Tijdreeksanalyse volgens Box-Jenkins-methode

7 oktober 2021

Paul Baggelaar

PB Icastat

Tijdreeksanalyse is een wiskundig deelgebied dat zich bezighoudt met het analyseren, voorspellen of beheren van tijdreeksen, met behulp van tijdreeksmodellering

Probleem

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

→ voorspelinterval te smal en te vaak onterechte detectie van statistisch significante relatie

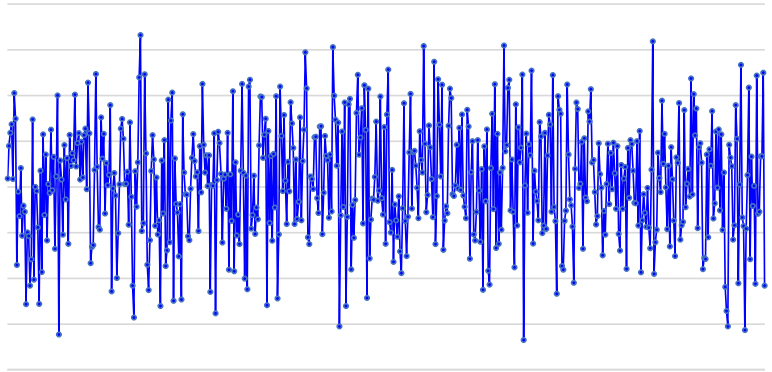
Vanaf Yule [1927] veel aandacht gekregen

Syntheses van oplossingen in Box and Jenkins [1970]:
Time Series Analysis: Forecasting and Control

Autocorrelatie (positieve)

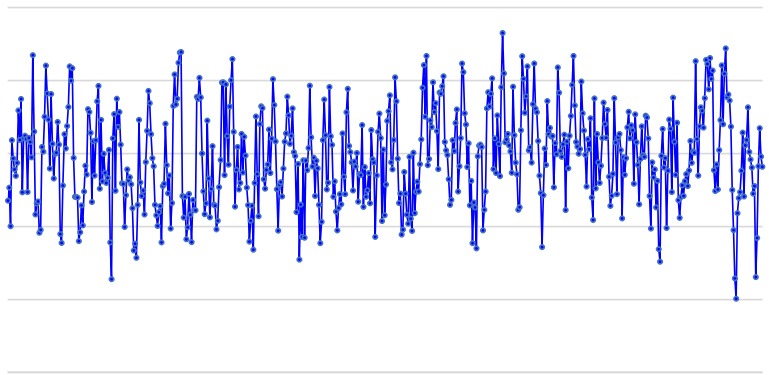
$\phi_1 = 0{,}0$

Z_t



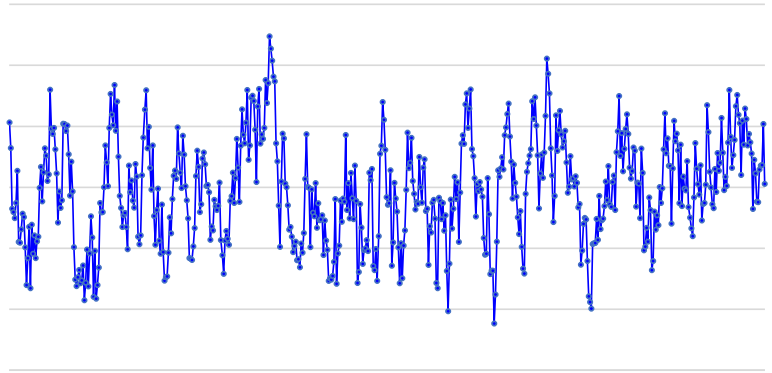
$\phi_1 = 0{,}5$

Z_t



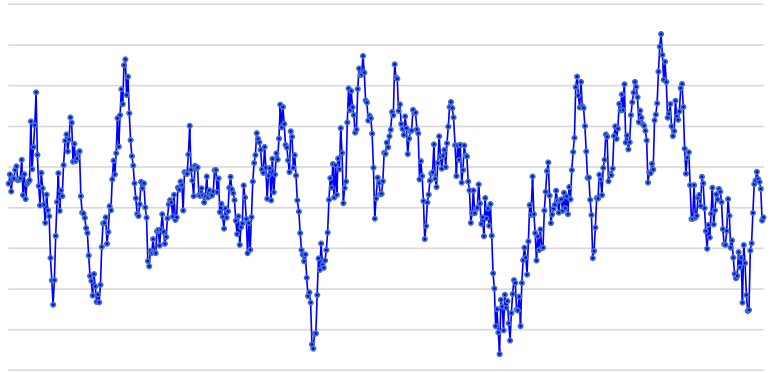
$\phi_1 = 0{,}8$

Z_t

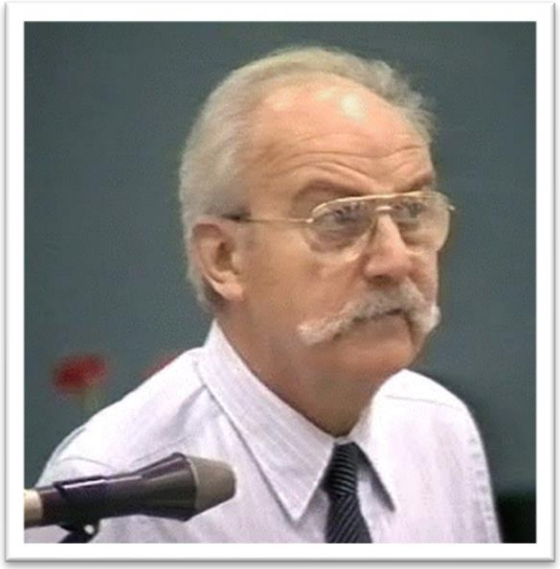


$\phi_1 = 0{,}95$

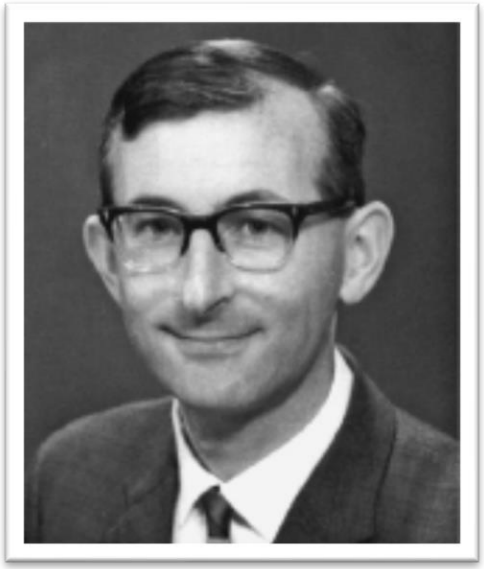
Z_t



Box en Jenkins

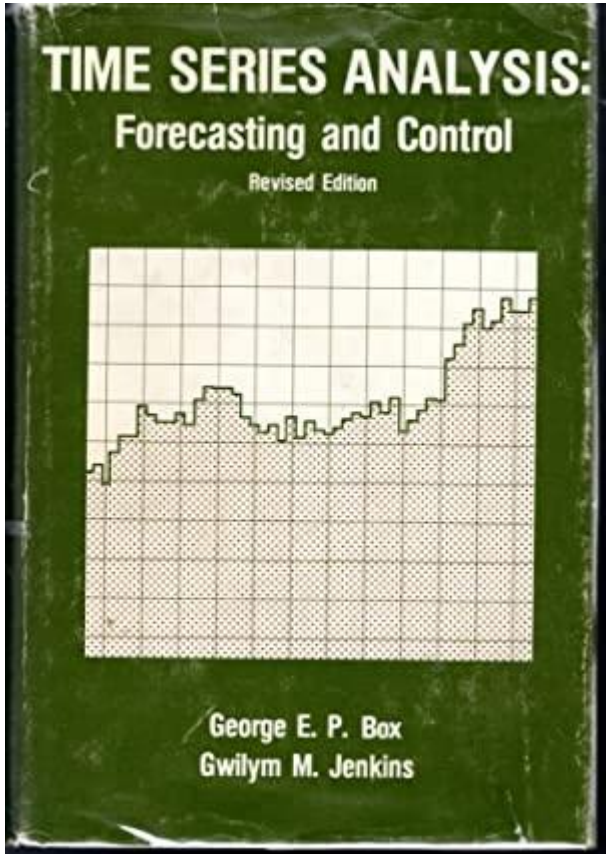


George E.P. Box
(1919 – 2013)



Gwilym M. Jenkins
(1932 – 1982)

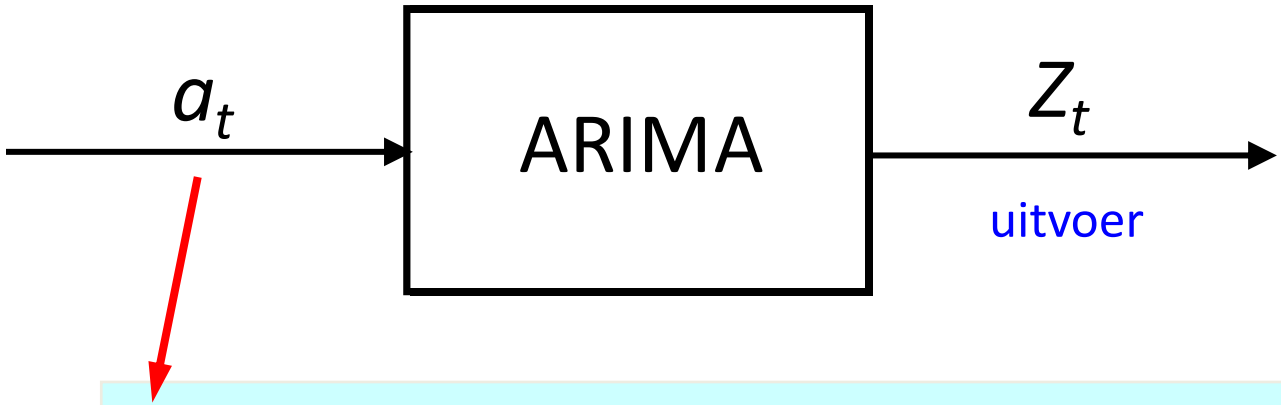
Hun bekende boek



Twee families van modellen:

1. Univariate model
2. Transfer-ruismodel

Univariate model



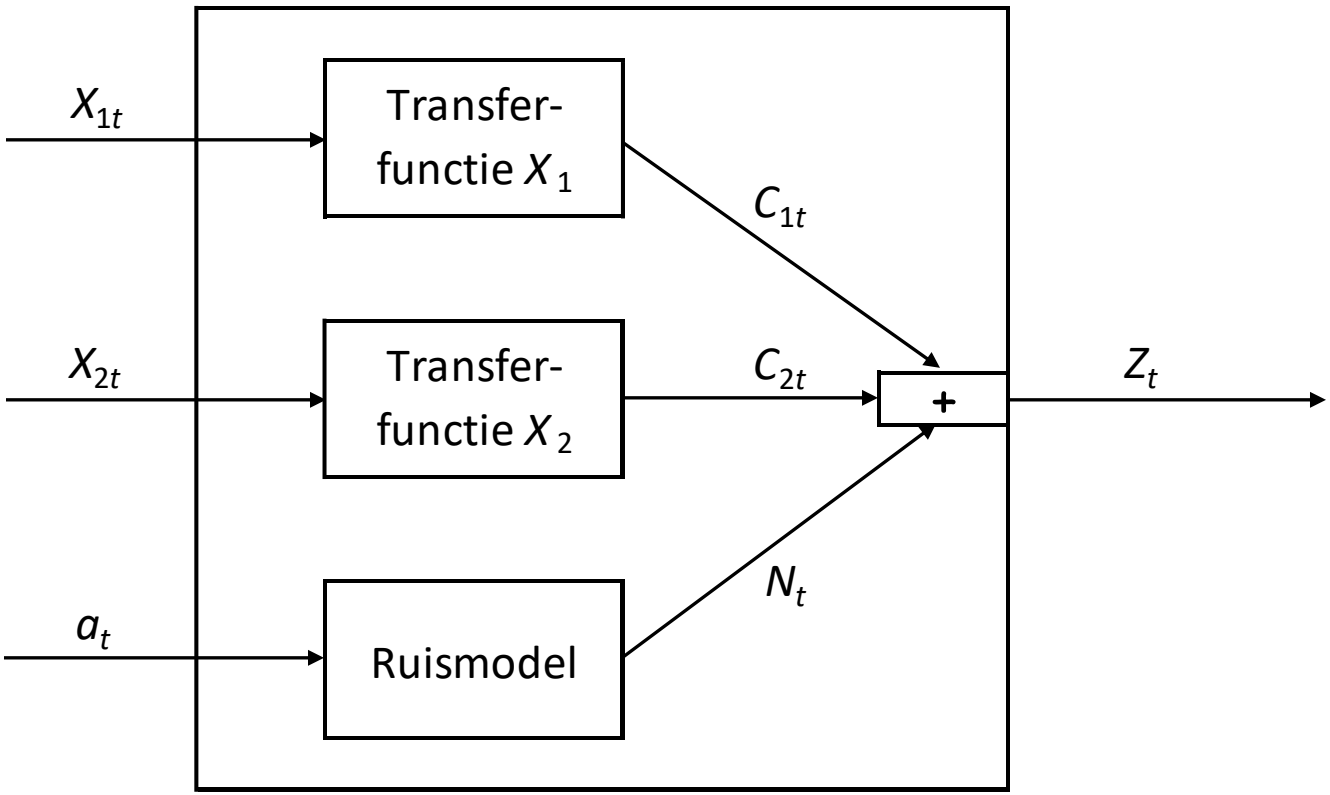
a_t moet kwalificeren als normaal verdeelde *witte ruis*:

