Goodness-of-fit of *Fitness-for-use*?

Pasvorm of geldigheid?
Gebruiksgemak of bruikbaarheid?

Presentatie voor Werkgroep Tijdreeksanalyse

16 juni 2022

Paul Baggelaar
PB Icastat

Martin Knotters
WUR

Eit van der Meulen
AMO

Probleemstelling

Voor veel tijdreeksanalisten is

goodness-of-fit van deterministische component

belangrijkste kwaliteitscriterium

Maar wat zegt die over fitness-for-use?

Box-Jenkins tijdreeksanalyse

- Tijdreeks bevat vaak onderling afhankelijke waarden, in de vorm van positieve autocorrelatie
- Klassieke statistische methoden zijn daarvoor niet geschikt en onderschatten dan de onzekerheid

Betrouwbaarheids- of voorspelinterval té smal en té vaak onterechte detectie van statistisch significante relaties

Vanaf Yule [1927] veel aandacht gekregen

Box and Jenkins: Time Series Analysis: Forecasting and Control (1970)
Ljung*:* System identification, theory for the user (1987)

Algemene vorm tijdreeksmodel

Deterministische component

$$Y_t = f(P, X, t) + N_t$$

Stochastische component (ruis)

Autoregressieve deel

ruismodel

'Moving-average'-deel

$$\phi(B)\Phi(B^S)(\nabla^d \nabla_S^D N_t - c) = \theta(B)\Theta(B^S)a_t$$

Differentiedeel

Constante

Modelresidu

Principe ruismodellering

Breng stochastische component terug tot verschijnsel dat:

- minimaal is
- bekende waarschijnlijkheidswetten volgt

Daardoor kunnen op basis van het model uitspraken worden gedaan met:

- minimale en
- kwantificeerbare onzekerheden

Randvoorwaarden residu a_t

Modelresidu (a_t) moet kwalificeren als normaal verdeelde *witte ruis*:

- gemiddeld nul
- onderling onafhankelijk
- zelfde variantie (homoscedastisch)
- normaal verdeeld

Nut van ruismodel

1. Geeft betere pasvorm tijdreeksmodel en betere voorspellingen (vermindert onzekerheid), door ook structuur in ruis te benutten bij het modelleren

2. Stelt in staat kwantitatieve uitspraken over onzekerheden te doen (zoals bij risico-analyses)

3. Kan aanvullende informatie geven over proces dat tijdreeks heeft gegenereerd (zoals niet-lineariteiten)

Onzekerheid resultaten meegeven

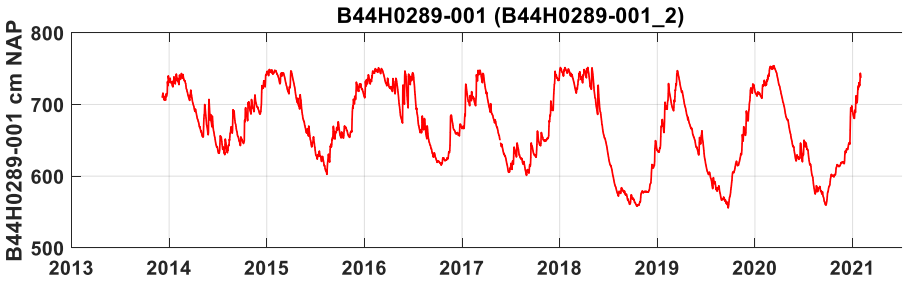
Voorbeelden

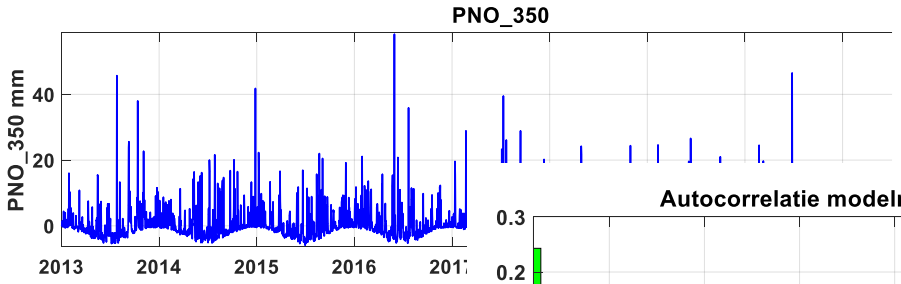
- 95%-betrouwbaarheidsinterval geschatte evenwichtsrelatie grondwaterstand en grondwaterwinning
- Resultaat statistische toets op relatie grondwaterstand en waterhuishoudkundige maatregel
- 95%-voorspelinterval van verloop grondwaterstand bij voorspeld verloop van neerslag en verdamping

