PIRFICT-PASTAS

Hoe kan het dat uitkomsten zónder en mét een ruismodel zo van elkaar kunnen verschillen?

Eit van der Meulen (AMO)

16 maart 2023

Wezenlijke verschillen Pirfict- en BJ-toepassing

- Niet-equidistante modellering versus equidistante modellering
- AR(1)- en ARMA(1,1)-ruismodel versus een uitgebreid ARMA(*,*)-ruismodel
- ❖ Verificatie op basis van ‘goodness-of-fit’ (EVP en R^2) en plausibiliteit versus statistische modelvoorwaarden en bruikbaarheid
- ❖ Klasse van voorgedefinieerde responsfuncties versus modelidentificatie met een grote set kandidaatmodellen
- Hydrologisch plausibiliteit versus gekwantificeerde onzekerheid
- Modellering bij voorkeur op dagbasis versus modellering op een tijdseenheid passend bij het proces
- ❖ Een indirecte schattingsroutine versus een directe schattingsroutine
- Toepassen van selectiecriteria voor bewerkstelligen plausibele modellen versus bruikbare modellen

Toepassing Pirfict-methode
‘Goodness-of-fit’ en plausibel

- Goede pasvorm/modelfit
- Fysische modellering (voorgedefinieerd, expert)
- EVP (‘verklaarde variantie’) of R^2 beoordelingsmaten voor de modelkeuze (70% grens)
- Beoordeling van het deterministische modeldeel
- Ruismodel alleen gewenst bij uitkomsten met schatting betrouwbaarheidsintervallen
- Hoogfrequente modellering veelal op dagbasis

Toepassing BJ-methode
‘Goodness-of-use’, bruikbaar

- Modelidentificatie, eerste modelkeuze uit klasse van BJ-modellen
- Modelverificatie, voldoen aan statistische voorwaarden, zoals residu/‘noise’ (a_t) is normaal verdeelde *witte ruis* en *minimaal*
- Geen correlatie tussen deterministische modeldeel en de ruis/’residuals’
- Kiezen uit de klasse van bruikbare modellen
- Onzekerheid kwantificeren
- Op basis van uitgebreid ruismodel en verschillende modelleringsfrequenties

Discussiepunten

- EVP, de verklaarde variantie
- Modelidentificatie
- Schattingsroutine PASTAS mét ruismodel
- Consequenties

Is EVP de verklaarde variantie?

Beschouw het model : $Z(t) = \mathrm{detmodel}(\mathrm{t}|\vartheta) + N(t)$

Verklaarde variantie : $\dfrac{var\left(\mathrm{detmodel}(\mathrm{t}|\vartheta)\right)}{var\left(Z(t)\right)}.100\%$

De EVP is bij PIRFICT gedefinieerd als: $EVP = \dfrac{var\left(Z(t)\right) - var\left(N(t)\right)}{var\left(Z(t)\right)}.100\%$

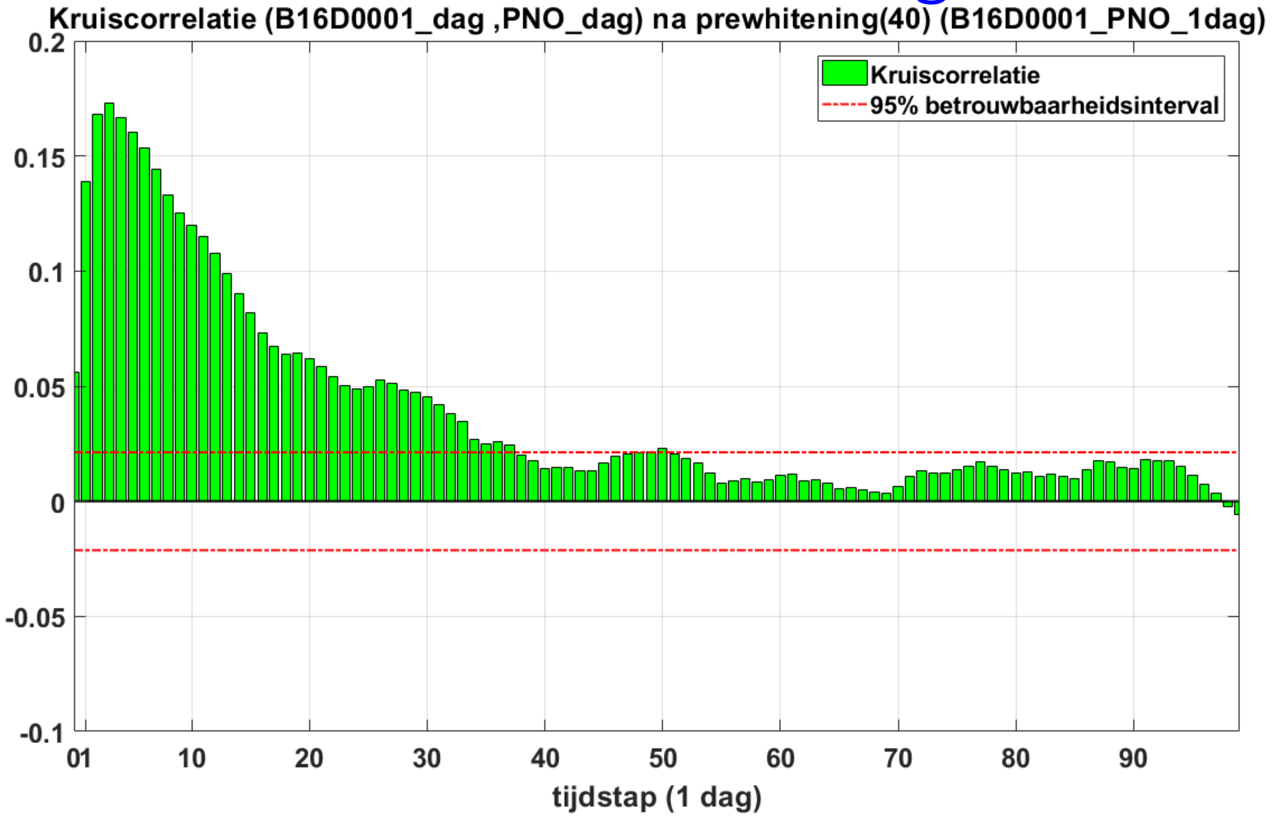
Nu geldt: $var\left(Z(t)\right) = var\left(\mathrm{detmodel}(\mathrm{t}|\vartheta)\right) + var\left(N(t)\right) + 2.cov\left(\mathrm{detmodel}(\mathrm{t}|\vartheta), N(t)\right)$

Dan geldt: Verklaarde variantie = $EVP - 2.\dfrac{cov\left(\mathrm{detmodel}(\mathrm{t}|\vartheta), N(t)\right)}{var\left(Z(t)\right)} * 100\%$

EVP is alleen gelijk aan de verklaarde variantie
als *N(t)* **niet** is gecorreleerd aan deterministische modeldeel,
maar dergelijke correlaties treden vaak juist wél op!
Ook bij R^2 wordt verondersteld dat deze correlatie nul is.