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

<u>Voorbeeld</u>

$$Z_t - c = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

Transfer-ruismodel



$$Z_t = \mathrm{Tf}_1(X_{1t}) + \mathrm{Tf}_2(X_{2t}) + N_t$$

Transfer-ruismodel

grondwaterstand

transferfuncties

ruis

$$Z_t = \boxed{\mathrm{Tf}_{\mathrm{PNO}}(PNO) + \mathrm{Tf}_{\mathrm{Q}}(Q)} + \boxed{N_t}$$

deterministisch deel

stochastisch deel

Voorbeeld ruismodel (AR) $N_t = \phi_1 N_{t-1} + a_t$ Is Univariaat model

Modelresidu moet kwalificeren als normaal verdeelde *witte ruis*:

- gemiddeld nul
- onderling onafhankelijk
- zelfde variantie
- normaal verdeeld

Principe Box-Jenkins-methode

Stochastische component (N_t) wordt teruggebracht tot verschijnsel (a_t) dat:

- minimaal is
- bekende waarschijnlijkheidswetten volgt

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

- minimale en
- kwantificeerbare onzekerheden

Van belang: gaat uit van equidistante tijdreeksen

Mijn achtergrond in deze

- 1984 Hydroloog bij Kiwa
- Had als mentor Johan Groennou (wiskundige/statisticus)
- Hydrologen pasten toen nog vaak regressieanalyse toe op tijdreeksen, was onverantwoord
- Onderricht gevolgd bij Gwylim Jenkins & Partners (Lancaster)
- Daarna veel ervaring opgedaan met modelleren hydrologische tijdreeksen (attributiestudies) en tijdreeksen consumptiegoederen (voorspellen afzet)

Onzekerheid resultaten meegeven

<u>Voorbeelden</u>

- 95%-betrouwbaarheidsinterval geschatte evenwichts-relatie grondwaterstand en grondwaterwinning

- Resultaat statistische toets op relatie grondwaterstand en waterhuishoudkundige maatregel

- 95%-voorspelinterval van verloop grondwaterstand bij voorspeld verloop van neerslag en verdamping

Meerwaarde van onzekerheid

- Objectieve en waardevolle maat voor gebrek informatie-inhoud van model + data
- Kan inspiratie geven voor verbetering en daar ook richting aan geven
- Managers kunnen met kennis van de onzekerheid veel beter besluiten onderbouwen
- Is meestal geen onkunde van modelleur

Niet iedereen houdt van onzekerheid

Belang van witte ruis

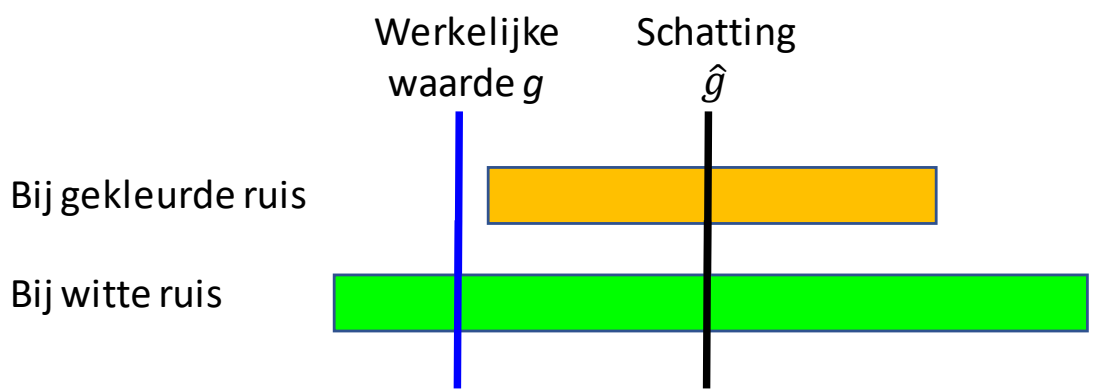
<u>Simulatie-exercitie</u>

1. Ga uit van relatie: $Z_t = \beta_0 + \beta_1 X_t + N_t$ Gekleurde ruis
2. Genereer N_t als: $N_t = \phi_1 N_{t-1} + a_t - \theta_1 a_{t-1}$ (ARMA)
3. Simuleer per combinatie van ϕ_1 en θ_1 5.000 reeksen, door aselect a_t's te trekken uit een normale verdeling
4. Schat met LR per reeks b_1 en zijn 95%-betrouwb.interval $95\%\text{b. i.}\,[b_1] \;=\; b_1 \pm t_{97,5\%;k-2} \cdot s_{b_1}$
5. Stel vast in welk percentage van de 5.000 simulaties dat interval de werkelijke waarde β_1 bevat => empirische dekkingsgraad van 95%-betrouwbaarheidsinterval

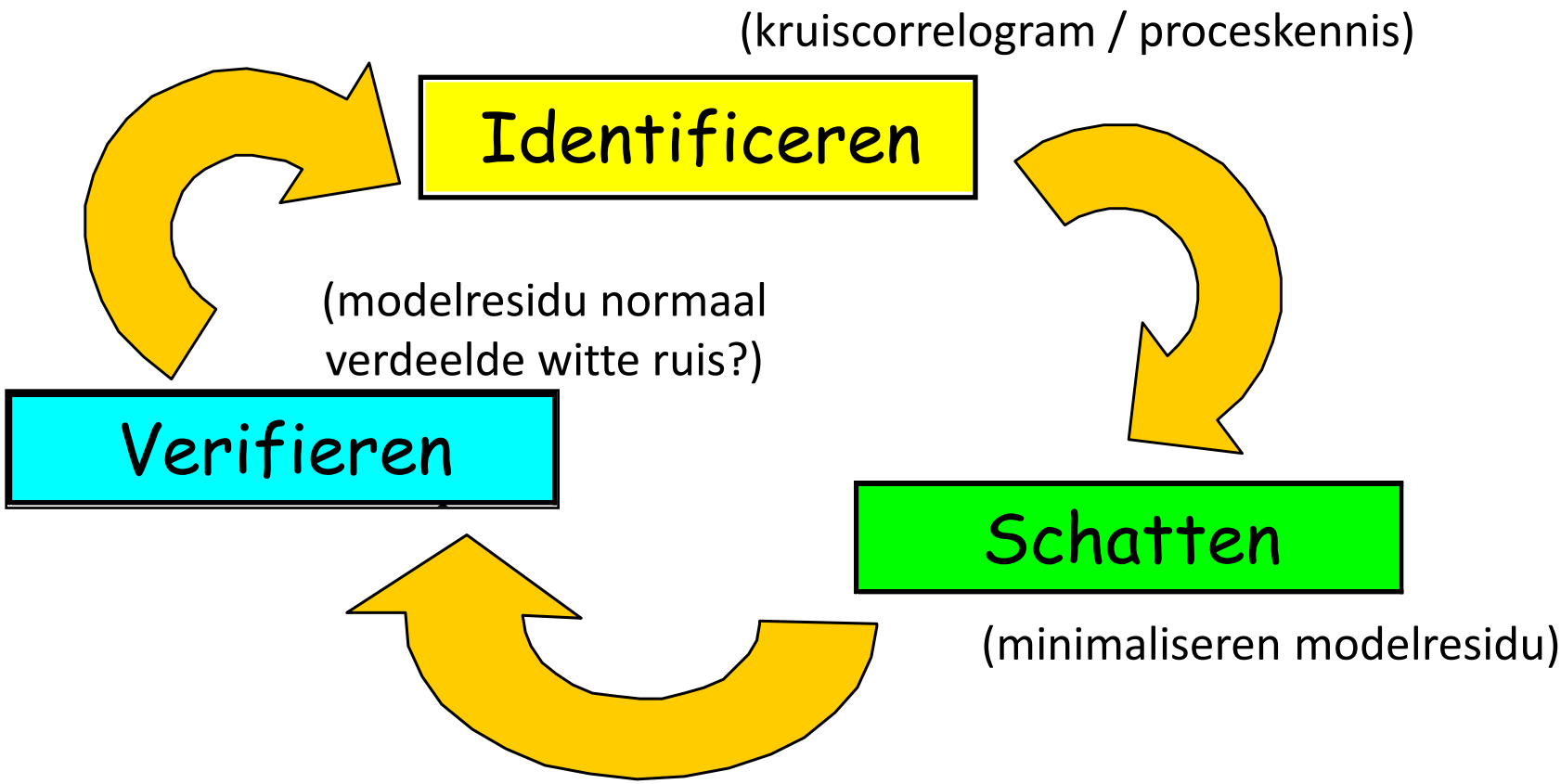
Empirische dekkingsgraad
(bij nominale dekkingsgraad van 95%)

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%

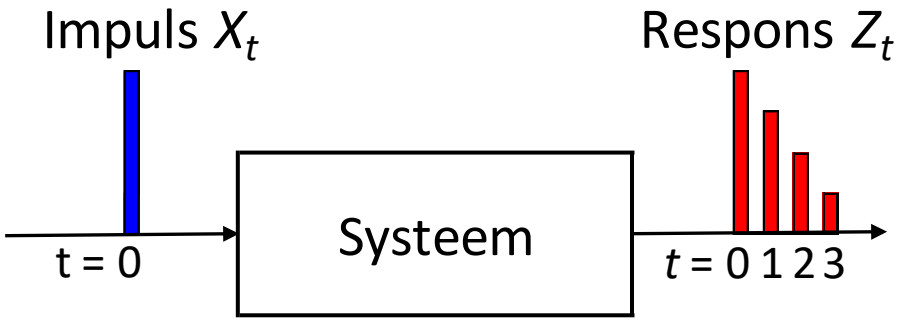
95%-betr.interval: $\hat{g} \pm t_{(97{,}5\%;vg)} \cdot s_{\hat{g}}$



Modelleercyclus



Ideale situatie voor identificatie



Mogelijke transferfuncties:

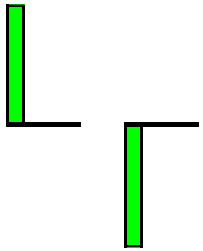
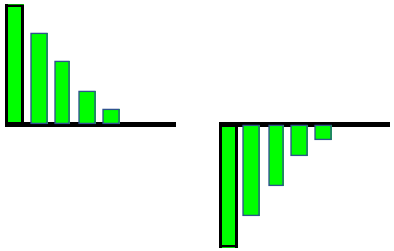
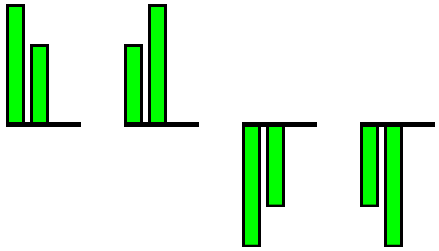
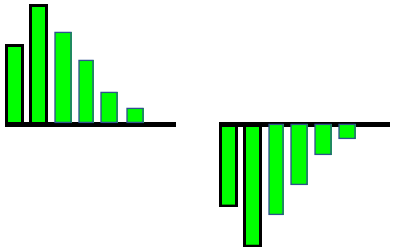
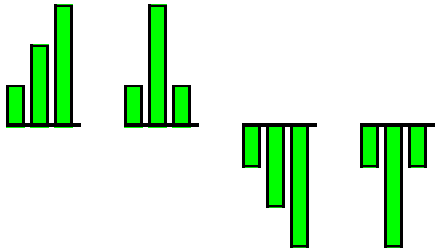
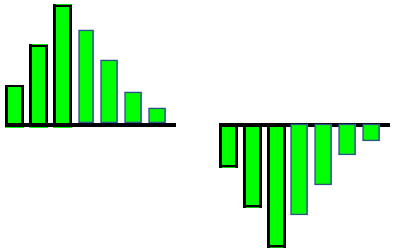
- ARMA(1,0):