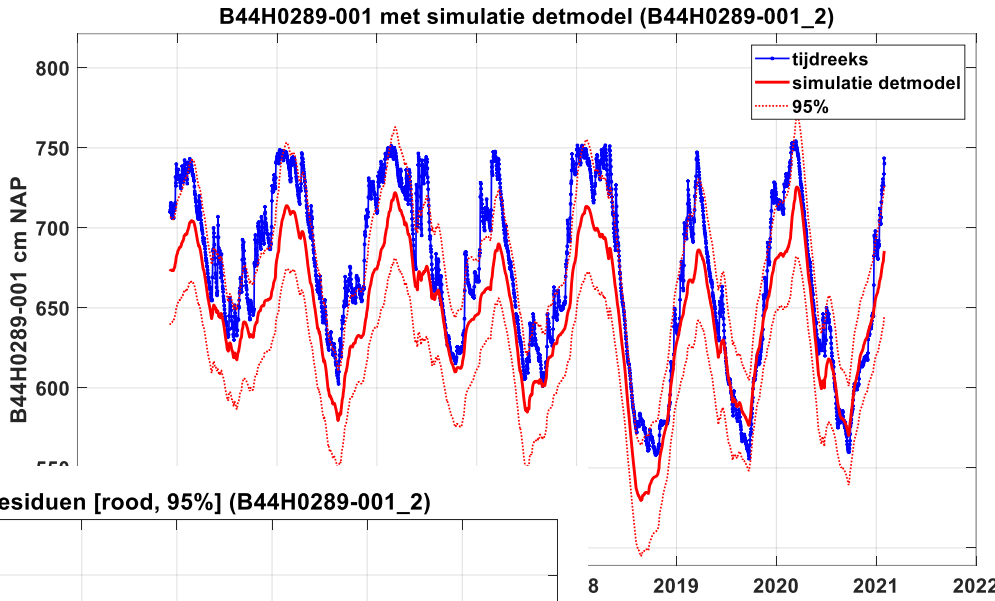
Meegeven onzekerheid geeft **altijd** meerwaarde

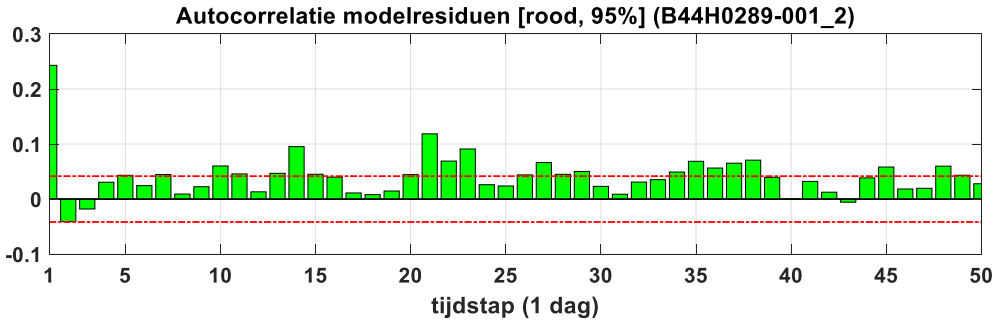
Je klant heeft daar gewoon recht op!

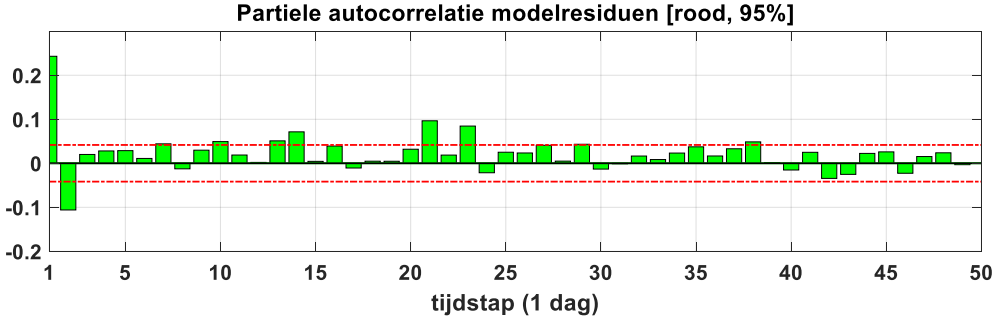
Goede pasvorm (EVP 88%), onbruikbaar model