Modelidentificatie ('laat de data spreken')

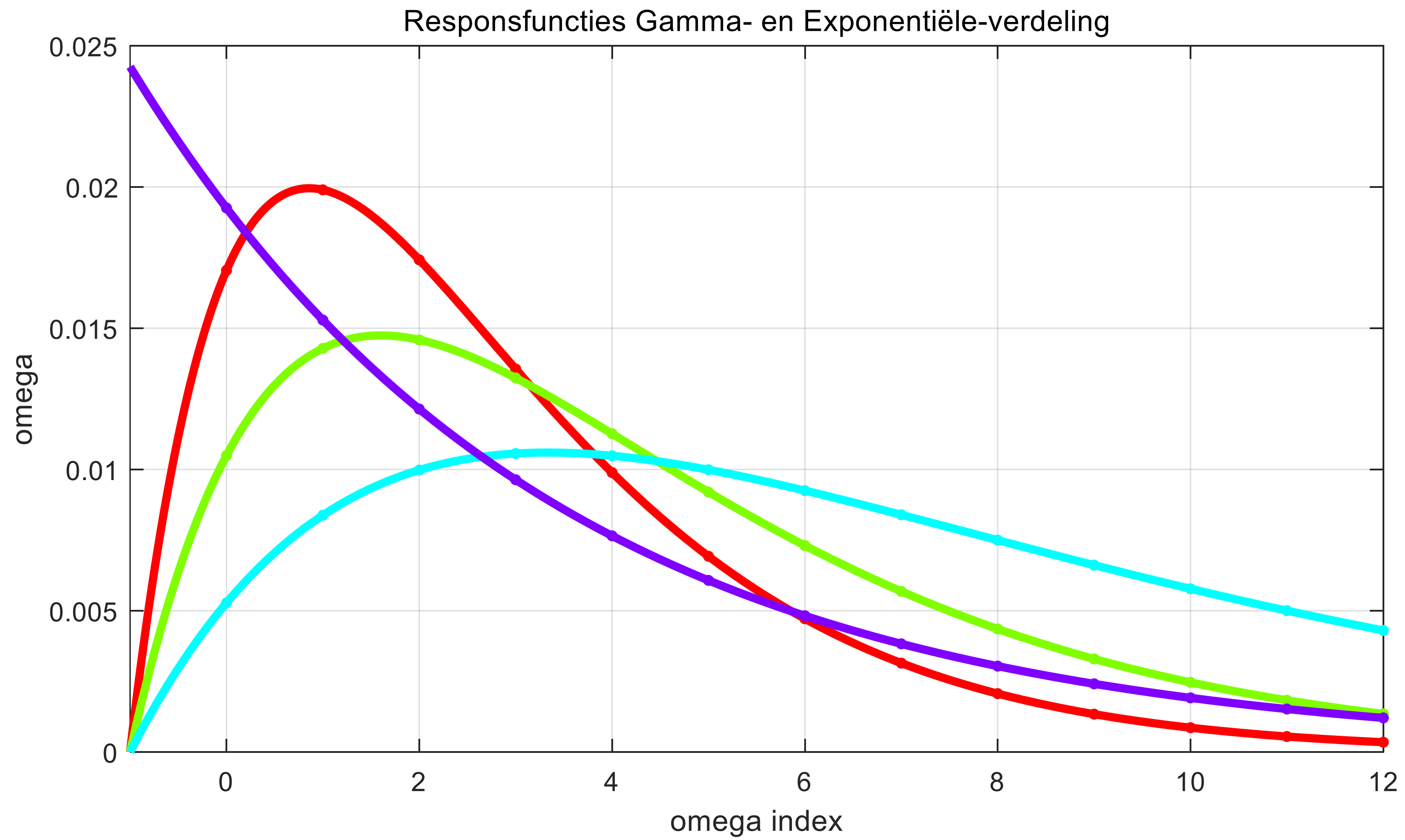
Gewit kruiscorrelogram



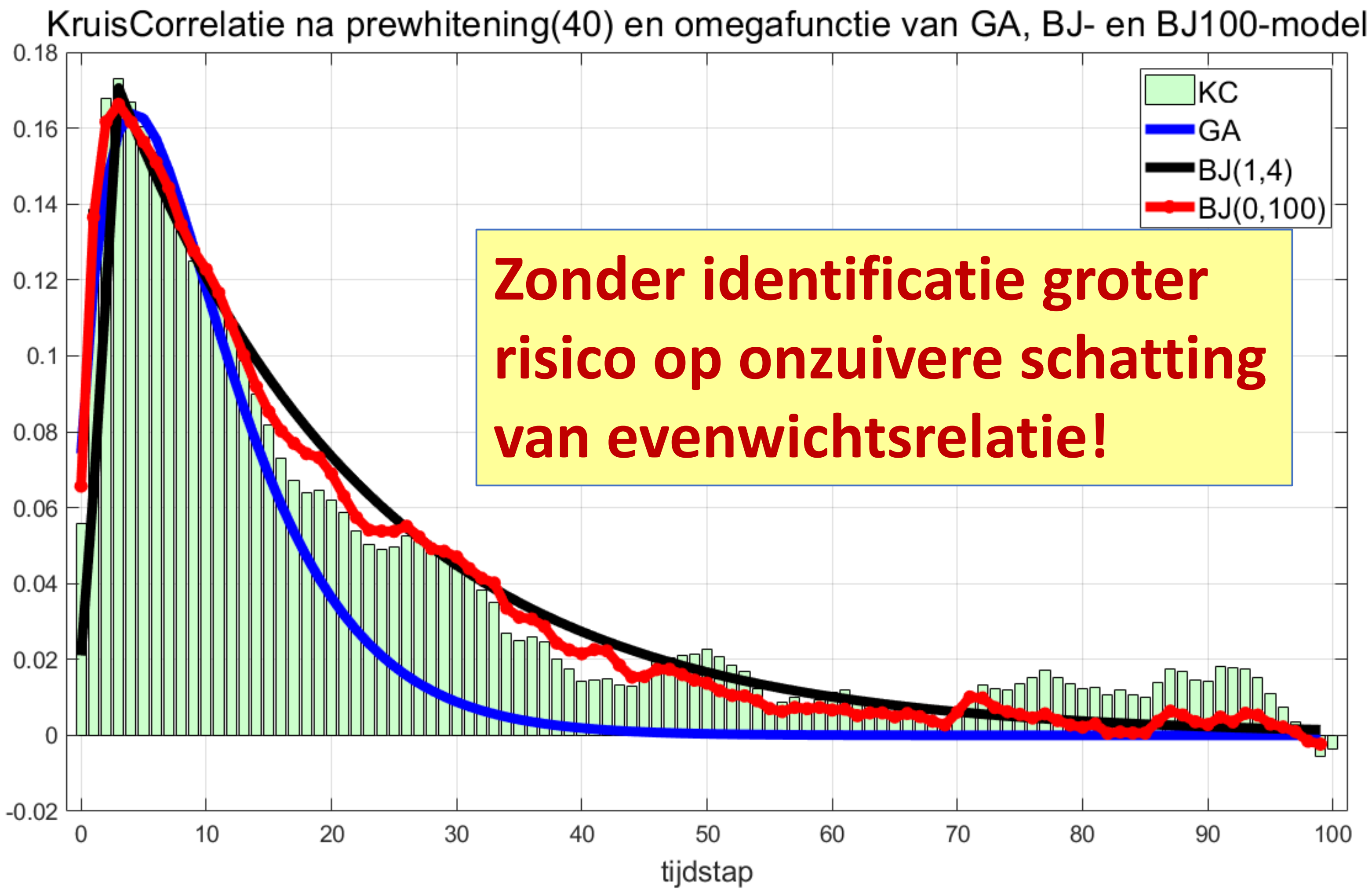
1. $C_t = \delta_1 C_{t-1} + \omega_0 X_t - \omega_1 X_{t-1} - \omega_2 X_{t-2} - \omega_3 X_{t-3}$