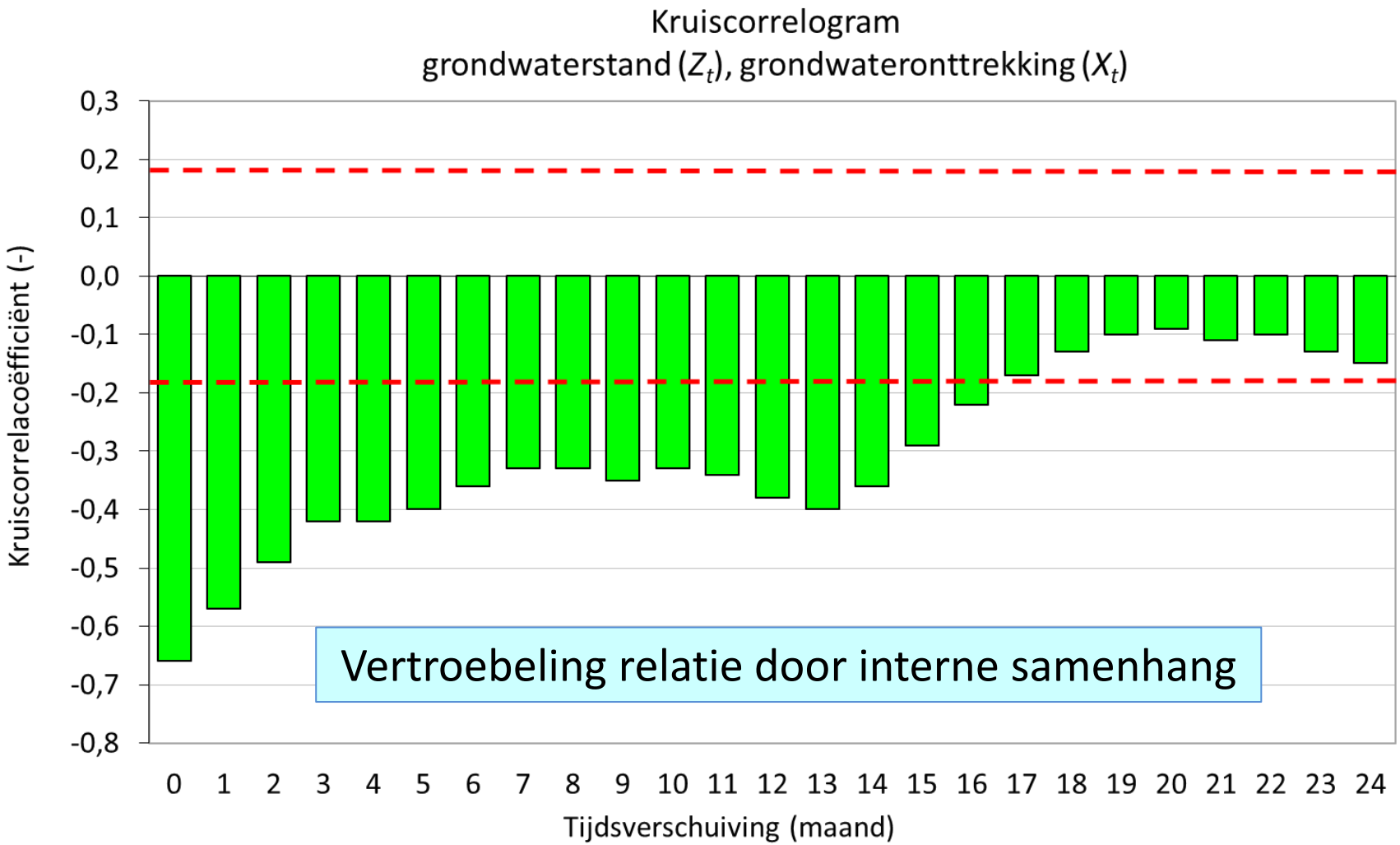
$$Z_t = \frac{\omega_0}{(1-\delta_1 \mathrm{B})} X_t + N_t$$

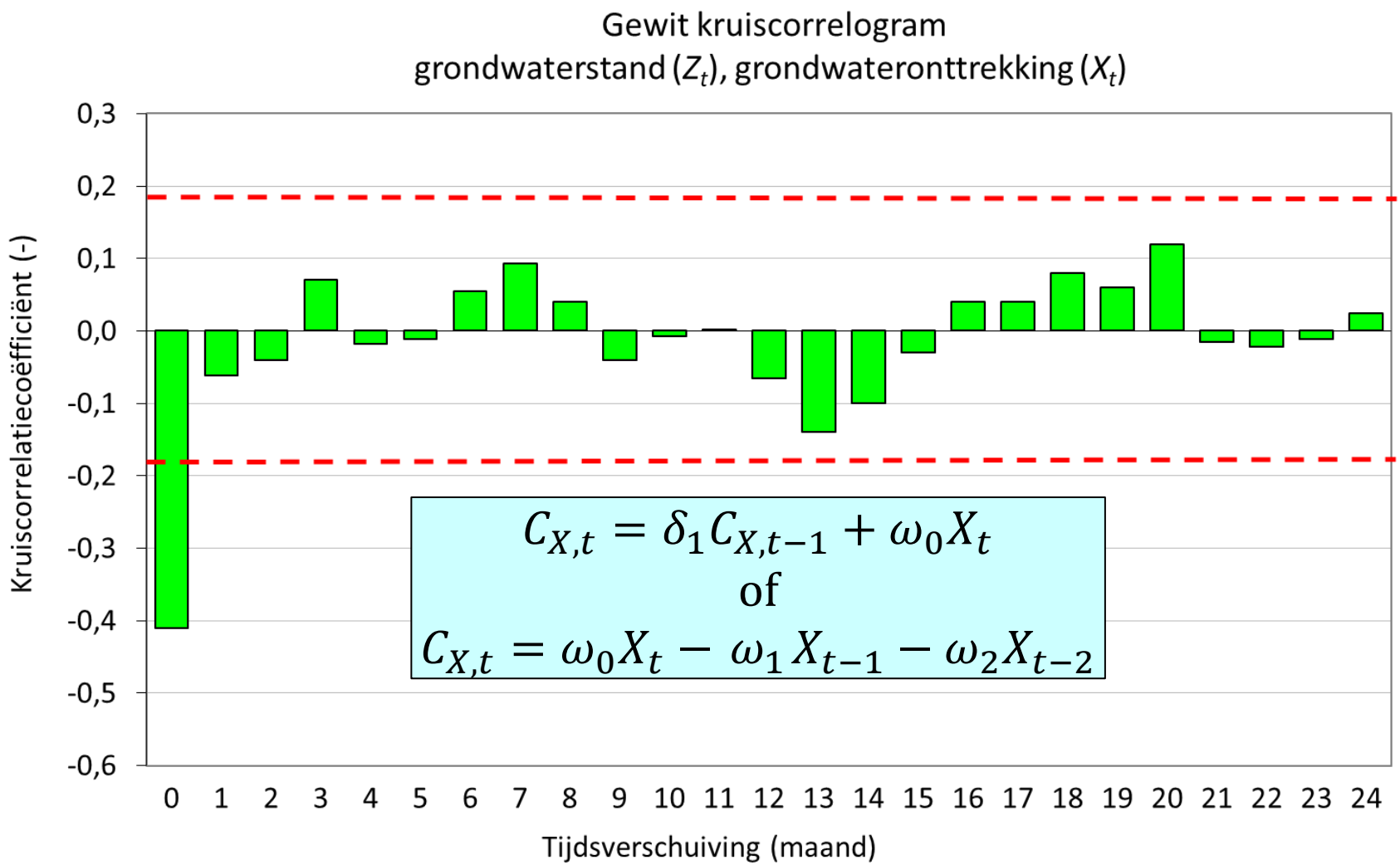
- ARMA(0,3):

$$Z_t = \omega_0 X_t - \omega_1 X_{t-1} - \omega_2 X_{t-2} - \omega_3 X_{t-3} + N_t$$

Identificatie transferfunctie

Transferfunctie	Vorm van respons	Transferfunctie	Vorm van respons
ARMA(0,0) $\omega_0 X_t$		ARMA(1,0) $\frac{\omega_0 X_t}{(1-\delta_1 \mathrm{B})}$	
ARMA(0,1) $\omega_0 X_t - \omega_1 X_{t-1}$		ARMA(1,1) $\frac{\omega_0 X_t - \omega_1 X_{t-1}}{(1-\delta_1 \mathrm{B})}$	
ARMA(0,2) $\omega_0 X_t - \omega_1 X_{t-1} - \omega_2 X_{t-2}$		ARMA(1,2) $\frac{\omega_0 X_t - \omega_1 X_{t-1} - \omega_2 X_{t-2}}{(1-\delta_1 \mathrm{B})}$	


Kruiscorrelogram
grondwaterstand (Z_t), grondwateronttrekking (X_t)
Kruiscorrelacoëfficiënt (-)
0,3
0,2
0,1
0,0
-0,1
-0,2
-0,3
-0,4
-0,5
-0,6
-0,7
-0,8
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Tijdsverschuiving (maand)
Vertroebeling relatie door interne samenhang



$$C_{X,t} = \delta_1 C_{X,t-1} + \omega_0 X_t$$

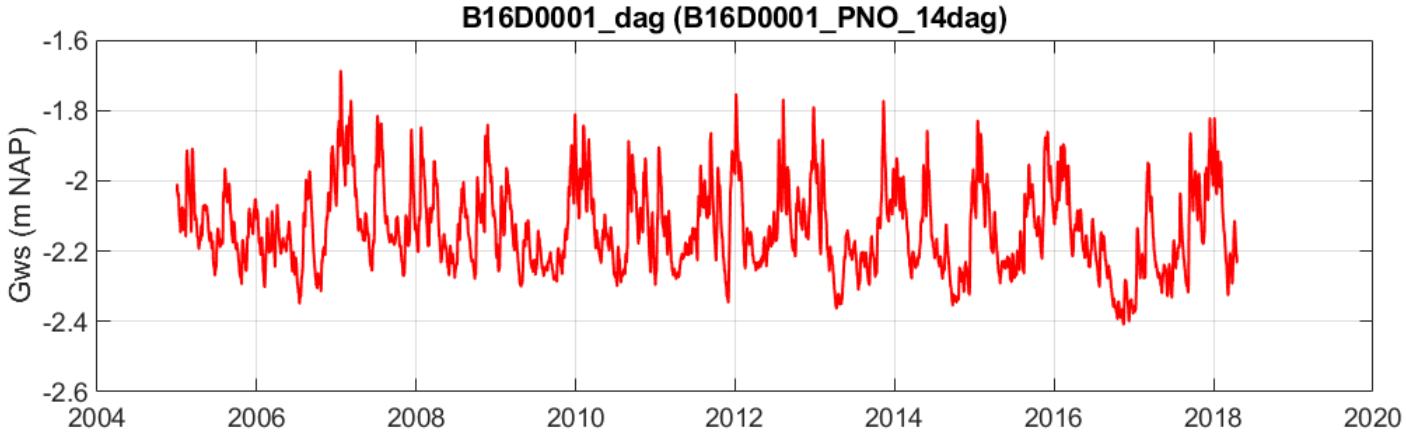
of

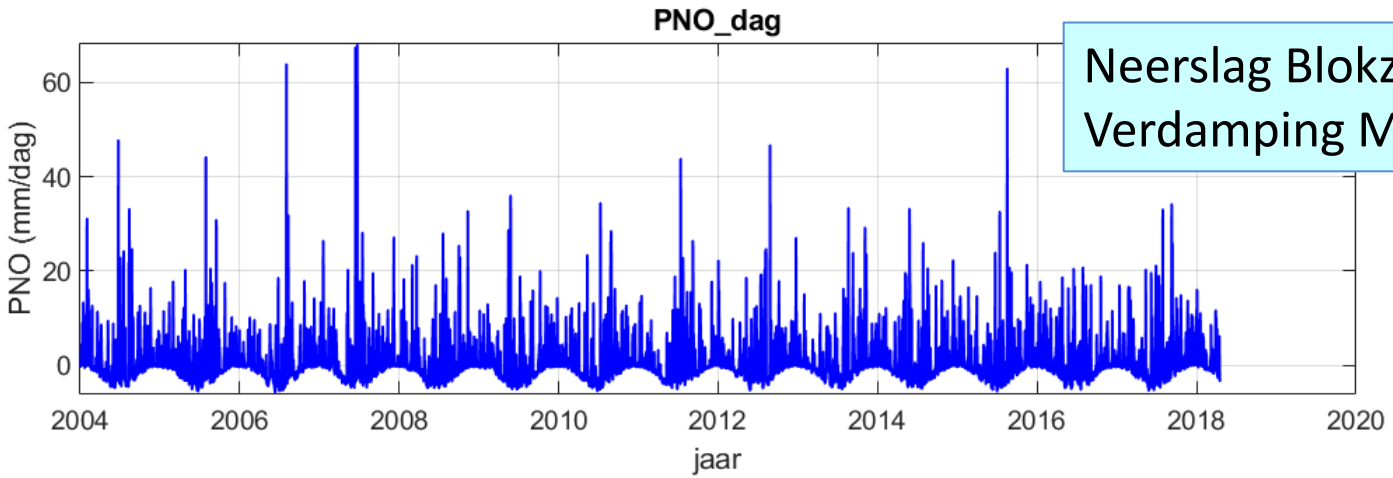
$$C_{X,t} = \omega_0 X_t - \omega_1 X_{t-1} - \omega_2 X_{t-2}$$