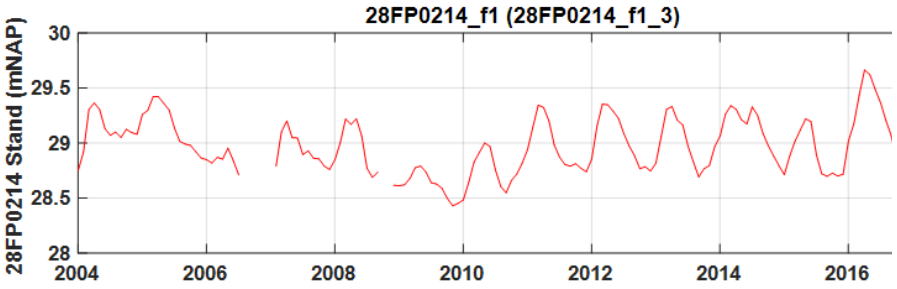


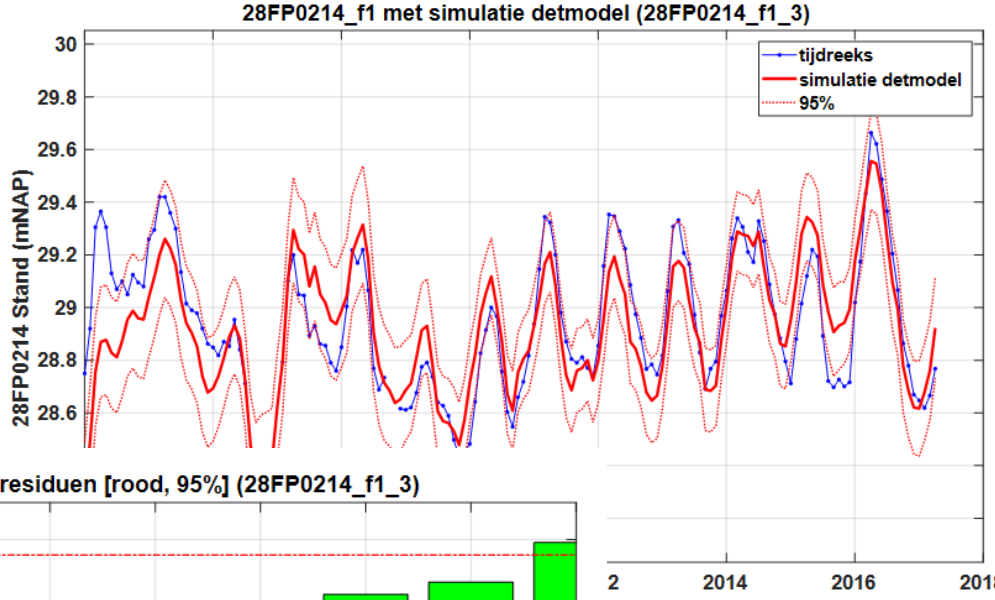


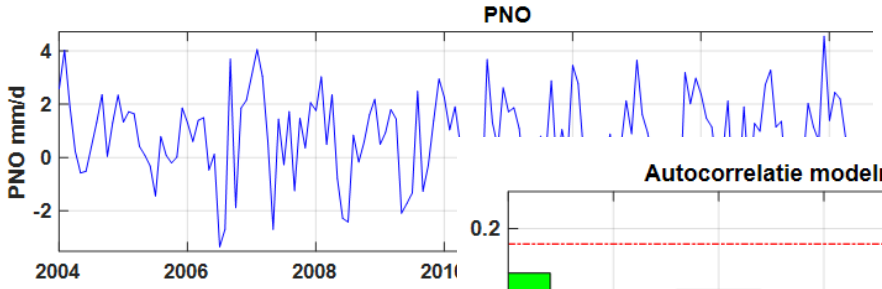


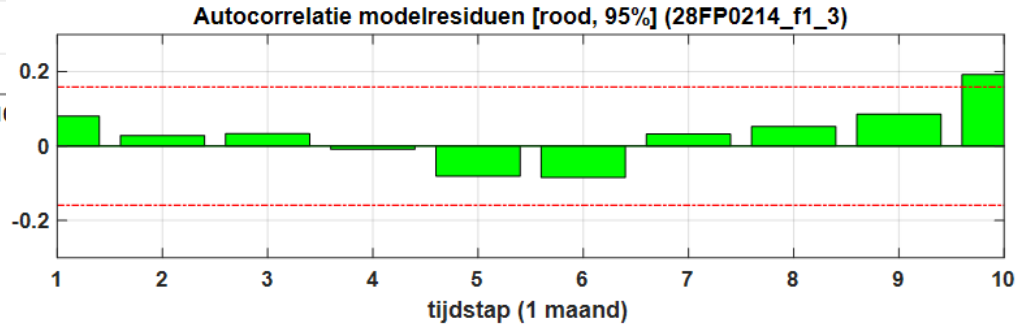


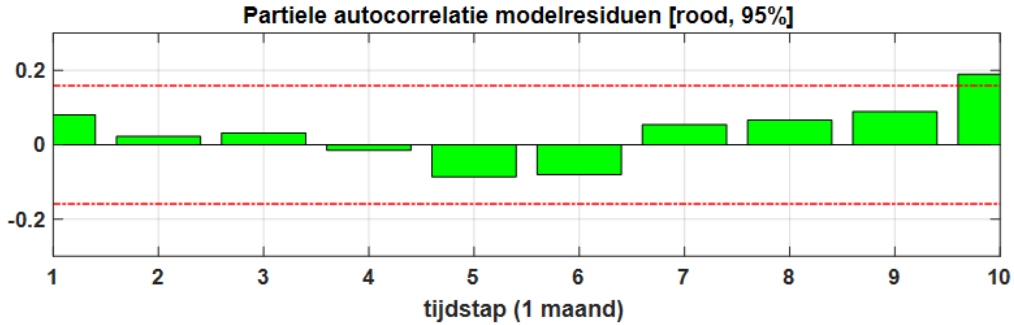
Slechte pasvorm (EVP 64%) bruikbaar model



