizoenseffect (SARIMA)		3
De problematiek van de slechte fit bij een klein tijdsinterval dient goed beschreven te worden, zodat deze niet verward wordt met andere zaken. Uitleg met een voorbeeldreeks ligt voor de hand. Mogelijke oplossingen (ander ruismodel, andere doelfunctie, apart de ruis optimaliseren) zouden getoetst moeten worden aan synthetische en praktijk-reeksen.	Artesia	1
Omdat Pastas zich juist op fysische relaties richt, heeft het ruismodel voor de gebruiker wellicht niet de prioriteit. Wellicht kan in het kader van de STOWA handleiding in een notebook een praktisch voorbeeld uitgewerkt worden om te demonstreren wat het nut en de noodzaak van het ruismodel is.	Artesia	2

Bijlage 1: Reactie op Artesia-studie

Invloed tijdstap bij tijdreeksanalyse

Paul Baggelaar

Eit van der Meulen

28 september 2020

Over rapport Invloed tijdstap bij tijdreeksanalyse (Artesia, 28 augustus 2020)

De studie omvat drie stappen:

1. Inbouwen toetsen ter verificatie van tijdreeksmodel – De verificatietoetsen die zijn gebruikt in de studie [Baggelaar en Van der Meulen, 2020] worden toegevoegd aan Pastas.

2. Nagaan wat het vergroten van de modeltijdstap betekent voor de toetsresultaten – Eerst wordt nagegaan of de toetsresultaten van [Baggelaar en van der Meulen, 2020] van de met Pastas afgeleide modellen van Overijsselse grondwaterstandreeksen op dagbasis worden gereproduceerd. Vervolgens wordt voor verschillende grotere tijdstappen en significantieniveaus nagegaan in welke mate de toetsresultaten veranderen.

3. Workshop, leidend tot voorstellen voor verder onderzoek.

Voorwoord

We zijn blij met dit rapport, omdat het illustreert dat de Pastas-ontwikkelaars daadwerkelijk belangrijke stappen zetten om het programma nog meer geschikt te maken als statistisch instrument. Dit zal de gebruiksmogelijkheden vergroten, wat Pastas nog waardevoller kan maken. Wij willen daar graag aan meewerken en hebben in het nu volgende getracht bruikbare adviezen te verstrekken over de onderwerpen die in het rapport worden behandeld. Hopelijk zijn het bruikbare adviezen. Maar blijf kritisch - ook over onze adviezen - en stel beslist vragen als jullie daarover twijfelen. Houd het er maar op dat ook wij feilbaar zijn, want deze materie heeft zeker complexe elementen.

ad 1) Inbouwen verificatietoetsen

Onze opmerkingen.

- De Ljung-Box-toets, ook wel aangeduid als de Portmanteau-toets op autocorrelatie, is bedoeld voor een equidistante reeks, waarin geen waarden ontbreken. Deze toets gaat er impliciet van uit dat de residuen afkomstig zijn uit een normale verdeling. Belangrijke toetsinstellingen zijn L, het aantal tijdsintervallen waarover de autocorrelatie wordt beschouwd en M, het aantal vrijheidsgraden, berekend als M = L – k, met k het aantal parameters van het tijdreeksmodel. Dit laatste is in de literatuur niet altijd eenduidig omschreven, want soms staat er dat k het aantal parameters van het ARMA- of ARIMA-model betreft, wat lijkt te suggereren dat het alleen om de parameters van het ruismodel zou gaan.[1] Als de residuen normaal verdeeld zijn zal de toetsingsgrootheid Q_M onder de nulhypothese ('er is géén autocorrelatie') afkomstig zijn uit een

χ^2-verdeling met M vrijheidsgraden.

De instelling van L heeft invloed op het onderscheidend vermogen van deze toets, maar de optimale instelling is moeilijk objectief vast te stellen en hangt ook af ook af van de doelstelling van de modelleur. Bij te grote L resteert er nog maar weinig onderscheidend vermogen en wordt de nulhypothese van geen autocorrelatie te vaak onterecht niet verworpen.

Artesia adviseert om bij modelleren op dagbasis voor zowel de Ljung-Box-toets als de Stoffer-Toloi-toets een L te hanteren van minstens 365 dagen, om zo ook een eventuele autocorrelatie op jaarbasis mee te kunnen nemen. Een dergelijk grote L kan echter leiden tot veel verlies van onderscheidend vermogen, al is bij deze studie voor elk van de beschouwde reeksen de nulhypothese van geen autocorrelatie verworpen.