Kan desgewenst ook met hydrologische functie

Enkele praktijkvoorbeelden

Meetreeksplots (beide op dagbasis)

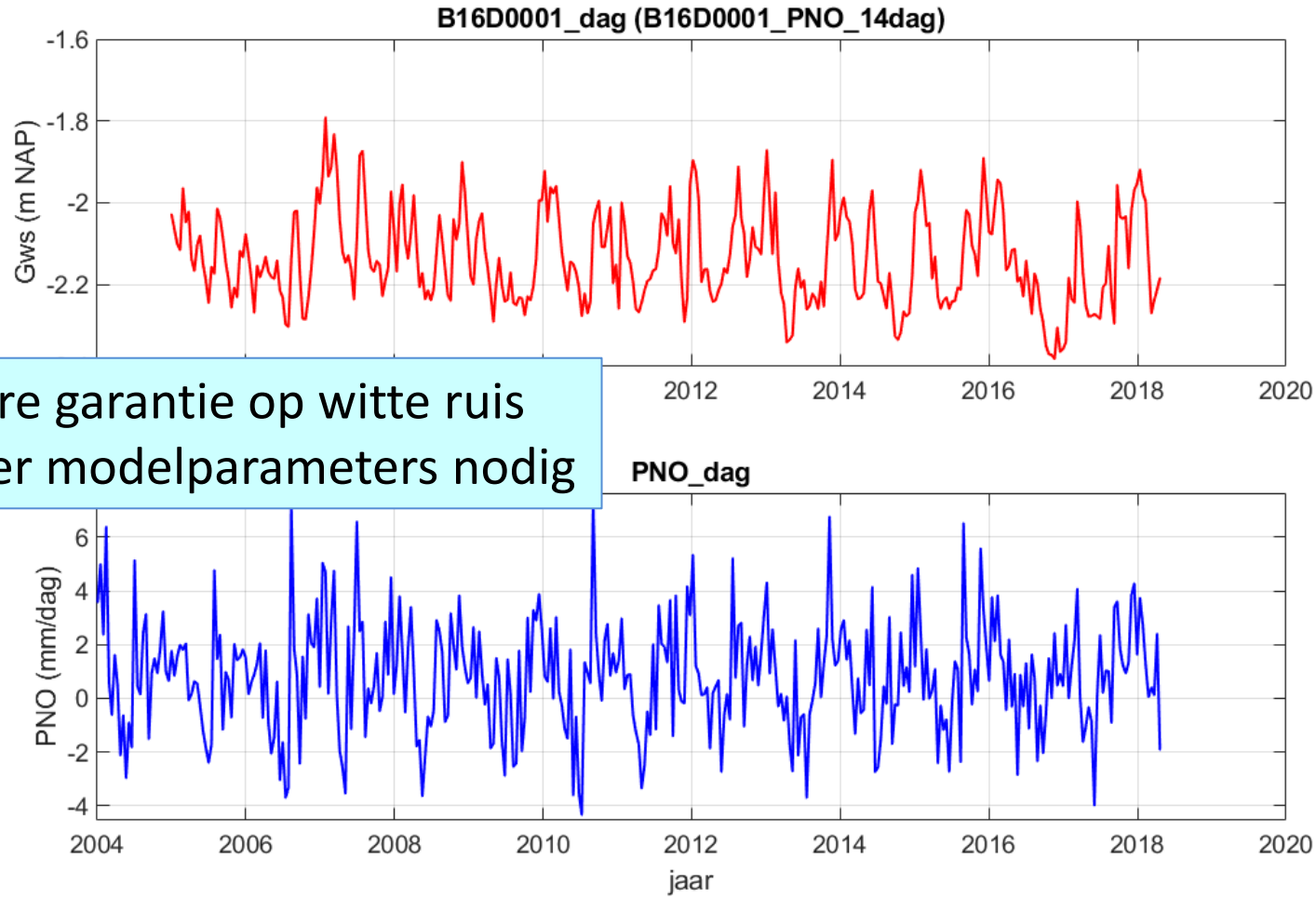




Neerslag Blokzijl
Verdamping Marknesse

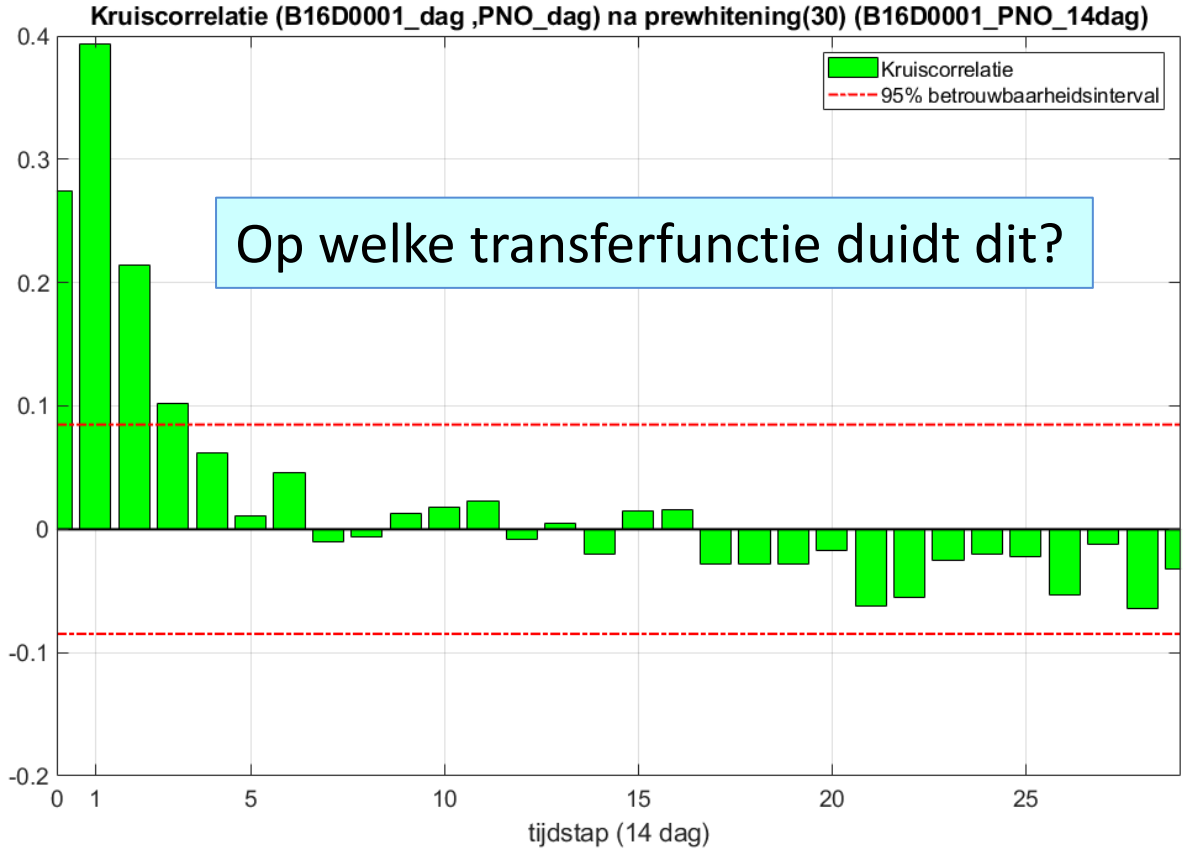
Geen duidelijke dataproblemen

Meetreeksplots 14-daagse basis



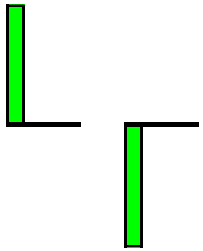
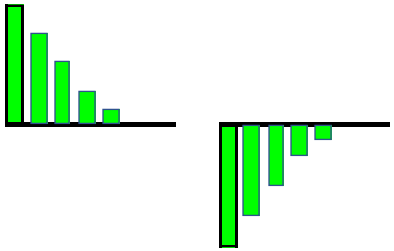
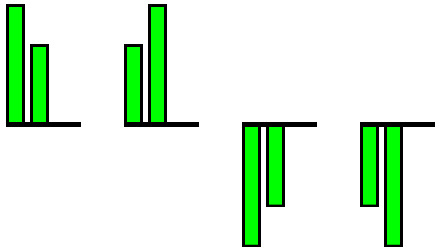
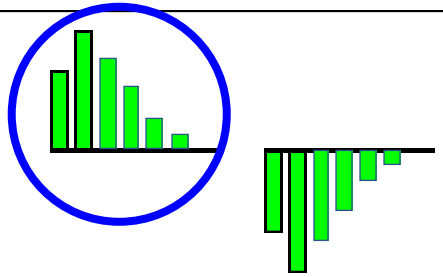
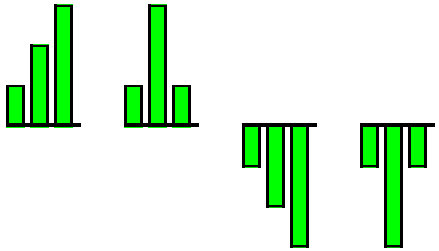
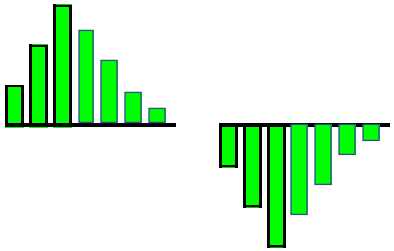
- Grotere garantie op witte ruis
- Minder modelparameters nodig

Gewit kruiscorrelogram

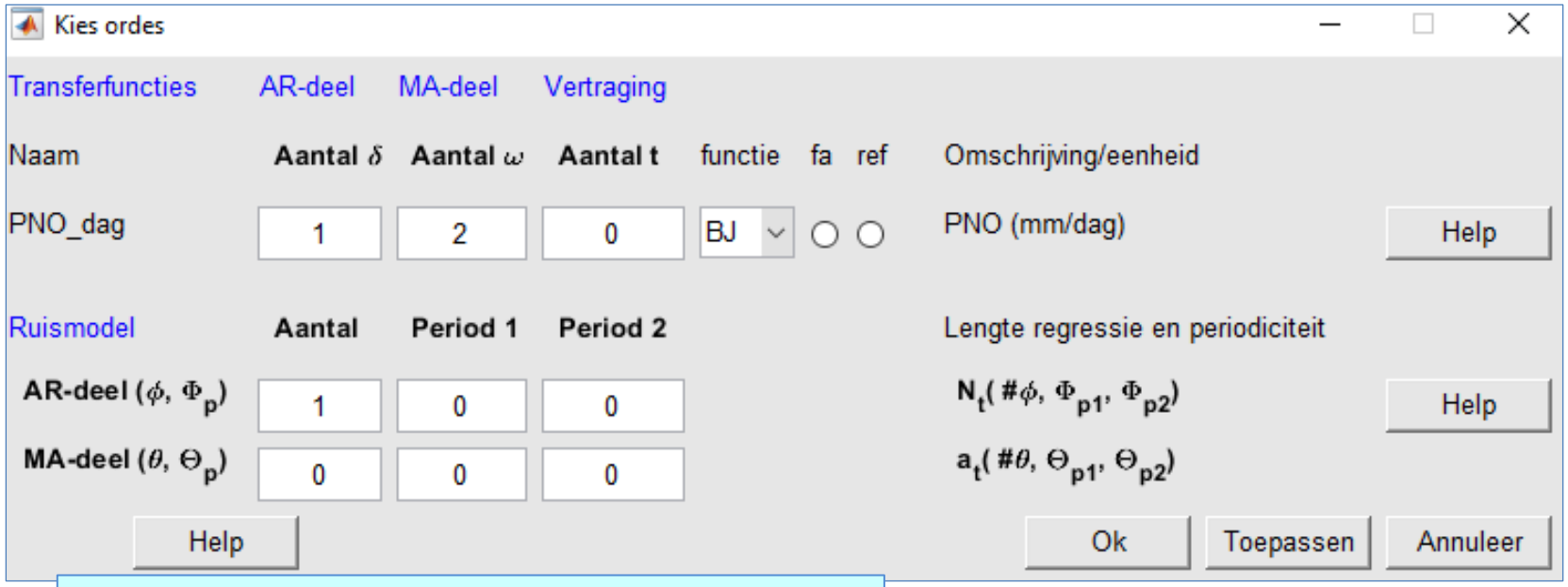


Op welke transferfunctie duidt dit?