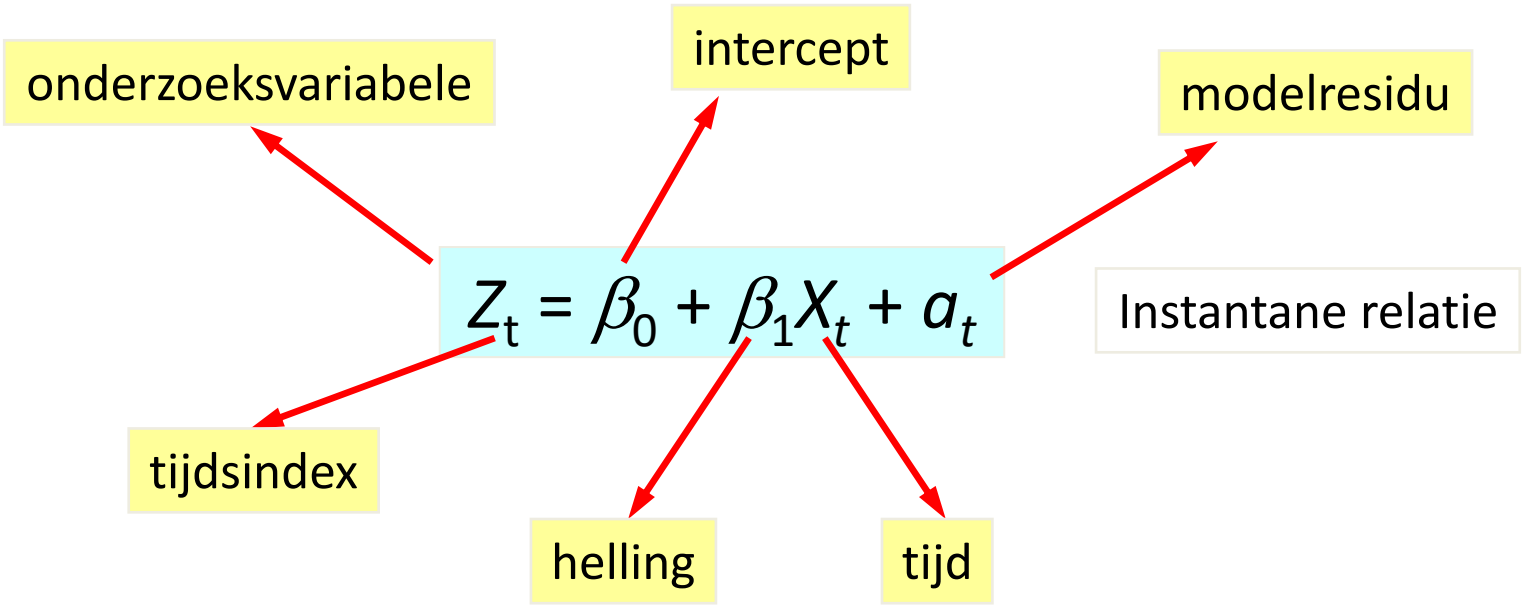








Wat als a_t geen witte ruis vormt?



Simulatie met gekleurde ruis

$$Z_t = \beta_0 + \beta_1 X_t + N_t$$

$$N_t = \phi_1 N_{t-1} + a_t - \theta_1 a_{t-1}$$

a_t: aselecte trekking uit normale verdeling, gemiddelde nul

Doe nu met vaste β_0 en β_1:

1. Kies combinatie van ϕ_1 en θ_1
2. Doe 5.000 maal het volgende: trek een aselecte reeks van a_t, bepaal daaruit de reeksen N_t en Z_t en schat vervolgens β_1 als b_1
3. Stel vast hoe vaak het 95%-betr.interval van b_1 de waarde van β_1 omvat. Dat geeft de empirische dekkingsgraad van dat interval
4. Kies nieuwe combinatie van ϕ_1 en θ_1 en ga verder met stap 2

Empirische dekkingsgraad 95%-b.i.

ϕ_1 \ θ_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%

Conclusie

Bij gekleurde ruis kan de empirische dekkingsgraad sterk afwijken van de ingestelde dekkingsgraad (hier zelfs tot 62%), het betrouwbaarheidsinterval kan dus sterk vertekend zijn en leiden tot verkeerde uitspraken over evenwichtsrelaties

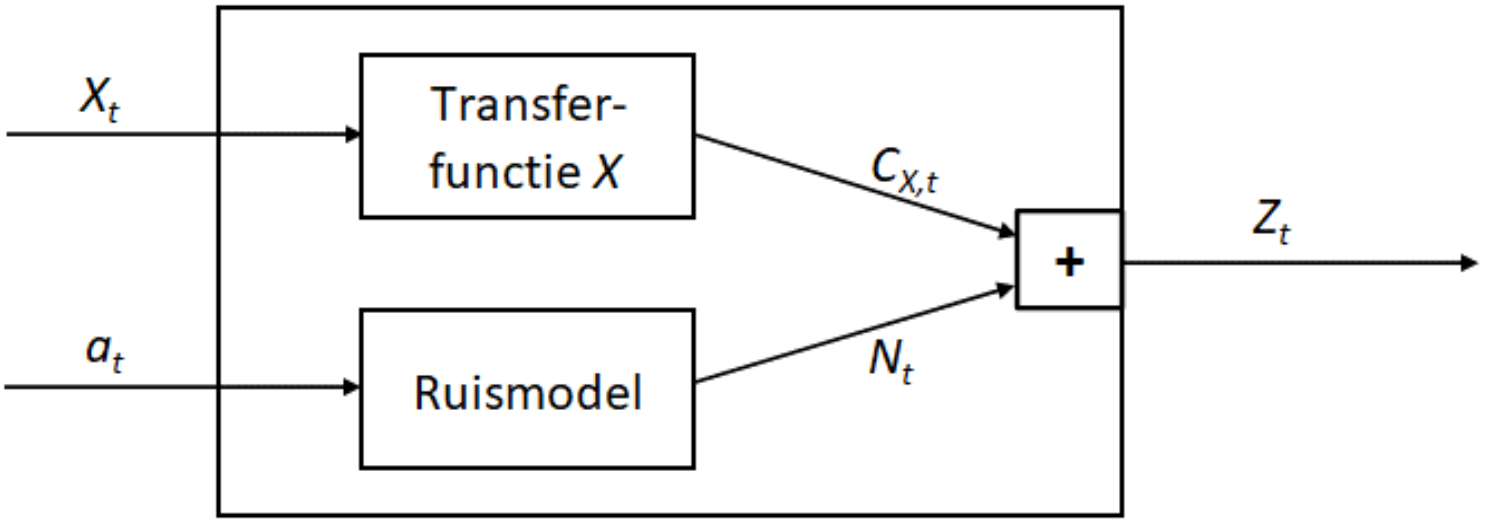
Grote toename van risico op onterechte detectie relatie!