Daarom wordt doorgaans voor de keuze van L een vuistregel gehanteerd. In de literatuur zijn bijvoorbeeld de volgende vuistregels te vinden:

- L = ln(T)
- L = minimum(20, T-1), waarin T het aantal reekswaarden (vuistregel die Matlab als standaard hanteert). Als het tijdreeksmodel veel parameters bevat kan dit bijvoorbeeld worden uitgebreid tot L = minimum(20+g, T-1), waarin g het aantal parameters.
- L = minimum(T/2-2,40) in de huidige versie van Python, maar dit wordt minimum(10, T/5) in de volgende versie. 'If lags is None, then the default maxlag is currently min((nobs // 2 - 2), 40). After 0.12 this will change to min(10, nobs // 5). The default number of lags changes if period is set.'

In de eerste versie van de verificatiemodule voor Pastas die we hebben aangeleverd, is gekozen voor L = 15. Bij modellering zonder ruismodel wordt het aantal vrijheidsgraden dan 15 en voor een modellering met een AR(1)-ruismodel wordt het 14. Dit klopt achteraf niet, want het aantal parameters van het empirische deel is dan nog niet meegenomen.

Mede gezien het soms geringe onderscheidend vermogen van de Ljung-Box-toets hanteren wij deze toets bij voorkeur in combinatie met een visuele beoordeling van de combinatie van een ACF (Autocorrelatiefunctie) en een PACF (Partiële autocorrelatiefunctie) van de te beoordelen reeks. Bij die visuele beoordeling gaan we dan vooral af op de eerste twee of drie geplotte coëfficiënten, want als er sprake is van een autocorrelatiestructuur zal die vooral naar voren komen bij dat eerste segment tijdsverschuivingen (lags). Verder gaan wij ook af op de ACF- en PACF-segmenten rond de tijdsverschuivingen van een jaar.

- De Stoffer-Toloi-toets is een aanpassing van de Portmanteau-toets, die overweg kan met een equidistante reeks waarin waarden ontbreken. Artesia vermoedt dat deze niet goed werkt door een verkeerde bepaling van het aantal vrijheidsgraden. Is de toets wellicht te checken door een aantal reeksen zowel te toetsen met de Pastas-uitvoering als met de Tijdreeksanalist-uitvoering?

- De Durbin-Watson-toets is alleen bedoeld voor de reeks van de restterm van een model dat geen autoregressieve term bevat. Deze kan dus alleen worden toegepast op de restterm van een model dat geen autoregressieve parameter omvat, zoals f van het ruismodel of d van het transfermodel (deze laatste zoals bij Box-Jenkinsmodellering).

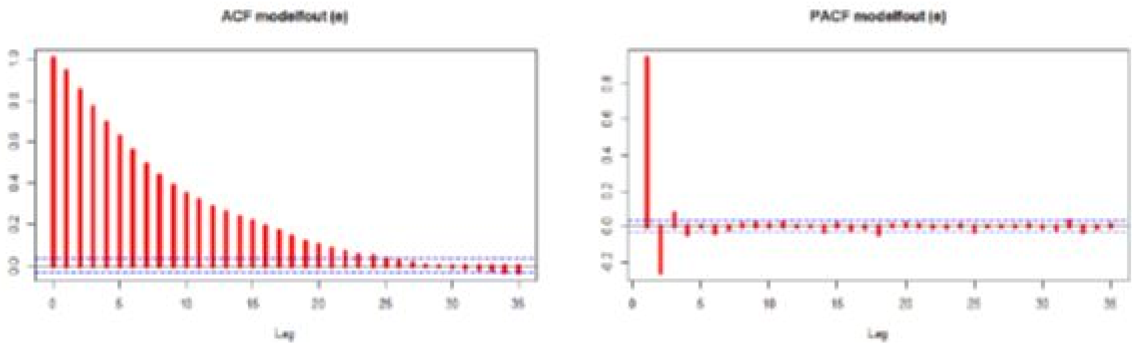
- De runs-toets is toegepast door runs ten opzichte van gemiddelde te bepalen. Maar omdat het een verdelingsvrije toets betreft is de mediaan daarvoor een betere keuze dan het gemiddelde (zie ook Wikipedia Wald-Wolfowitz runs test). De uitkomsten van de runs-toets toegepast in Pastas en onze verificatiemodule verschillen daardoor.

- Hoe wordt bij toepassing van ps.stats.acf of ml.plots.diagnostics op een tijdreeks met ongelijke tijdstappen het 95%-betrouwbaarheidsinterval van een geschatte autocorrelatiecoëfficiënt geschat? Is die methode theoretisch en praktisch (met simulaties) geschikt gebleken?