Identificatie transferfunctie

Transferfunctie	Vorm van respons	Transferfunctie	Vorm van respons
ARMA(0,0) $\omega_0 X_t$		ARMA(1,0) $\frac{\omega_0 X_t}{(1-\delta_1 \mathrm{B})}$	
ARMA(0,1) $\omega_0 X_t - \omega_1 X_{t-1}$		ARMA(1,1) $\frac{\omega_0 X_t - \omega_1 X_{t-1}}{(1-\delta_1 \mathrm{B})}$	
ARMA(0,2) $\omega_0 X_t - \omega_1 X_{t-1} - \omega_2 X_{t-2}$		ARMA(1,2) $\frac{\omega_0 X_t - \omega_1 X_{t-1} - \omega_2 X_{t-2}}{(1-\delta_1 \mathrm{B})}$	

Formuleren tijdreeksmodel

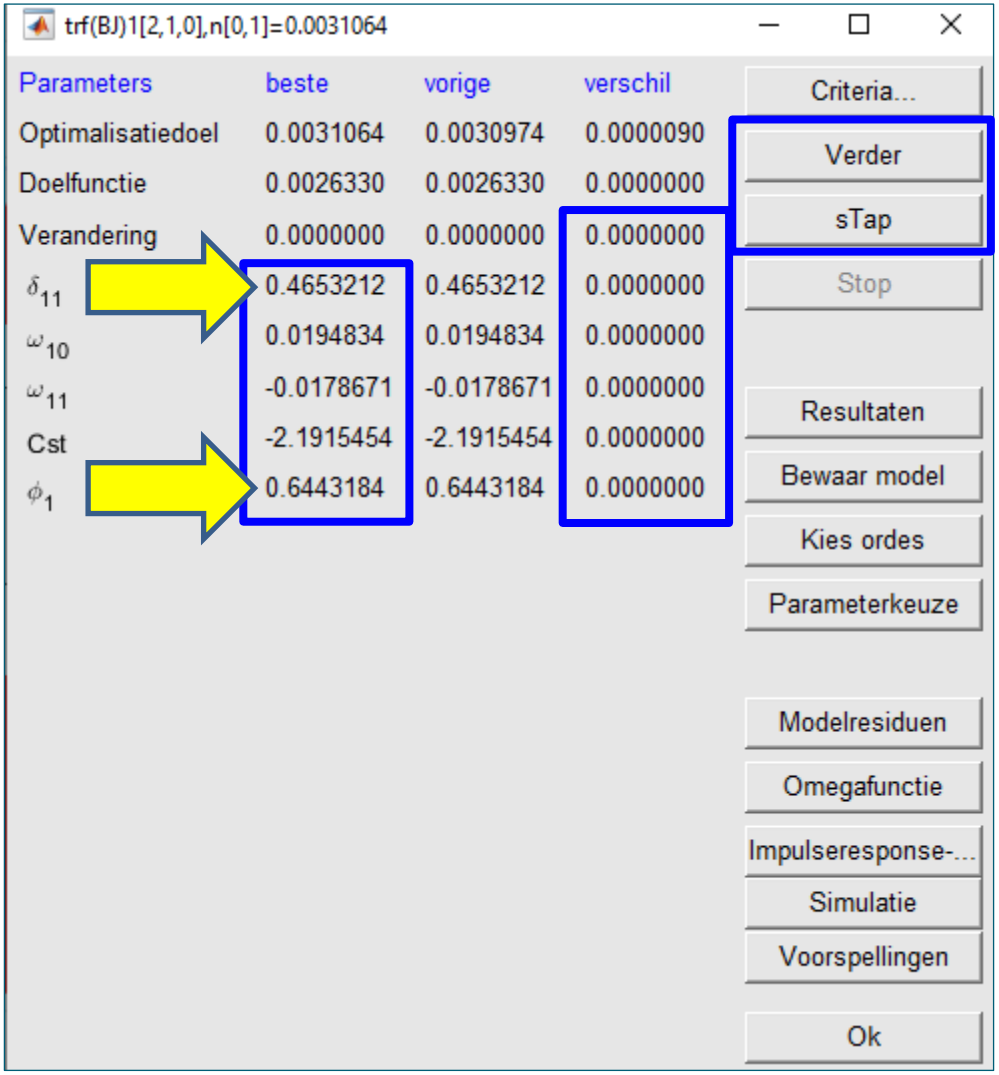


Initiële formulering ruisdeel:

$$N_t - c = \phi_1(N_{t-1} - c) + a_t$$

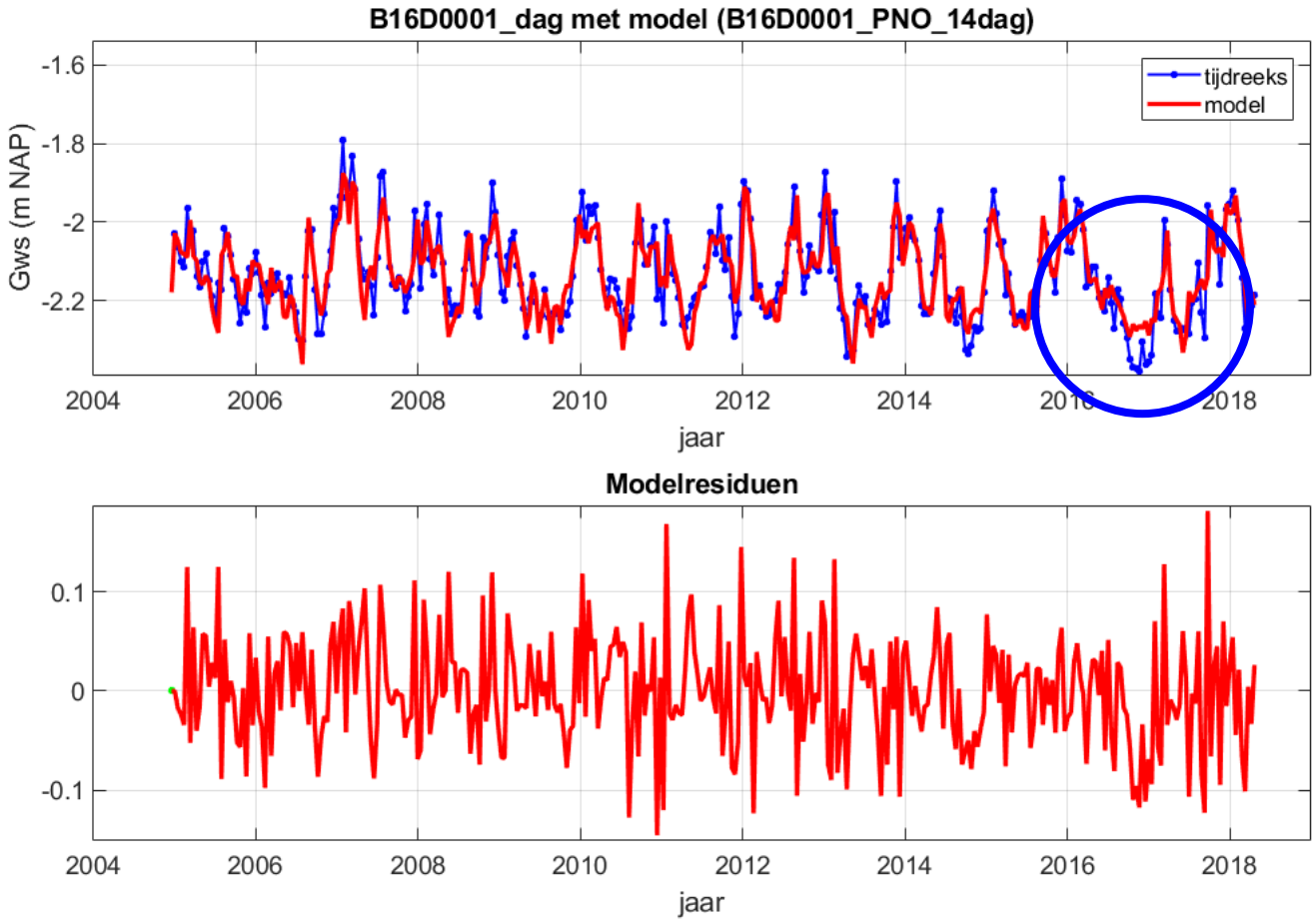
Formulering ruis verifiëren en eventueel aanpassen na 1e modelschatting. Ruis is immers latente variabele, die pas al modellerend tevoorschijn komt.

Parameterschattingen



Parameters	beste	vorige	verschil
Optimalisatiedoel	0.0031064	0.0030974	0.0000090
Doelfunctie	0.0026330	0.0026330	0.0000000
Verandering	0.0000000	0.0000000	0.0000000
δ_{11}	0.4653212	0.4653212	0.0000000
ω_{10}	0.0194834	0.0194834	0.0000000
ω_{11}	-0.0178671	-0.0178671	0.0000000
Cst	-2.1915454	-2.1915454	0.0000000
ϕ_1	0.6443184	0.6443184	0.0000000

Plot reeks, model en residu



Resultaten en diagnoses

De evenwichtsrelatie en verklaarde varianties

Transfermodel	omschrijving/eenheid	evenwichtsrela	standaardfout
PNO_dag(BJ)	PNO (mm/dag)	0.06985590	0.00766664

m per mm/d

Naam	omschrijving/eenheid	variantie	percentage
B16D0001_dag	Gws (m NAP)	0.01275098	
Transfermodel PNO_dag(BJ)	PNO (mm/dag)	0.00872261	62.10 %
Modelruis	verschil tussen meting en simulatie	0.00532411	37.90 %
Modelresiduen	verschil tussen meting en fit	0.00310640	
Akaike FPE	modelresiduen gewogen naar het aantal parameters	0.00319723	
Simulatie	gegenereerd met de invoervariabelen	0.00872261	

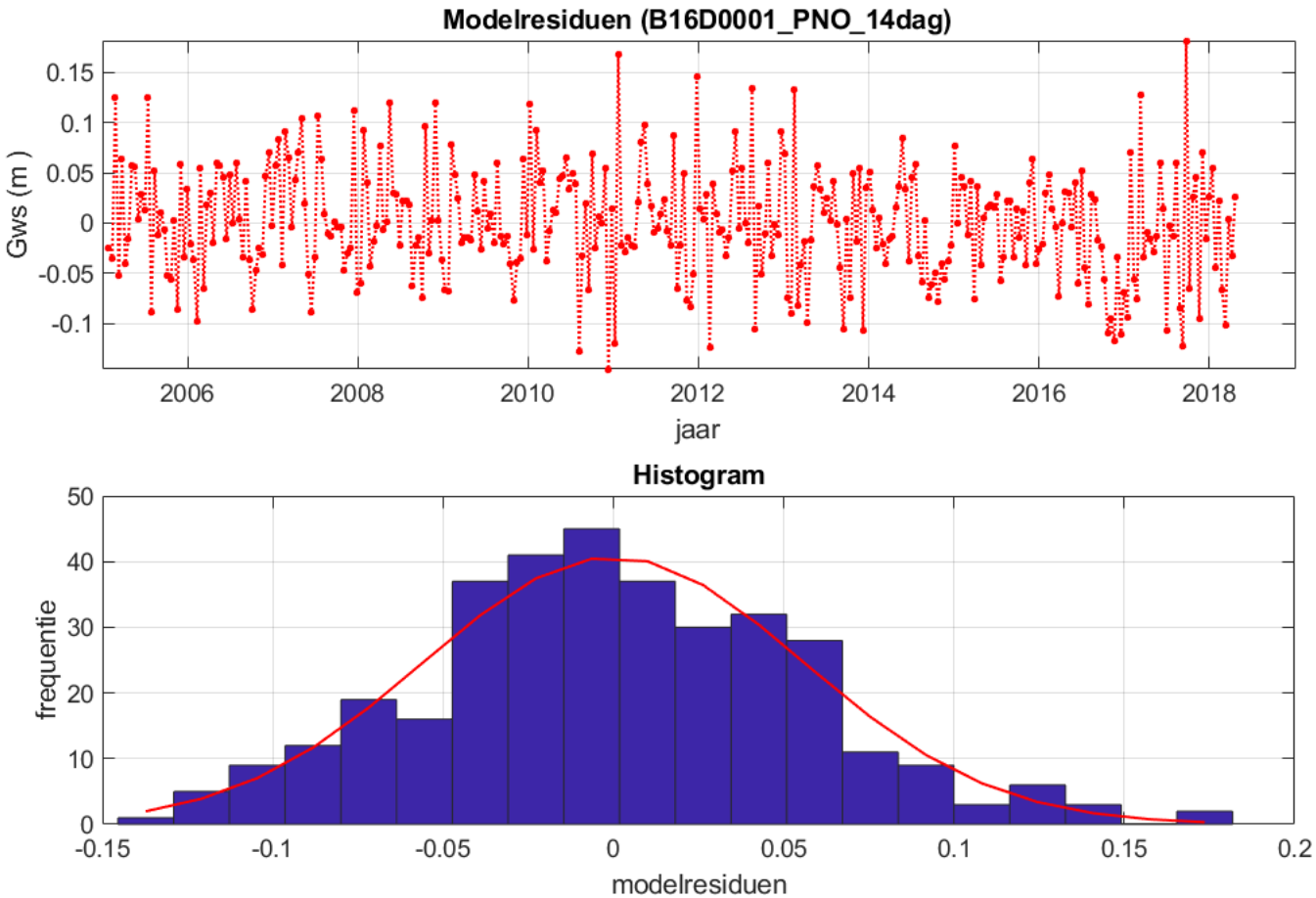
R^2

Naam	omschrijving/eenheid	gemiddelde
Modelresiduen	B16D0001_dag - fit	0.00001586
Modelruis	B16D0001_dag - simulatie	0.00088812