2. $C_t = \omega_0 X_t - \omega_1 X_{t-1} - \omega_2 X_{t-2} - \ldots - \omega_{100} X_{t-100}$

Responsfunctie



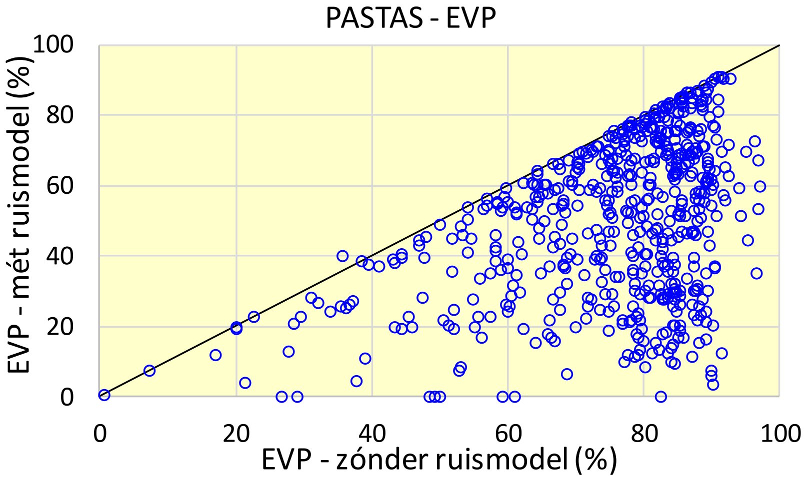
Vergelijking geschatte responsen

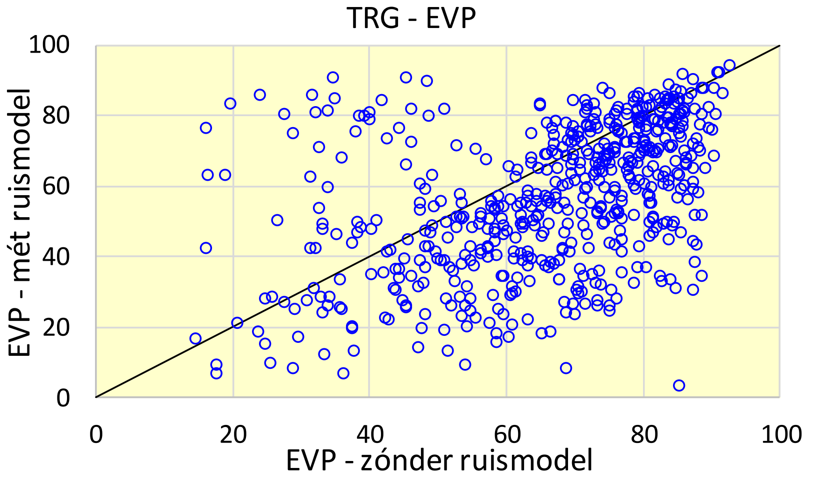


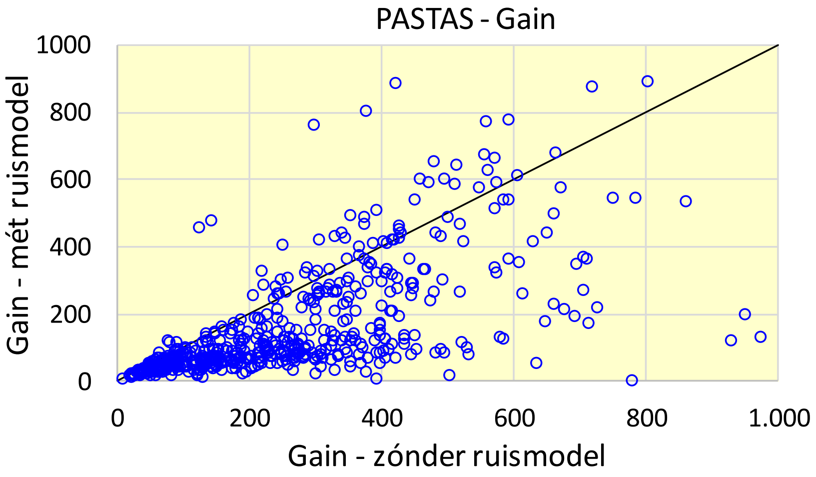
Schattingsroutine en optimalisatiecriterium

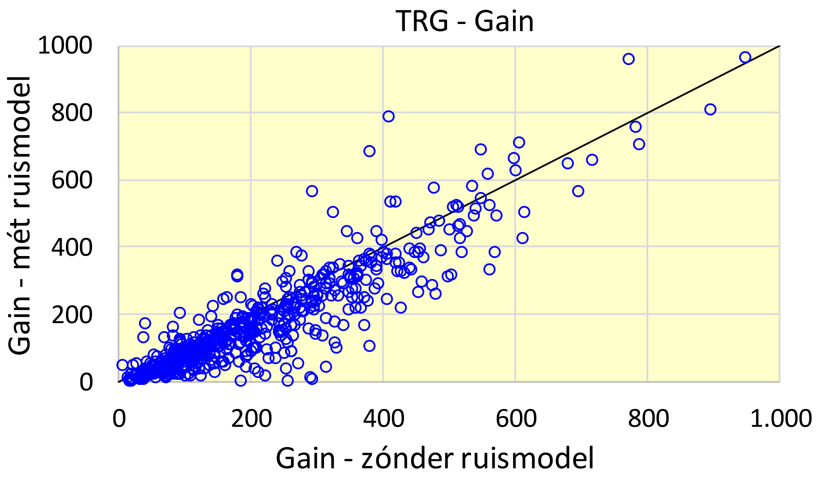
- Een goede pasvorm met plausibele uitkomst. Is dat een goed criterium?
- Voldoet dit model aan alle wiskundig/statistische voorwaarden van een goed model?
- Functioneert het ruismodel ook zodanig dat er geen correlatie is tussen het deterministische model en de ruis/'residuals' (N(t))en de residuen/'noise' (a(t))?
- Is er sprake van zuiverheid? Zijn uitspraken met kwantificeerbare betrouwbaarheden mogelijk?
- Een ruismodel levert schade op voor de EVP. Hoe is dit mogelijk?
- Worden modelparameters wel optimaal geschat?