Goodness of fit

Mate van door deterministisch deel van model verklaarde variantie

Tegenwoordig vaak uitgedrukt als EVP

Transfer-ruismodel



$$Z_t = C_{X,t} + N_t$$

Is EVP de verklaarde variantie?

$$Z_t = C_{X,t} + N_t$$

EVP wordt bepaald als: EVP = $1 - \frac{\mathrm{Var}(N_t)}{\mathrm{Var}(Z_t)}$

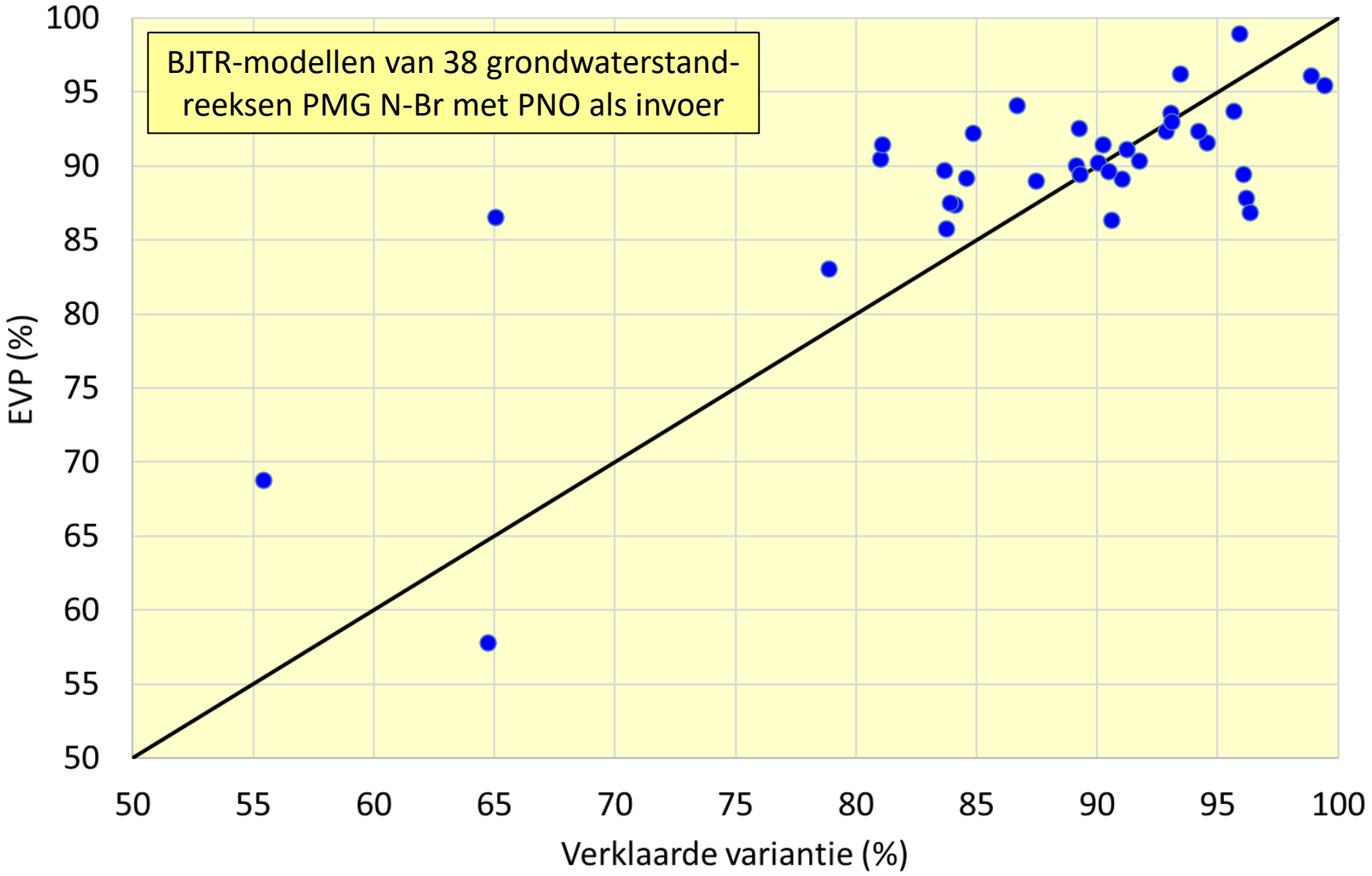
Checken: $\mathrm{Var}(Z_t) = \mathrm{Var}(C_{X,t} + N_t)$

$$\mathrm{Var}(Z_t) = \mathrm{Var}(C_{X,t}) + \mathrm{Var}(N_t) + 2 \cdot \mathrm{Covar}(C_{X,t}, N_t)$$