Toets modelresiduen	toetst	uitkomst	detail
Lilliefors-toets	normaal verdeeld (99%)	Ja	p = 0.44
Jacque-Bera-toets	normaal verdeeld (99%)	Ja	p = 0.33
Portmanteau-toets	geen autocorrelatie (95%)	Ja	14.20<23.68 (14)
Runs-toets	geen autocorrelatie (95%)	Ja	p = 0.87
Engle-toets	geen heteroscedasticiteit residuen (95%)	Ja	p = 0.35
Breusch-Pagan-toets	geen heteroscedasticiteit residuen-fit (95%)	Ja	p = 0.62

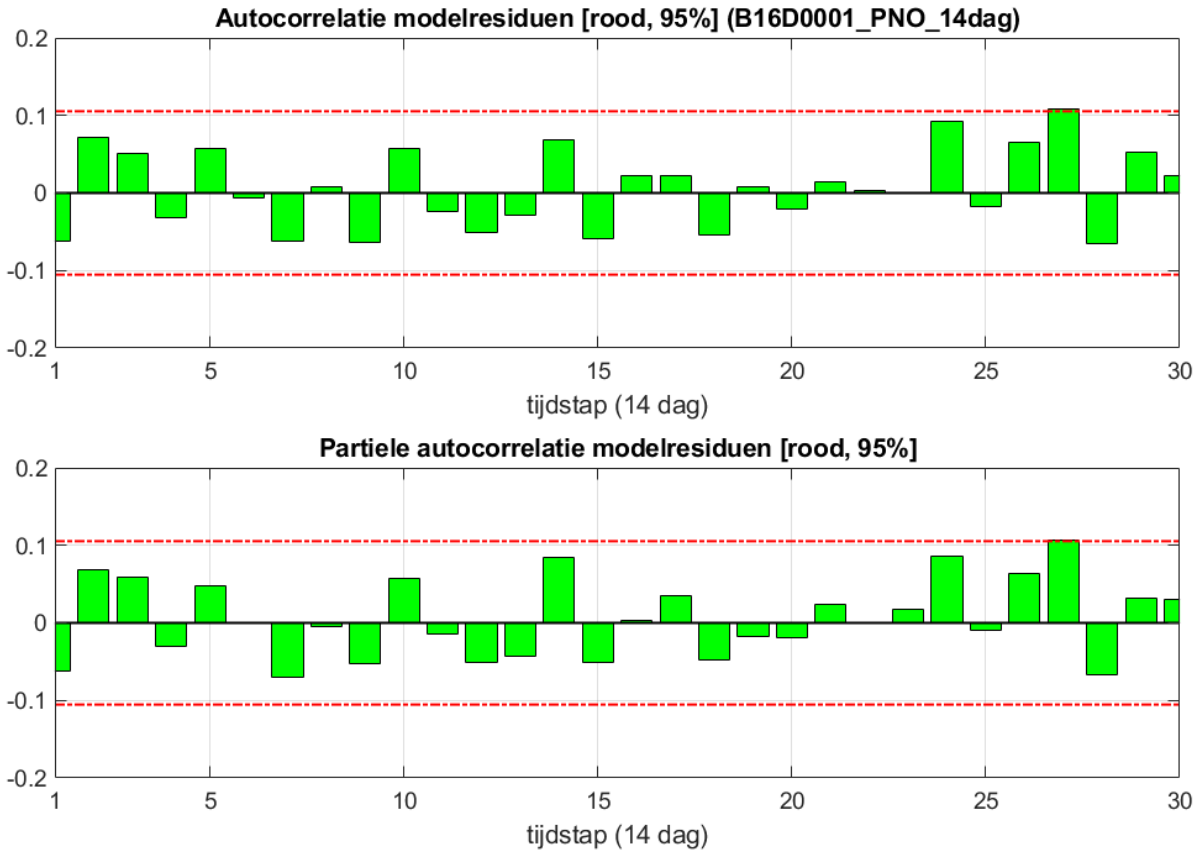
Ok

Modelverificatie



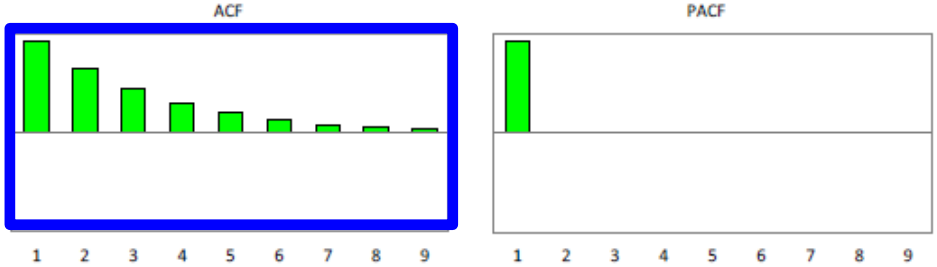
Geen duidelijke scheefheid

Geen autocorrelatie residu a_t?

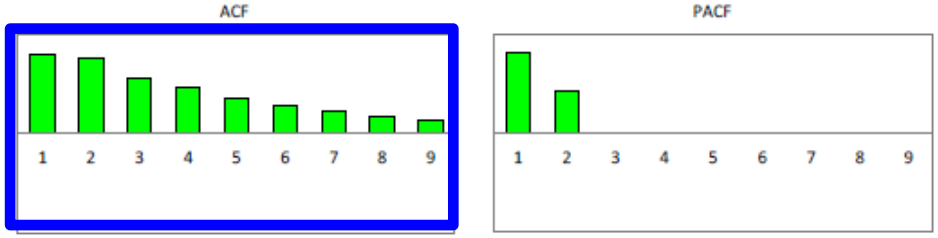


Geen autocorrelatie

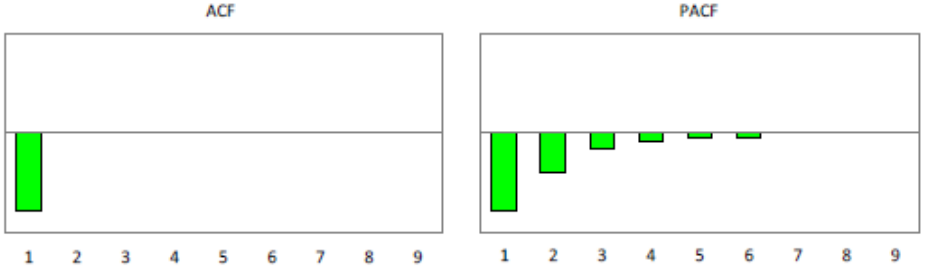
1. AR(1): $N_t = \phi_1 N_{t-1} + a_t$



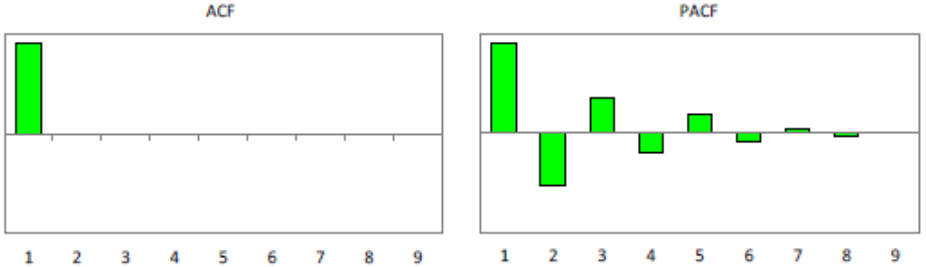
2. AR(2): $N_t = \phi_1 N_{t-1} + \phi_2 N_{t-2} + a_t$



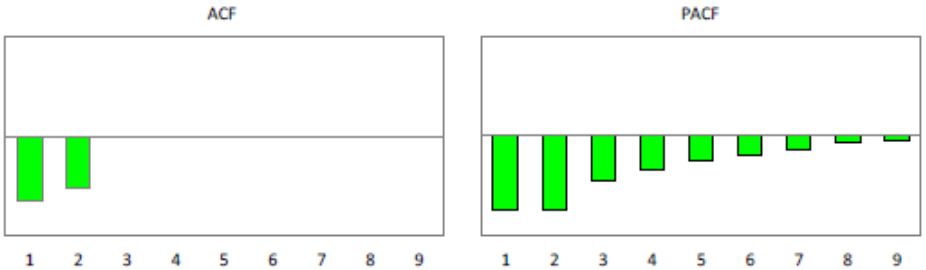
3. MA(1): $N_t = a_t - \theta_1 a_{t-1}$ (bij positieve θ_1)



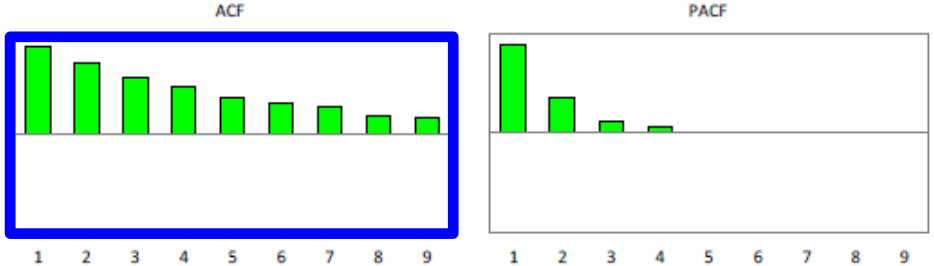
4. MA(1): $N_t = a_t - \theta_1 a_{t-1}$ (bij negatieve θ_1)



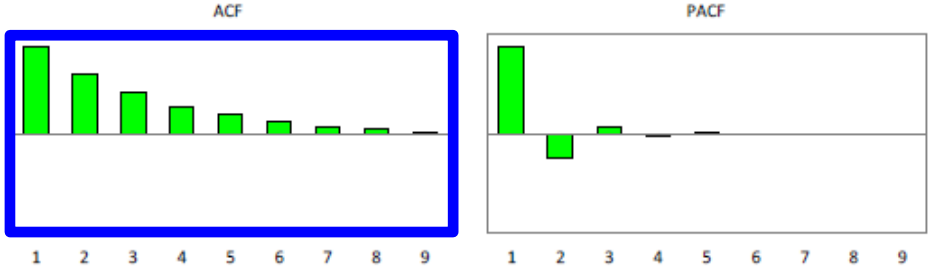
5. MA(2): $N_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$