Voorgeschiedenis





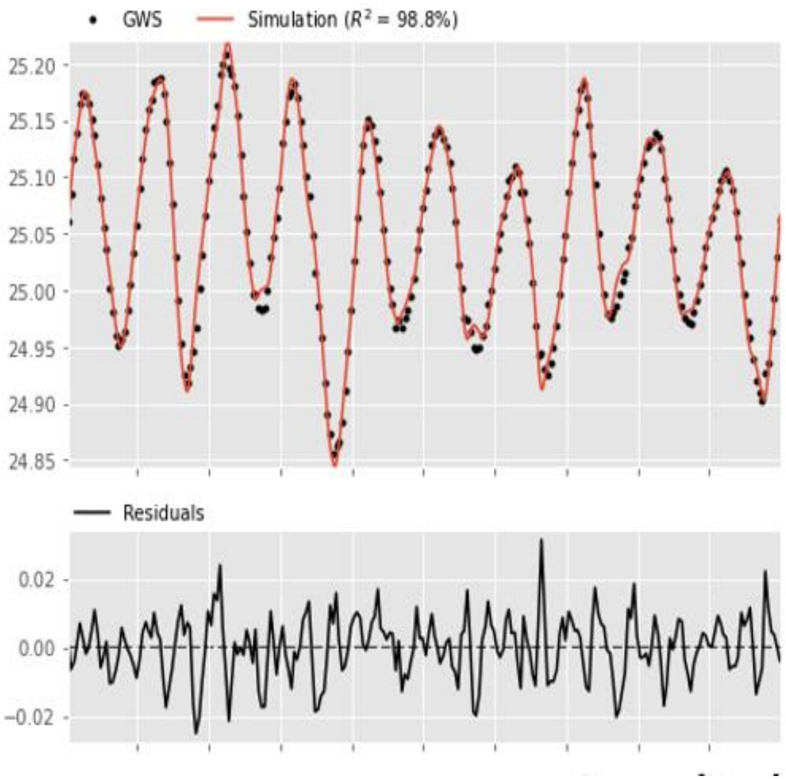




Test-notebook ‘Hoezo plausibel?’

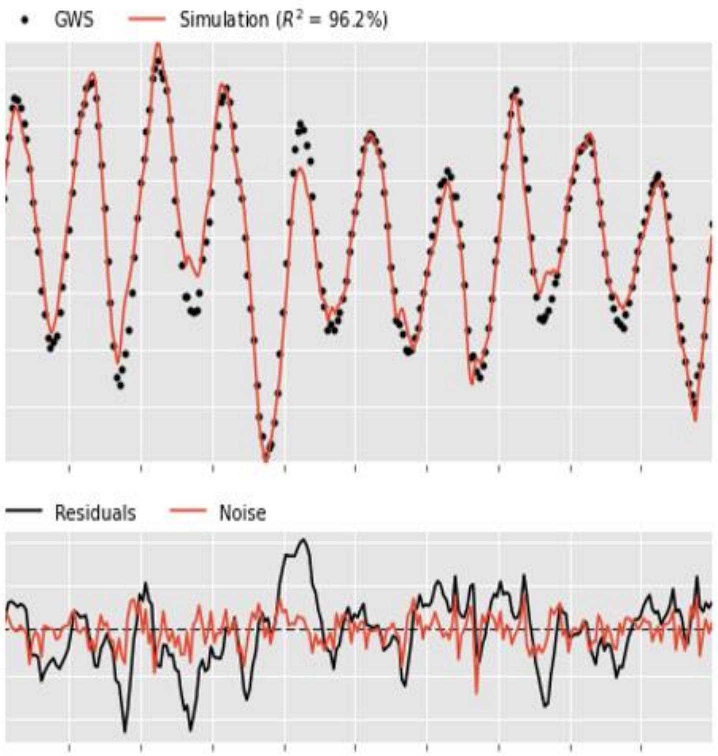
- De PASTAS-modellering zónder en mét ruismodel
- De evenwichtsrelatie bij extra ruis
- De EVP en de werkelijke verklaarde variantie
- De empirische dekkingsgraad
- De zuiverheid
- De correlatie tussen deterministische modeldeel de ruis/’residuals’ en de residuen/’noise’
- Vergelijking van de werkelijke met de geschatte transferfunctie

Zonder en met ruismodel



Model Parameters

name	optimal	stderr
recharge_A	105.93	1.61%
recharge_n	2.90	2.71%
recharge_a	37.31	3.39%
constant_d	25.00	0.00%