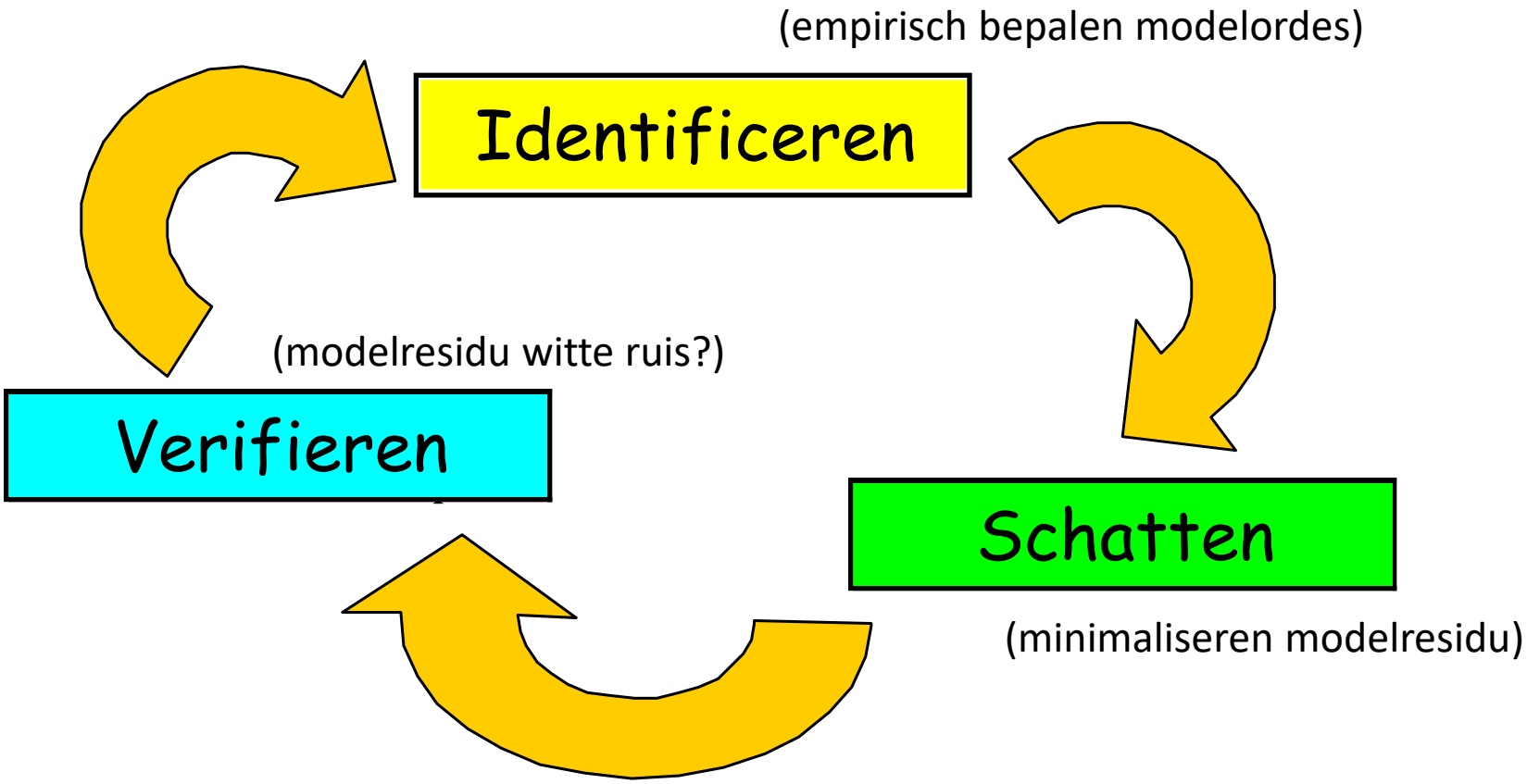
Verklaard door $C_{X,t}$: $\frac{\mathrm{Var}(C_{X,t})}{\mathrm{Var}(Z_t)} = 1 - \frac{\mathrm{Var}(N_t)}{\mathrm{Var}(Z_t)} - \frac{2 \cdot \mathrm{Covar}(C_{X,t}, N_t)}{\mathrm{Var}(Z_t)}$

EVP is alleen gelijk aan de verklaarde variantie
als N_t **niet** is gecorreleerd aan deterministische component,
maar dergelijke correlaties treden vaak juist wél op!