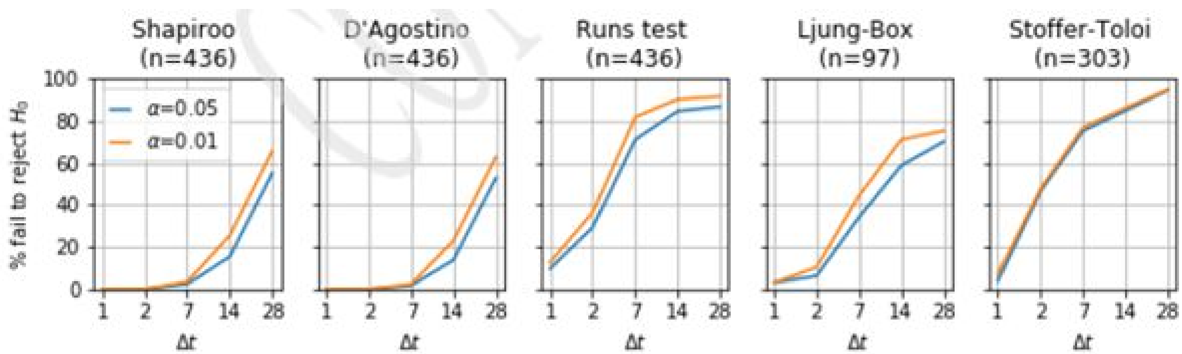
In onze verificatiemodule zijn de acf- en pacf-functie van Python toegepast, waarbij reeksen equidistant worden gemaakt door tussenvoegen van missende waarden.

- Waarom wordt de ACF (Autocorrelatiefunctie) niet getoond in combinatie met een PACF (Partiële autocorrelatiefunctie)? Deze combinatie is immers nodig om de juiste AR- en MA-ordes van het ruismodel te kunnen identificeren. Met alleen een ACF lukt dat vaak niet.

Ter illustratie onderstaand voorbeeld. Uit de ACF kan hooguit worden afgeleid dat er sprake is van autocorrelatie, maar uit de aanvullende informatie van de PACF blijkt dat er ook sprake is van een moving-average-structuur. Het betreft hier dan ook een ARMA(1,1)-model.



ad 2) Invloed vergroten tijdstap



- Artesia heeft bij vergroten van de tijdstap gekozen voor uitdunning van de tijdreeks. In principe zouden de resttermen van het model dan bij elke tijdstap uit hetzelfde type kansverdeling moeten komen, want er vindt geen middeling plaats. Maar volgens de bovenstaande twee linkerplots wordt naarmate de tijdstap groter is vaker voldaan aan normaliteit. Dit kan er op wijzen dat naarmate het aantal reekswaarden toeneemt kleine afwijkingen van normaliteit makkelijker worden gedetecteerd.

- Bij vergroten van de tijdstap kan worden gekozen voor uitdunning van de tijdreeks, maar een andere optie is aggregeren, zoals door het gemiddelde of de mediaan te nemen van alle meetwaarden in een tijdstap. Dit heeft als voordelen dat: i) eerder zal worden voldaan aan de randvoorwaarde van normaliteit en ii) er meer meetinformatie wordt gebruikt bij het modelleren.

- De normaliteitstoetsen verwerpen bij een tijdstap van 28 dagen voor circa 40% van de reeksen de nulhypothese van normaliteit van het modelresidu. Komt dit door het grote aantal meetwaarden, of is er echt sprake van een scheve kansverdeling? Dit kan beoordeeld worden door histogrammen van de betreffende restterm (ruis/residu) te plotten.

Aangezien de normaliteitstoetsen bij een grote reekslengte (veel meetwaarden) de nulhypothese al verwerpen bij praktisch gezien geringe afwijkingen van normaliteit, is aan te raden om bij het verifiëren nooit alleen af te gaan op de toets, maar ook op visuele beoordeling van het histogram.

Over de afsluitende vragen in het Artesia-rapport

- Hoe goed werken toetsen voor normaliteit bij ongelijke tijdstappen?

Dit kan worden vastgesteld met simulaties van kunstmatige reeksen. Vermoedelijk maken ongelijke tijdstappen weinig uit, omdat die niets veranderen aan het soort kansverdeling waar de meetwaarden uit afkomstig zijn.

- Welke toetsen zouden er beschikbaar zijn voor het toetsen van heteroscedasticiteit en werken deze ook voor ongelijke tijdstappen? (Breush-Pagan, White test?)

Dit kan worden vastgesteld met literatuuronderzoek, zonodig aangevuld met simulaties van kunstmatige reeksen.

- Kan het nieuwe ARMA(1,1) model helpen?