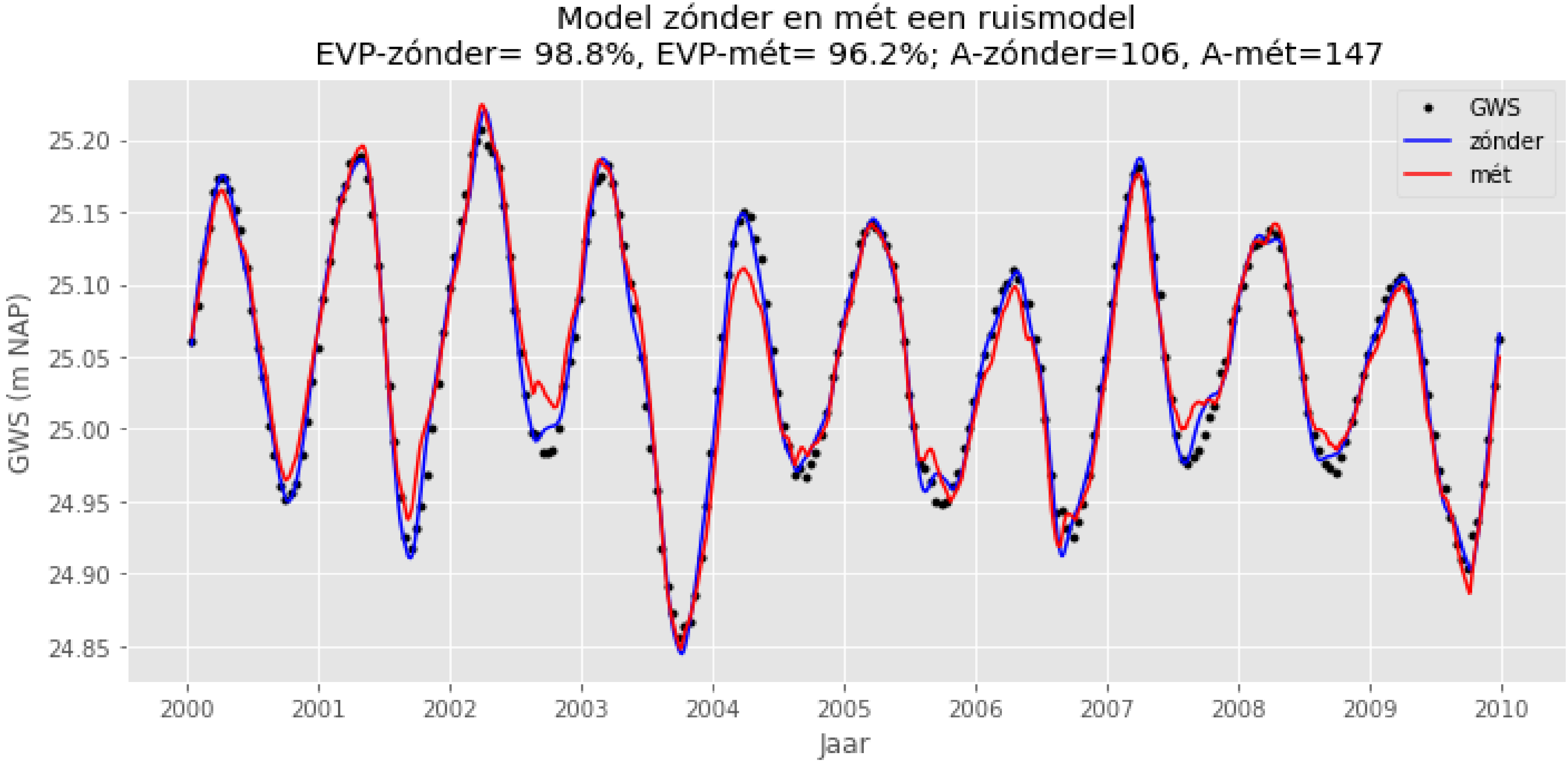
Model Parameters

name	optimal	stderr
recharge_A	146.66	5.32%
recharge_n	1.89	2.55%
recharge_a	74.37	5.99%
constant_d	24.98	0.02%
noise_alpha	165.82	35.37%

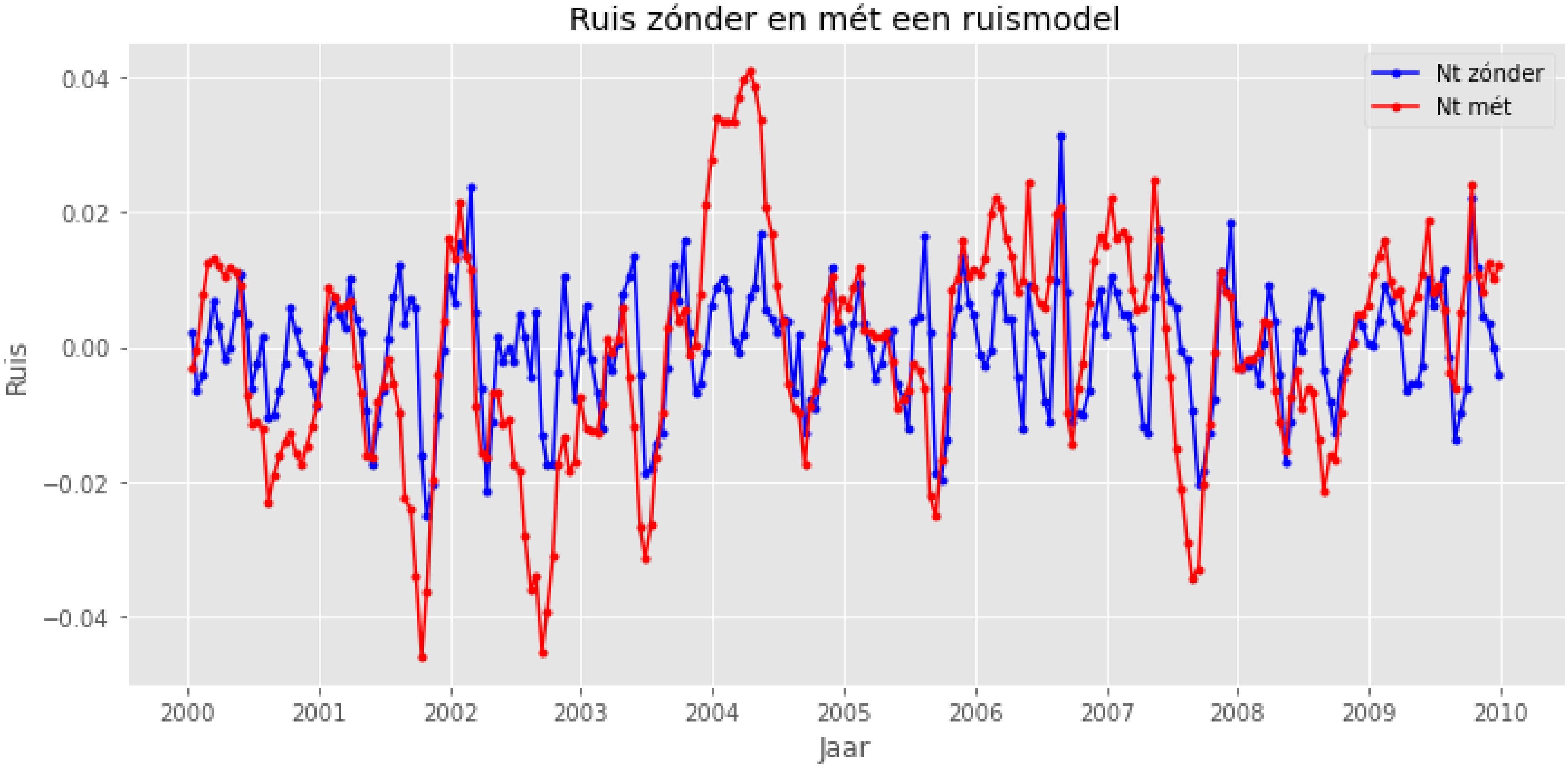
Evenwichtsrelatie zónder ruismodel = 106