EVP versus verklaarde variantie

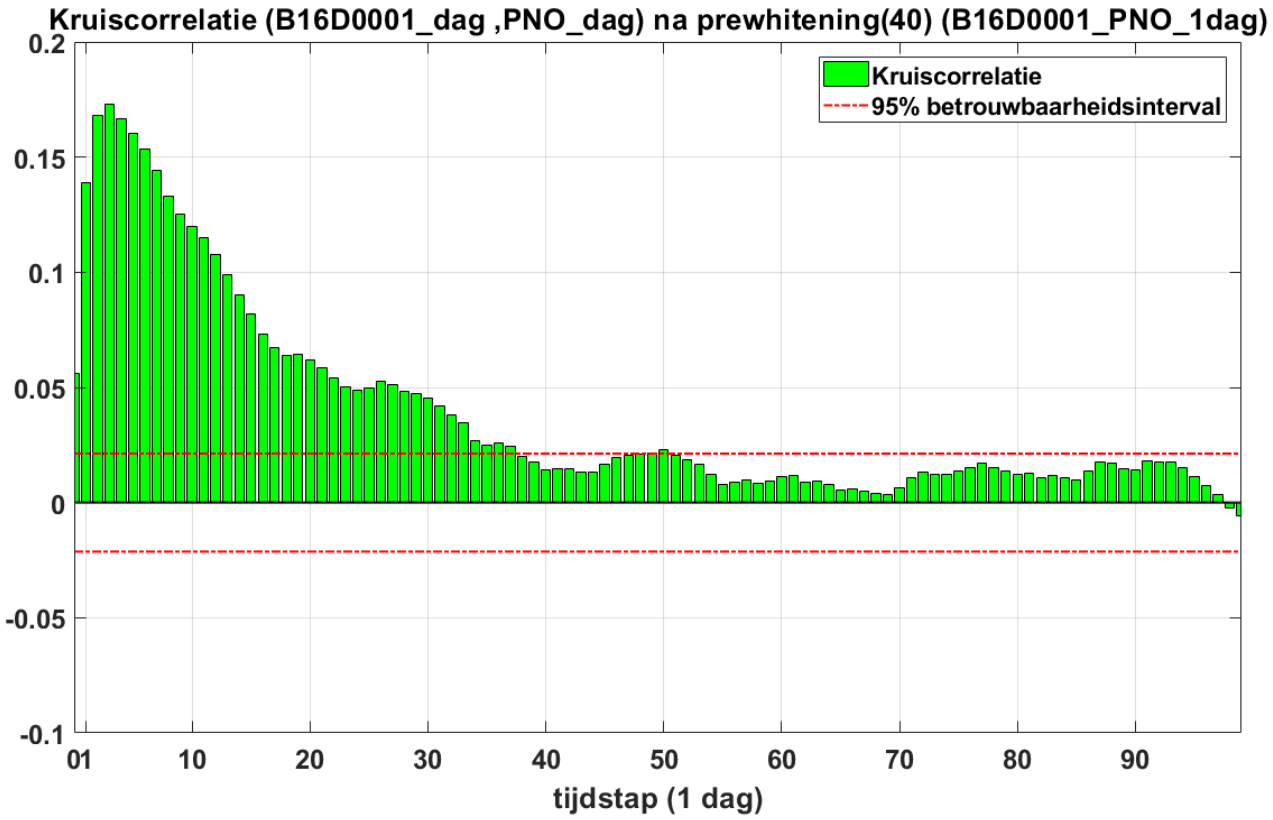


Modelleringscyclus



Modelidentificatie ('laat de data spreken')