Slechts gedeeltelijk. In zijn huidige vorm kan het namelijk slechts overweg met één teken voor q, de 1e-orde-moving-average parameter, terwijl die parameter in de praktijk zowel positieve als negatieve waarden kan aannemen. Verder kunnen er ook diverse complexere ARMA-modellen nodig zijn om witte ruis te kunnen bewerkstelligen, zoals ARMA(1,2), ARMA(1,3), etcetera.

- Wat is het effect van het wel of niet vergroten van de tijdstap op de toepassing van een model, zoals het schatten van de invloed van pompen, of van oppervlaktewater, of het voorspellen?

Dit zal afhangen van meerdere factoren, waaronder de dynamiek van het onderzochte systeem. Het kan in beeld worden gebracht met simulaties van kunstmatige reeksen.

- Hoe kan het toch dat neerslag-afvoer-modellen vrijwel nooit diagnostisch worden getoetst en er toch zoveel mee gewerkt wordt? Staren we ons niet dood op een bepaalde methode en zijn er geen alternatieven die beter werken?

We kunnen dit ook omdraaien. Het verbaast ons dat dergelijke belangrijke modellen blijkbaar zo weinig kritisch onder de loep worden genomen en dat er hooguit alleen wordt beoordeeld op de pasvorm. Als er zuivere voorspellingen mee moeten worden gedaan, met tevens een indicatie van het onzekerheidsinterval, is het onontbeerlijk dat het model voldoet aan bepaalde randvoorwaarden, zoals gebleken bij verificatie.

- Hoe (goed) werken Bayesiaanse methoden in plaats van frequentistische methoden?

Deze vraag is niet zo eenvoudig te beantwoorden, omdat het dispuut hierover niet is beslecht in het voordeel van één van de kampen. Mogelijk zal het antwoord ook nogal verschillen naar gelang het een aanhanger van de ene stroming of de andere betreft.

Over de html Diagnostic checking of many Pastas models

- The p-value of a test represents the probability that the Null-hypothesis is rejected. Dit klopt niet helemaal.

Betere formulering (uit Wikipedia): In null hypothesis significance testing, the p-value is the probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis is correct. A very small p-value means that such an extreme observed outcome would be very unlikely under the null hypothesis.

- Over de relatie tussen het meetinterval en de geschatte evenwichtsrelatie (en standaardfout) van de grondwaterstand met de recharge. In deze exercitie komt een positieve relatie naar voren tussen het meetinterval en zowel de evenwichtsrelatie als zijn relatieve standaardfout. Kan dit komen door de niet-optimale schattingsmethode (zie ook Mogelijkheden voor vervolg, hieronder)?

- De exercitie met de 436 Overijsselse tijdreeksen is geschikt om te kunnen beoordelen wat de relatie is tussen het meetinterval van de grondwaterstand en het kunnen voldoen aan de randvoorwaarden van witte ruis. Deze is ook geschikt om de relatie te kunnen beoordelen tussen het meetinterval en de geschatte evenwichtsrelatie van de grondwaterstand met de recharge. Maar het geeft nog geen inzicht in de zuiverheid van de schatter van die evenwichtsrelatie en van de dekkingsgraad van zijn bijbehorend 95%-betrouwbaarheidsinterval. Dat kan alleen met een simulatiestudie met kunstmatige reeksen en een daarbij ingestelde evenwichtsrelatie met de recharge.

Mogelijkheden voor vervolg

Deze studie van Artesia is een goede stap naar verbreding van de mogelijkheden van Pastas. Enkele van de bij deze studie opgemerkte onduidelijkheden kunnen wellicht worden opgehelderd met simulaties van kunstmatige reeksen.

Voor wat betreft prioriteiten van aanvullend onderzoek is het wellicht handig na te gaan of de schattingsmethode van Pastas verbeterd kan worden. Deze is namelijk nog niet optimaal. Dit is bijvoorbeeld in ons onderzoek naar voren gekomen en ook Thomas heeft het aangetoond in zijn issue#235 (zie https://github.com/Pastas/Pastas/issues/235). Daarin zijn grondwaterstandreeksen gesimuleerd op basis van de Pastas-Gamma-functie, met daarop gesuperponeerd ruis op basis van een ARMA(1,1)-model. Maar de schattingsmethode van Pastas kon de modelparameters van de Gamma-functie en van het ruismodel niet goed terugschatten. De onzuiverheid van de parametersschattingen bleek groter te zijn bij hogere waarden van de AR(1)-parameter, waardoor de onzuiverheid van de schatting van de evenwichtsrelatie kan oplopen tot meer dan 50%.

---

[1] In ons rapport Naar betere tijdreeksanalyse met Pastas (28 april 2020) staat dit dan ook niet goed omschreven.