Evenwichtsrelatie mét ruismodel = 147

Zonder en met ruismodel

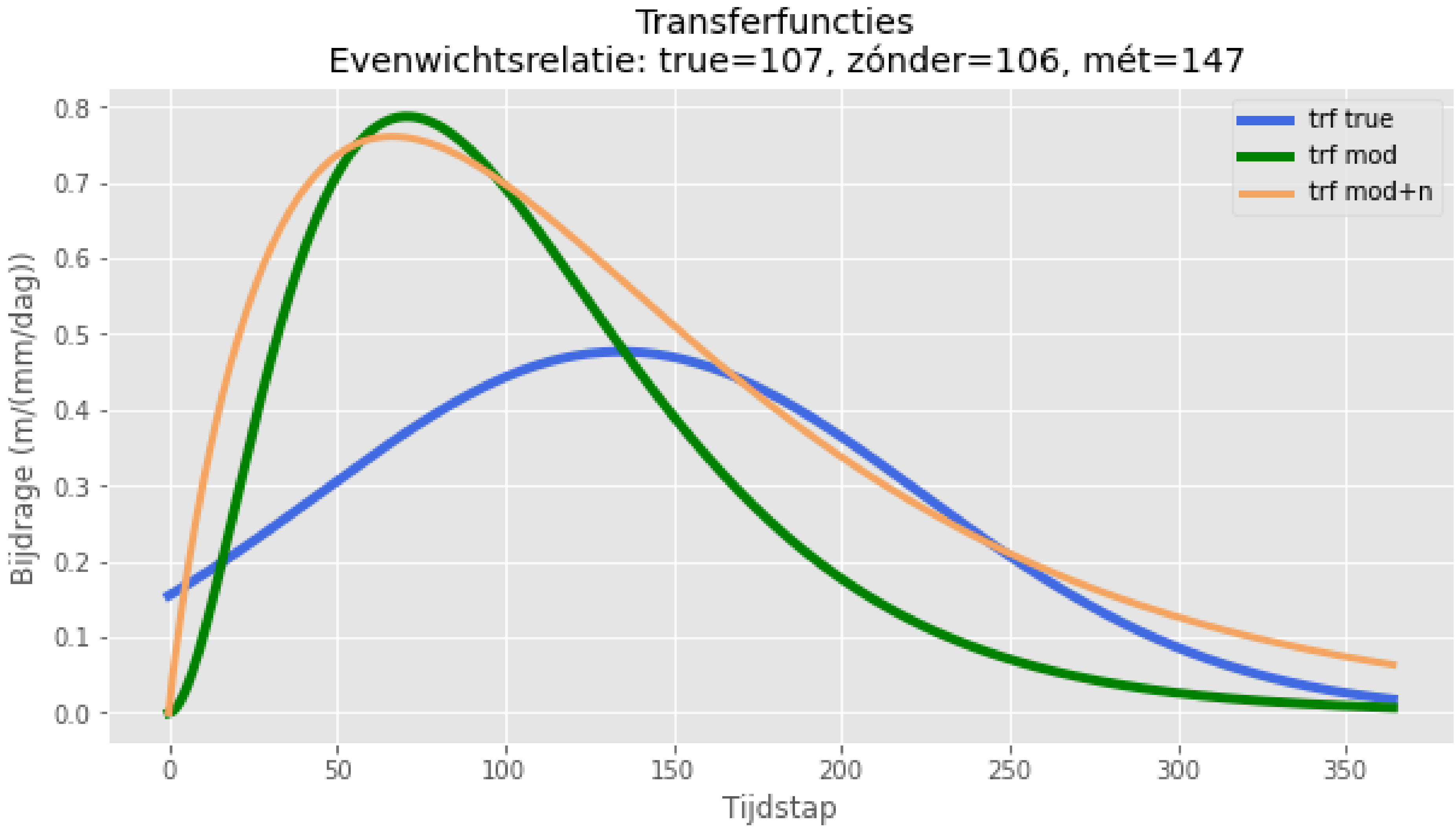


Ruis zónder en mét een ruismodel

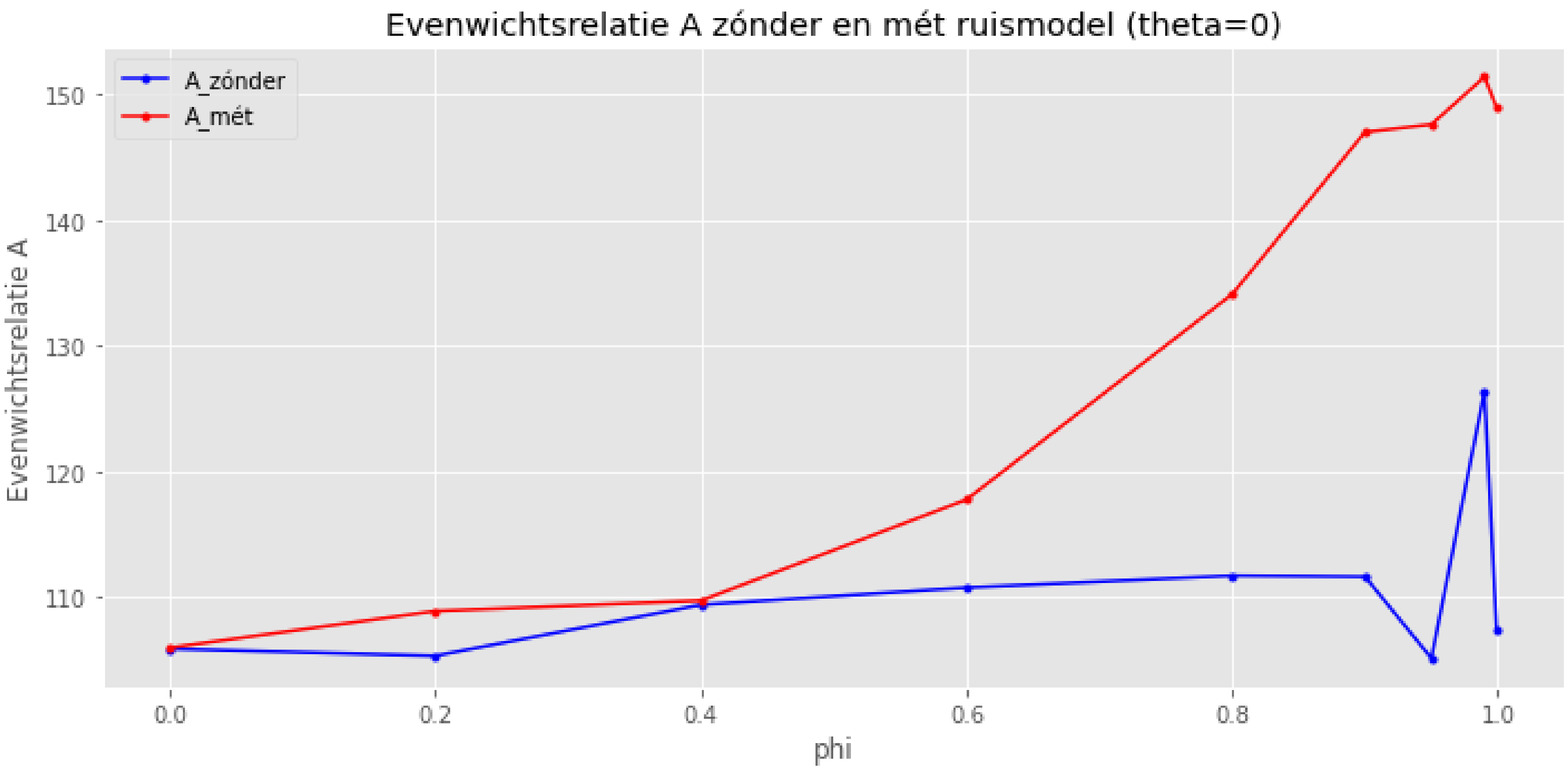