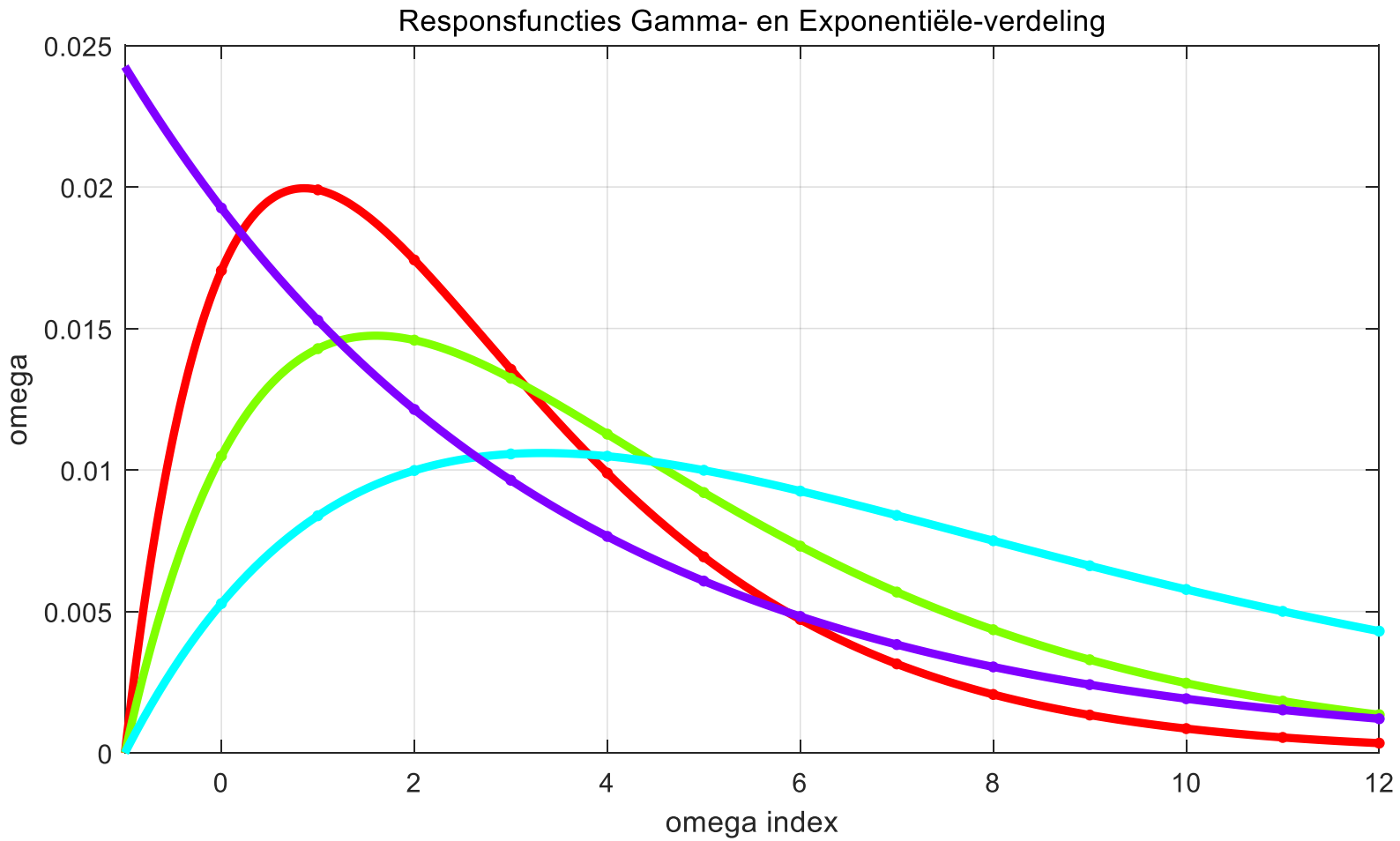
Gewit kruiscorrelogram



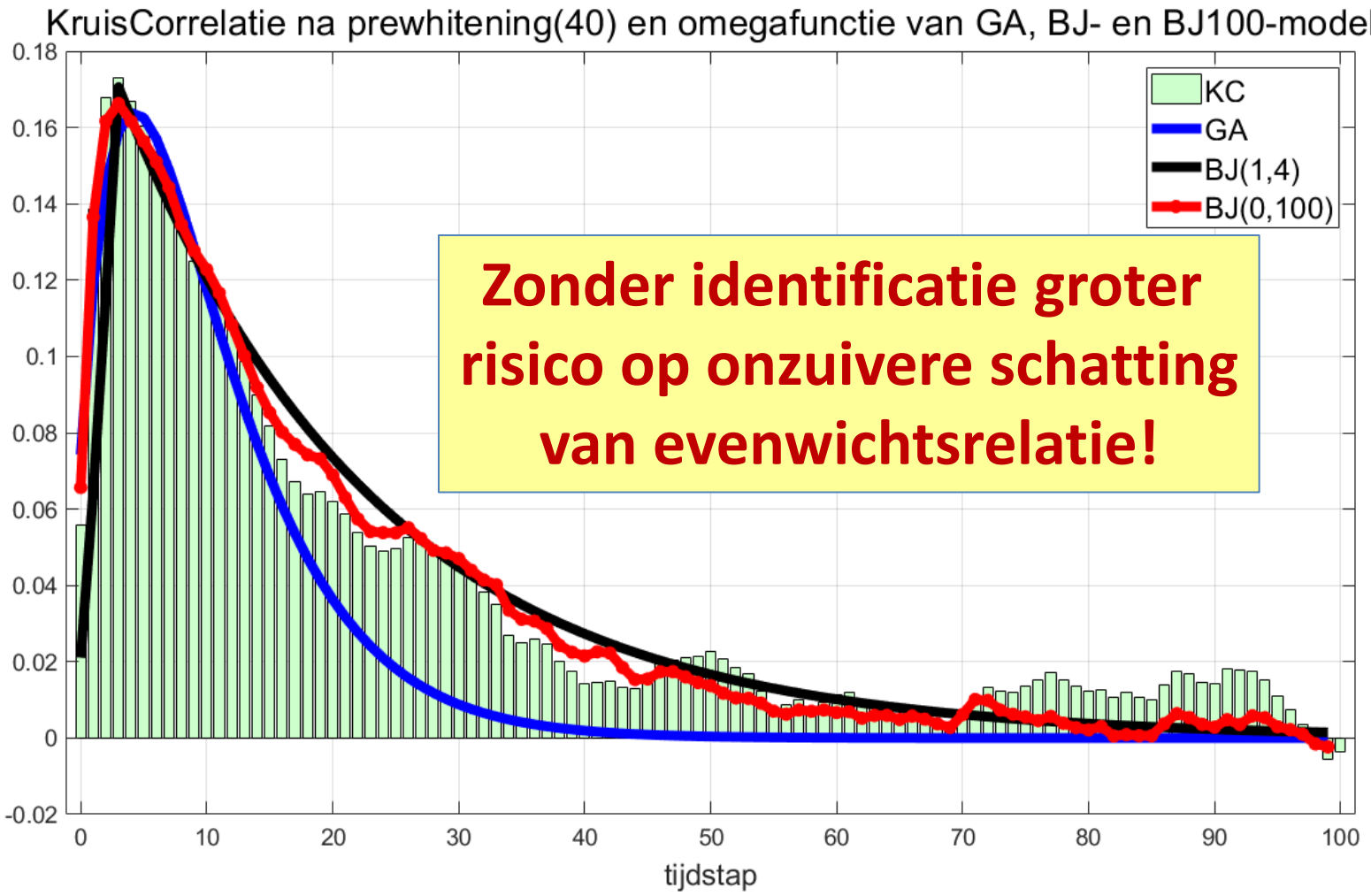
1. $C_t = \delta_1 C_{t-1} + \omega_0 X_t - \omega_1 X_{t-1} - \omega_2 X_{t-2} - \omega_3 X_{t-3}$

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

PIRFICT-opties



Vergelijking geschatte responsen



Zonder identificatie groter risico op onzuivere schatting van evenwichtsrelatie!

Wat is verontrustend?

- Een model wordt nu vooral beoordeeld op pasvorm (EVP), terwijl die weinig zegt over de bruikbaarheid van het model
- De ‘gebruiksvriendelijke’ software is teveel gaan bepalen hoe tijdreeksanalyse moet worden uitgevoerd en biedt weinig mogelijkheden daarbij te kunnen voldoen aan alle statistische randvoorwaarden die gelden voor toepassing van een model
- Gebruiksgemak is gaan prevaleren en dat is ten koste gegaan van de bruikbaarheid
- Bij veel gebruikers is er te weinig kennis van de achtergronden van tijdreeksanalyse om deze valkuil te kunnen vermijden

Hoe moet het dan?

- Beoordeel altijd of een model voldoet aan de statistische randvoorwaarden voor toepassing (fitness-for-use)
- Ga alleen dán af op de pasvorm als kan worden gekozen uit meerdere modellen die voldoen aan de statistische randvoorwaarden
- Valideer zonodig je model met criteria die pragmatisch aansluiten op het beoogde gebruiksdoel van het model
- Breng de identificatiefase weer terug! En laat zo de data je vertellen welke transferfunctie bij ze past

Hoe moet het dan?

- Als de kwaliteit van je tijdreeksanalyse wordt beperkt door de software, spreek de ontwikkelaars daar op aan, alleen zo kunnen we samen het algehele niveau verhogen

Of heeft de ontwikkelaar ‘zorgplicht’ in deze?

Meer weten?

https://www.amo-nl.com/nieuws