Het ruismodel voegt extra correlatie toe

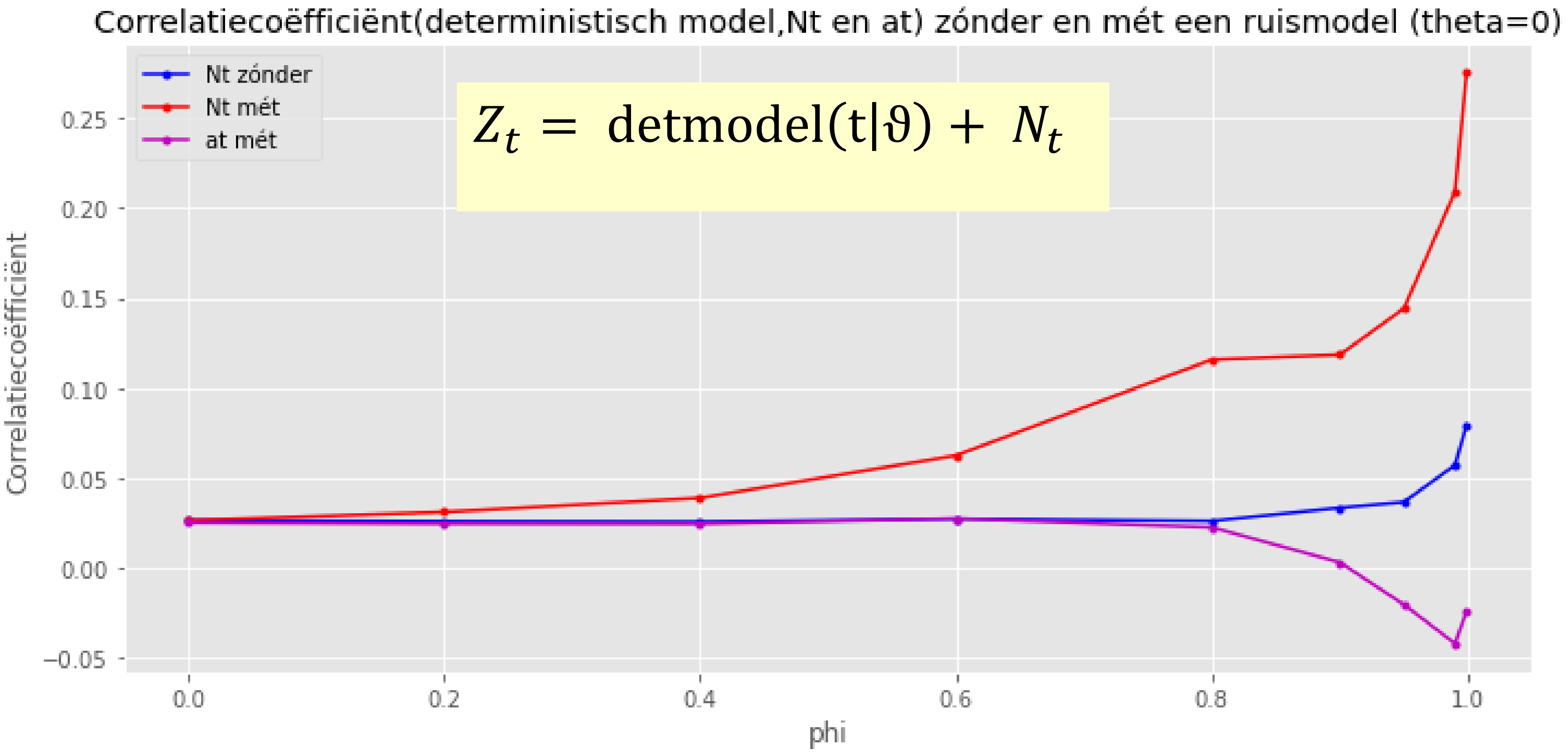
Transferfuncties



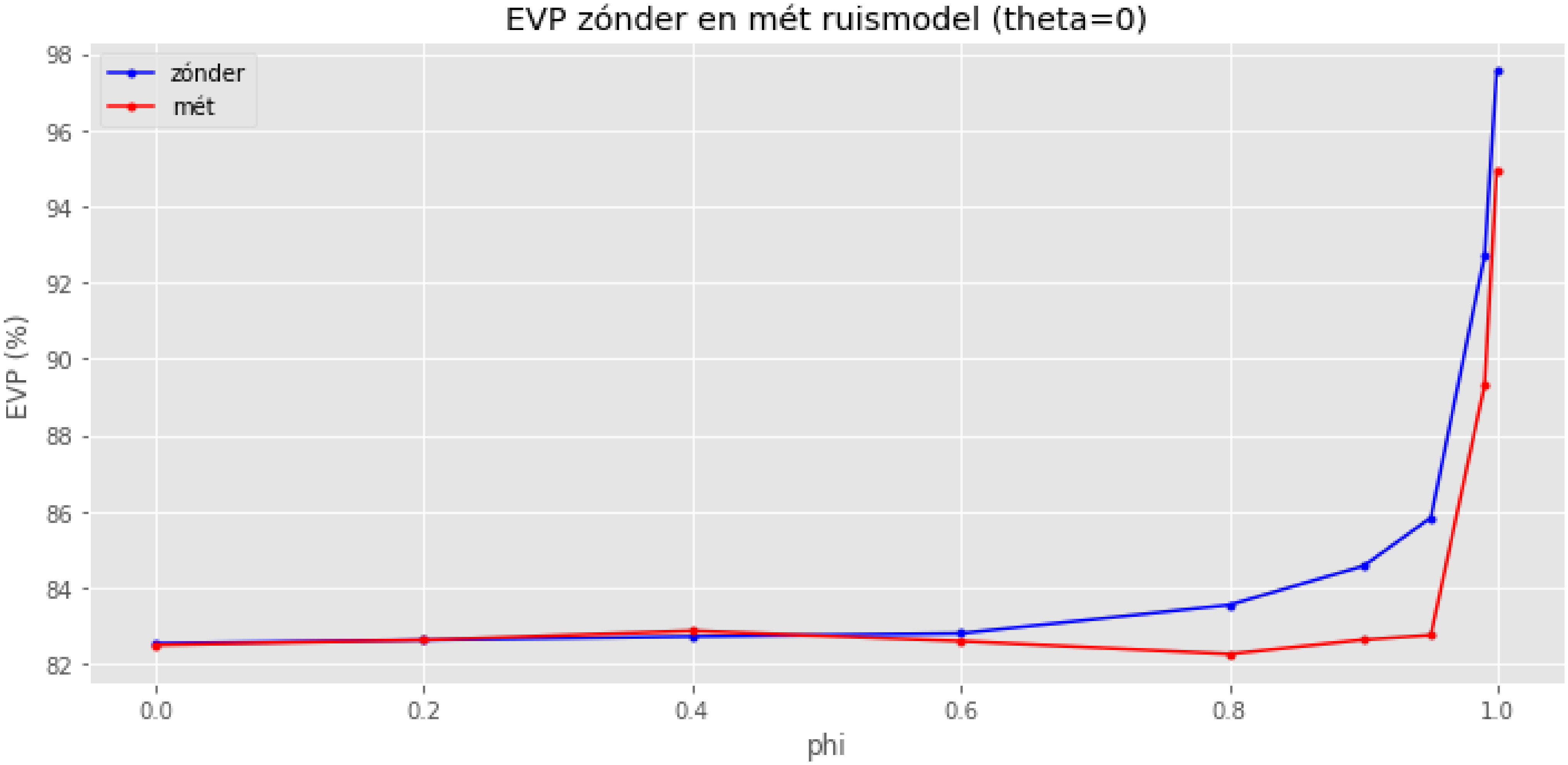
Evenwichtsrelatie als functie van toegevoegde ruis



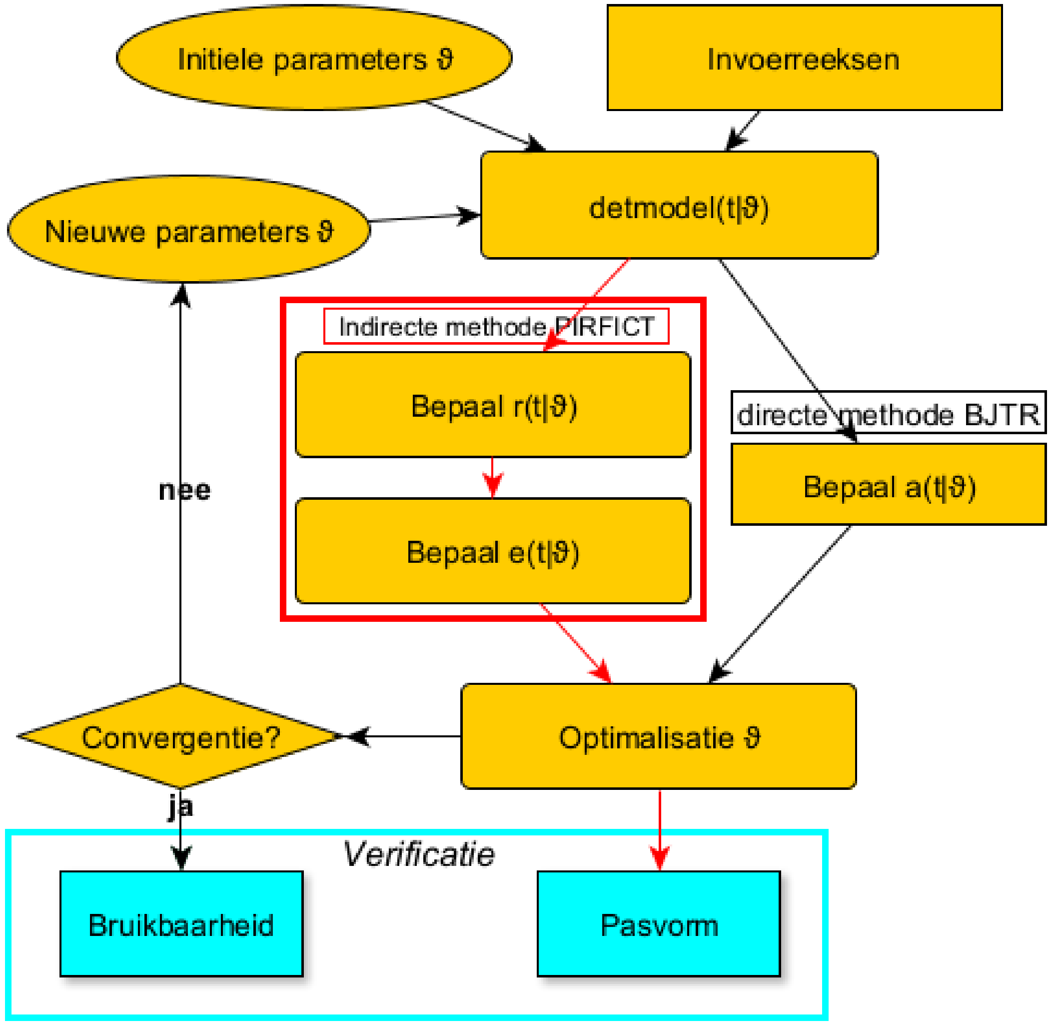
Correlatiecoëfficiënt detmodel en ruis



EVP zónder en mét ruismodel



Stroomschema parameteroptimalisatie PASTAS en Tijdreeksanalist



Algemeen Pirfict:

$r(t) = z(t) - \mathrm{detmodel}(\mathrm{t}|\vartheta)$

$e(t) = r(t) - \phi(\Delta t).r(t - \Delta t)$

$met\ \phi(\Delta t) = exp\left(-\frac{\Delta t}{\alpha}\right)$

Vergelijkbaar BJTR:

$Z_t = \mathrm{detmodel}(\mathrm{t}|\vartheta) + N_t$

$N_t = \phi.N_{t-1} + a_t$

Of herschreven:

$Z_t = \mathrm{detmodel}(\mathrm{t}|\vartheta) + f(a_t, \phi)$

Oplossing AR(1)-ruismodel

Beschouw het ruismodel:

$$N_t = \phi . N_{t-1} + a_t$$

$$N_t = \phi . BN_t + a_t$$

$$N_t = \frac{1}{(1-\phi B)} a_t, \text{ rekenkundige reeks met rede } \phi B$$

$$N_t = \sum_{k=0}^{\infty} \phi^k . B^{-k} a_t$$

$$N_t = a_t + \phi . a_{t-1} + \phi^2 . a_{t-2} + \cdots$$

Beschouw het model

$$Z_t = \text{detmodel}(\mathrm{t}|\vartheta) + a_t + \phi . a_{t-1} + \phi^2 . a_{t-2} + \cdots$$

Eenduidig kunnen de modelparameters ϑ en ϕ geoptimaliseerd worden.

De afleiding gaat uit van equidistante en gelijk verdeelde a_t

Ruismodel

- Een ruismodel dient gezond en nuttig te zijn
- Bij een AR(1)-ruismodel op dagbasis en een $\phi>0.9/\alpha>10$, kies voor een uitgebreid ruismodel of een lagere modelleringsfrequentie
- Geeft betere pasvorm tijdreeksmodel en betere voorspellingen (het rekenen met onzekerheid vermindert onzekerheid)
- Stelt in staat kwantitatieve uitspraken over onzekerheden te doen (zoals bij risico-analyses)

Meegeven onzekerheid geeft **altijd** meerwaarde.

Bezinning

- Het nalaten van modelidentificatie levert schade/onzuiverheid op. Ga bij voorbaat er *niet* vanuit dat de werkelijkheid past bij het huidige arsenaal aan verdelingsfuncties van PASTAS
- Parameterschattingen mét het PASTAS-ruismodel zijn onzuiver
- Geef geen onzekerheden als er niet aan de statistische voorwaarden wordt voldaan, geeft onzuivere schattingen van die onzekerheden
- De EVP en R^2 zijn geen bruikbare modelleringsmaten aangezien ze uitgaan van geen correlatie van de deterministische modeldeel met de ‘residuals’/ruis
- Wat zijn de consequenties van onzuivere parameterschattingen?
- Is kwantificatie van onzekerheid niet van belang?

PASTAS-tijdreeksmodellen zijn (nog) niet geschikt voor het doen van goed onderbouwde kwantificeerbare uitspraken.

Is discrete tijdreeksanalyse geen oplossing?

- Discrete tijdreeksanalyse op basis van equidistantie
- Uitgebreid ruismodel wordt mogelijk
- Statistische toetsen zijn beschikbaar (*modelverificatie*)
- Kwantificatie van onzekerheid, risicoanalyse is mogelijk
- Groot arsenaal aan kandidaatmodellen (*modelidentificatie*)
- Optimale parameterschatting
- Wetenschappelijk verantwoord
- Software is beschikbaar, kan ook vertaald worden Python