6. ARMA(1,1): $N_t = \phi_1 N_{t-1} + a_t - \theta_1 a_{t-1}$ (bij positieve θ_1)

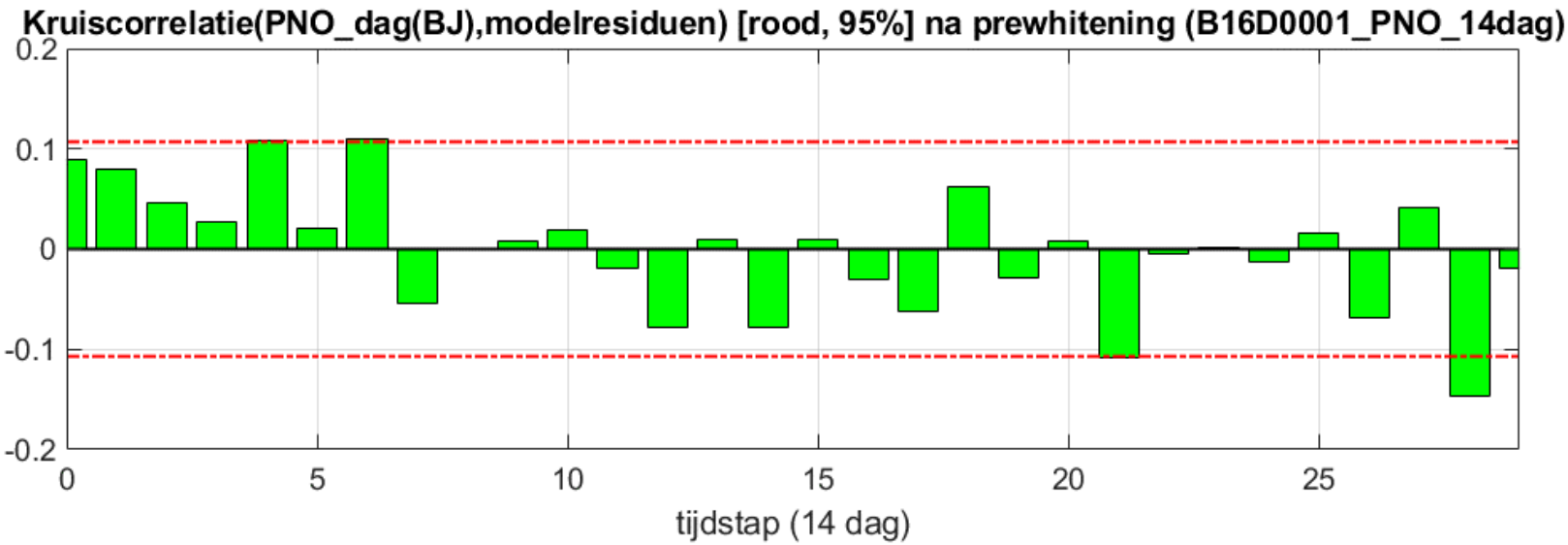


7. ARMA(1,1): $N_t = \phi_1 N_{t-1} + a_t - \theta_1 a_{t-1}$ (bij negatieve θ_1)



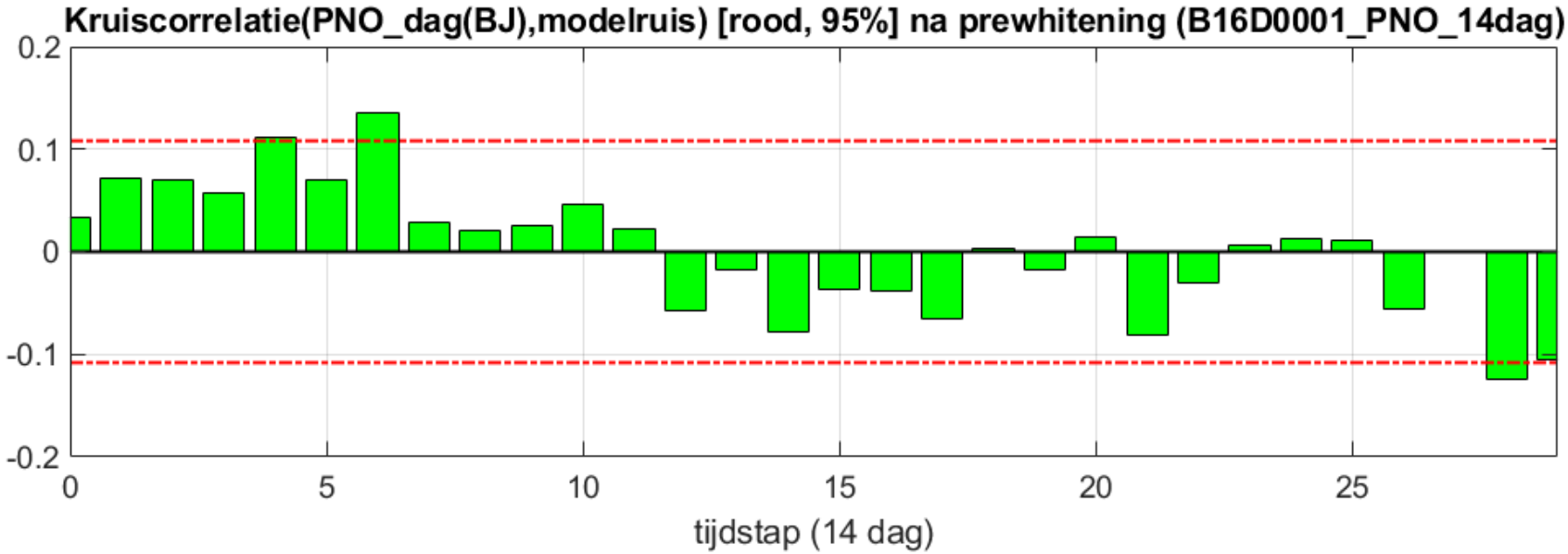
Identificatie ruismodel vergt combinatie van ACF en PACF

Geen kruiscorrelatie PNO en a_t?



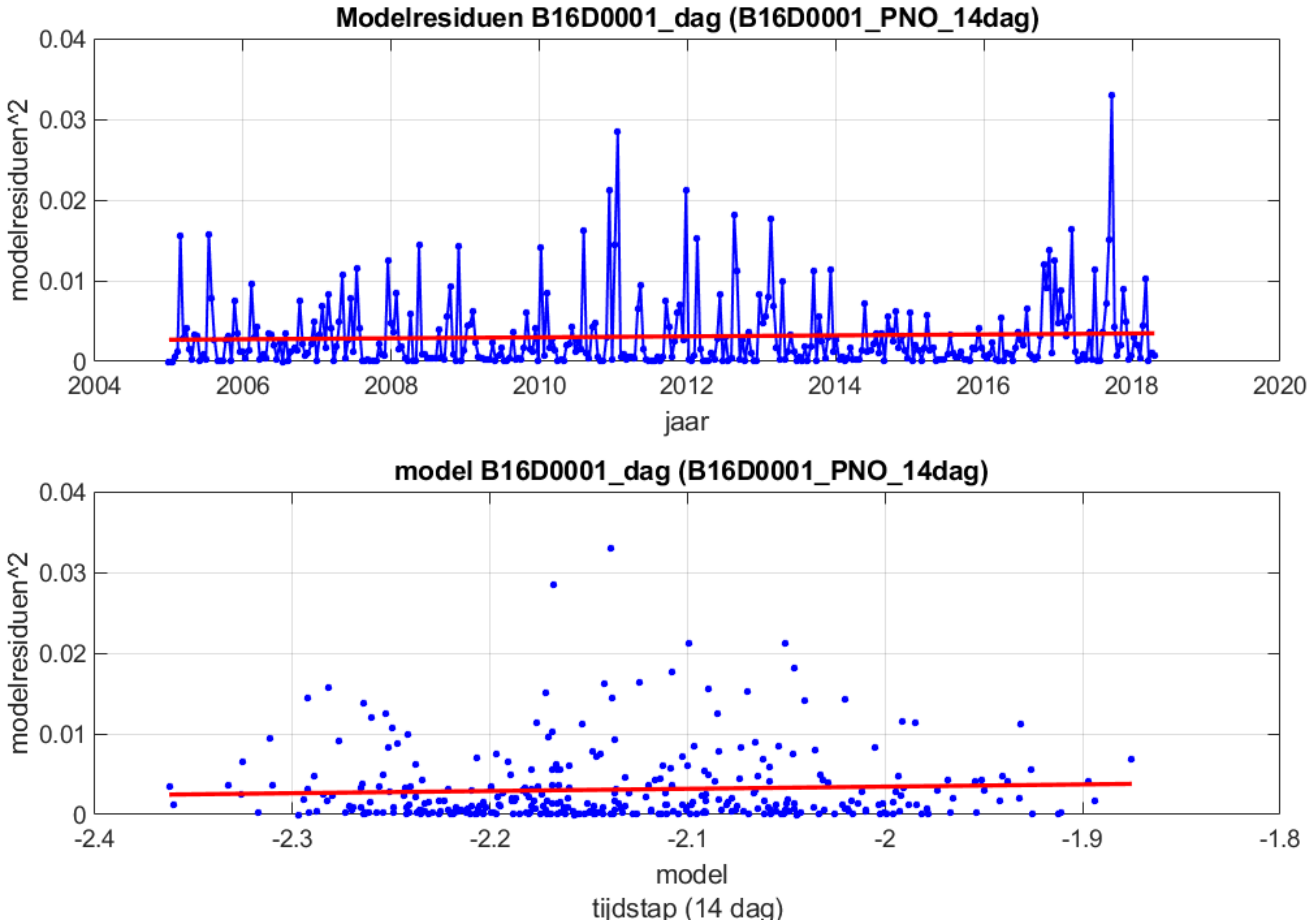
Geen kruiscorrelatie

Geen kruiscorrelatie PNO en N_t?



Geen kruiscorrelatie

Geen heteroscedasticiteit?



Geen heteroscedasticiteit

Voorspellen

- Deterministisch

Alleen gebaseerd op deterministische deel van model

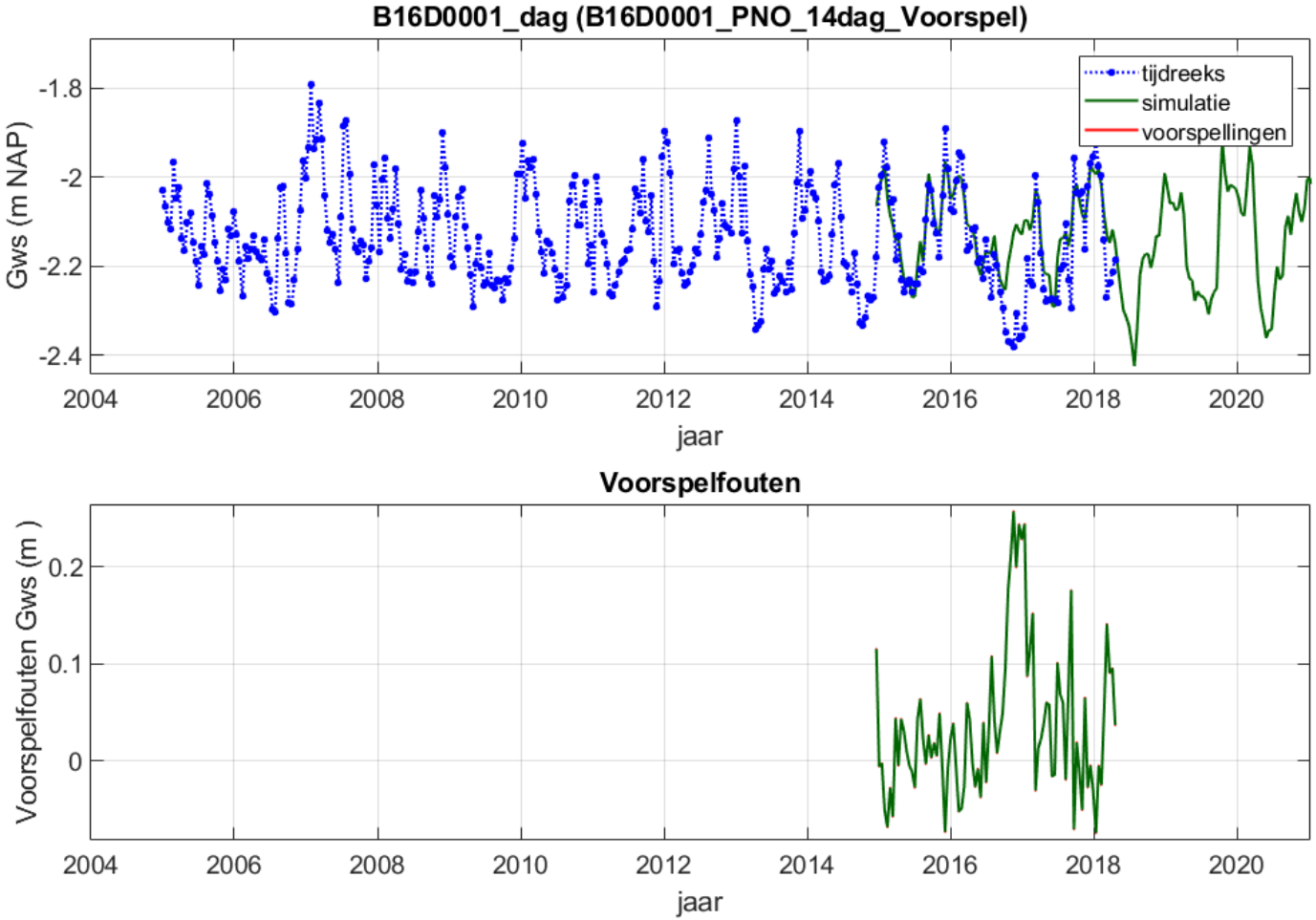
- Deterministisch én stochastisch, dus gebaseerd op álle modelonzekerheden:

  - onzekerheden parameters (met verdiscontering covarianties)

  - ruis N_t (als functie van modelresidu a_t)

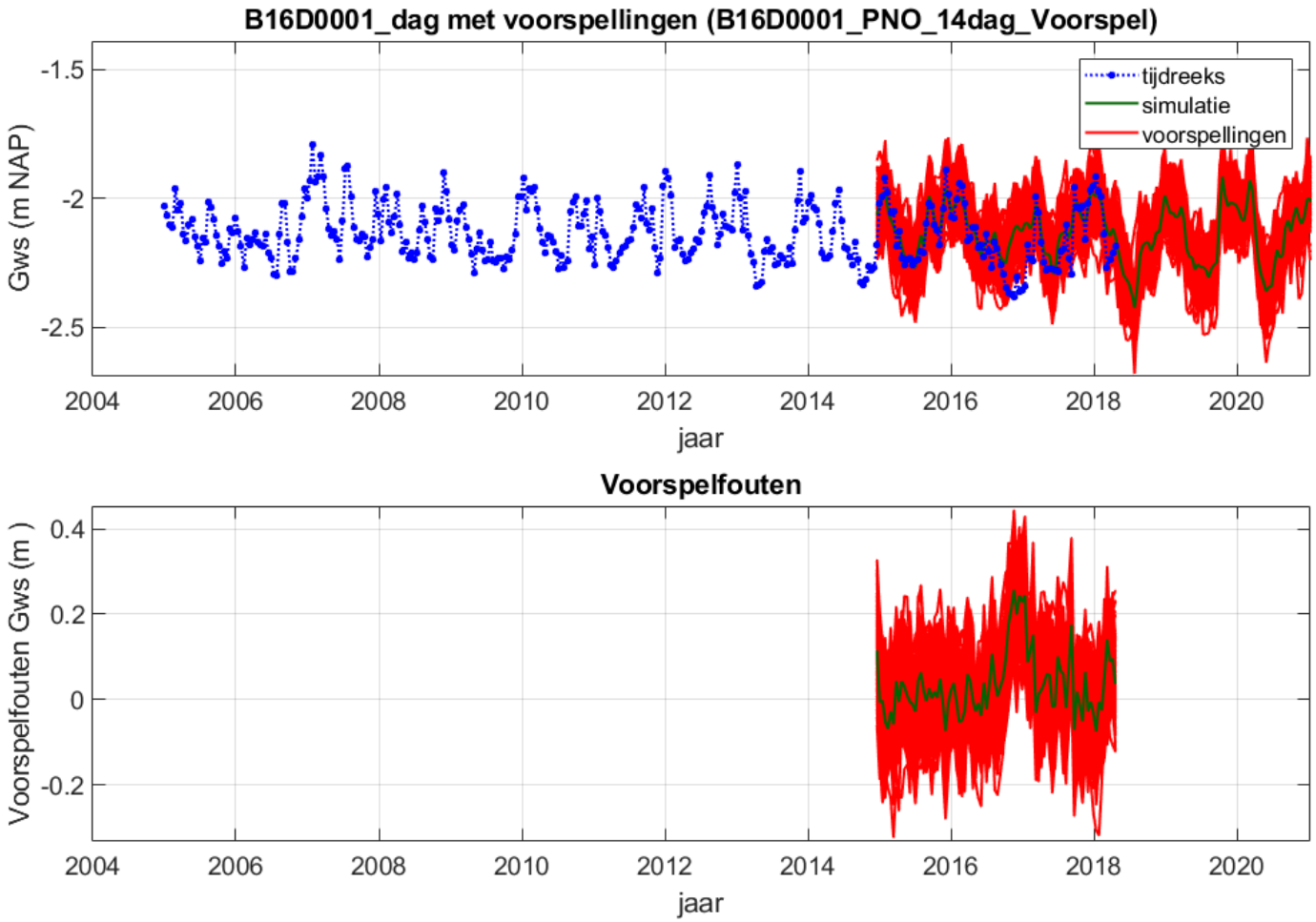
  Eventueel ook meenemen onzekerheden invoerreeksen

Deterministisch voorspellen



GHG en GLG onbetrouwbaar

Stochastisch voorspellen



Waar maak ik me zorgen om?

- Software is teveel gaan bepalen hoe wordt geanalyseerd
- Meer hydrologische beoefenaars, maar velen met nog maar beperkte kennis van achtergronden en valkuilen
- Doorgaans is er ook te weinig kennis van statistiek
- Geen of weinig benutting van de grote meerwaarde van het meegeven van modelonzekerheden
- Het percentage verklaarde variantie is een soort graadmeter voor de modelkwaliteit geworden (70%-grens), maar is daar totaal niet voor geschikt

Waar maak ik me zorgen om?

- Opdrachtgevers beschikken vaak niet over de kennis om de zeggingskracht en bruikbaarheid van modelresultaten te kunnen beoordelen
- In de softwarehoek ligt de focus nog veel op oplossingen voor niet-equidistante tijdreeksen, terwijl die steeds minder voorkomen

Hoe word je een betere analist?

1. Lees de Handleiding (van kaft tot kaft) en andere literatuur
2. Ga zelf verschillende soorten tijdreeksen genereren (univariaat en TFR) - bv met Excel – en zie wat verandert door het wijzigen van parameters
3. Maak altijd écht contact met een tijdreeks, ervaar de reactie op modelwijzigingen, voel of het model robuust is en ga na of het voldoet aan de vooronderstellingen
4. Onderzoek wat er gebeurt als je het interval wijzigt
5. Laat niet de software je verkenning en je contact met tijdreeksen beperken, gebruik flexibele toolboxen
6. Kijk eens verder en ervaar hoe in andere vakgebieden tijdreeksanalyses worden gedaan, zoals econometrie

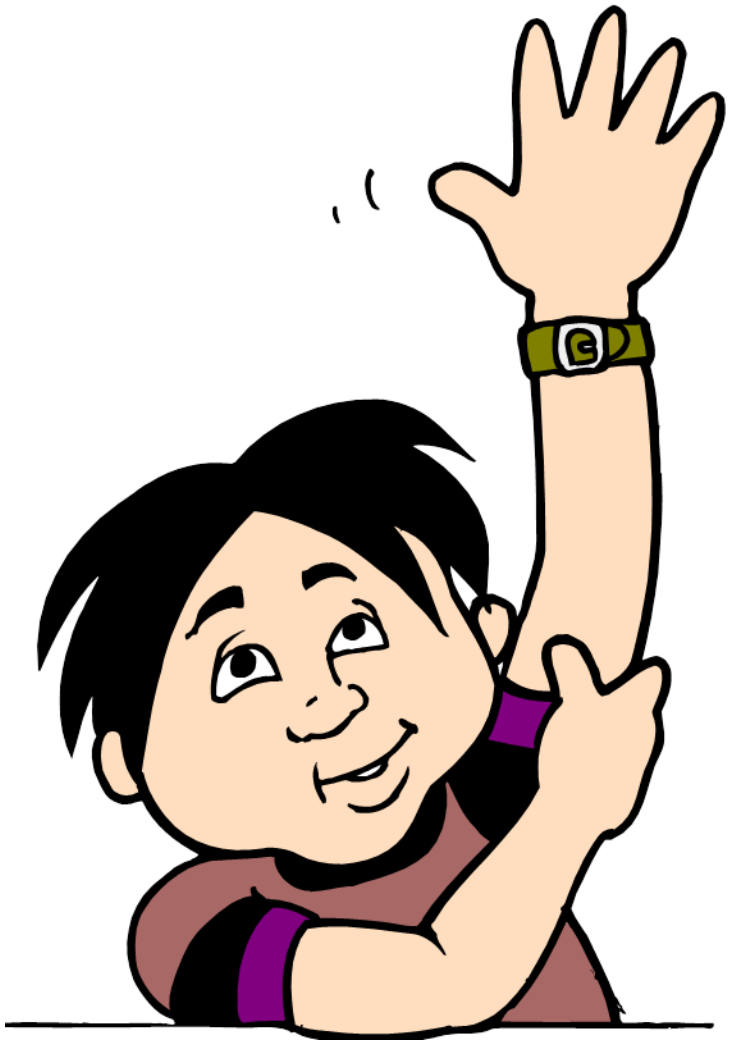
Slotadviezen

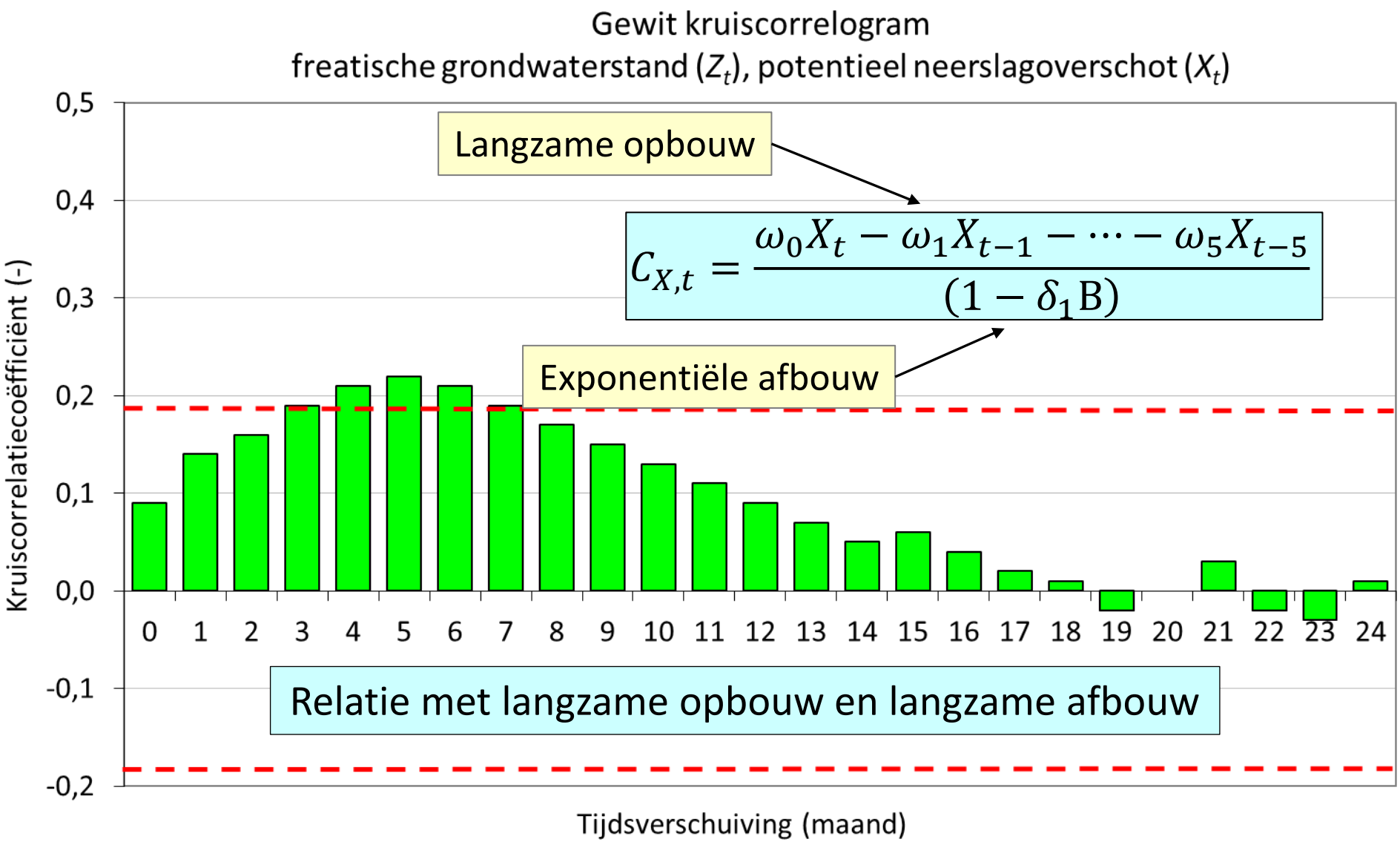
- Tijdreeksen zijn kostbaar en kunnen relevante informatie verschaffen, ga dan niet per sé besparen op hun verwerking (zoals bij bulkverwerking)
- Pas op voor subjectieve onzekerheidsreductie, laat je dat ook niet opdringen, je zult er beter door slapen
- Een statistisch significante relatie is nog geen hard bewijs voor een causale relatie: vermeld dat altijd!

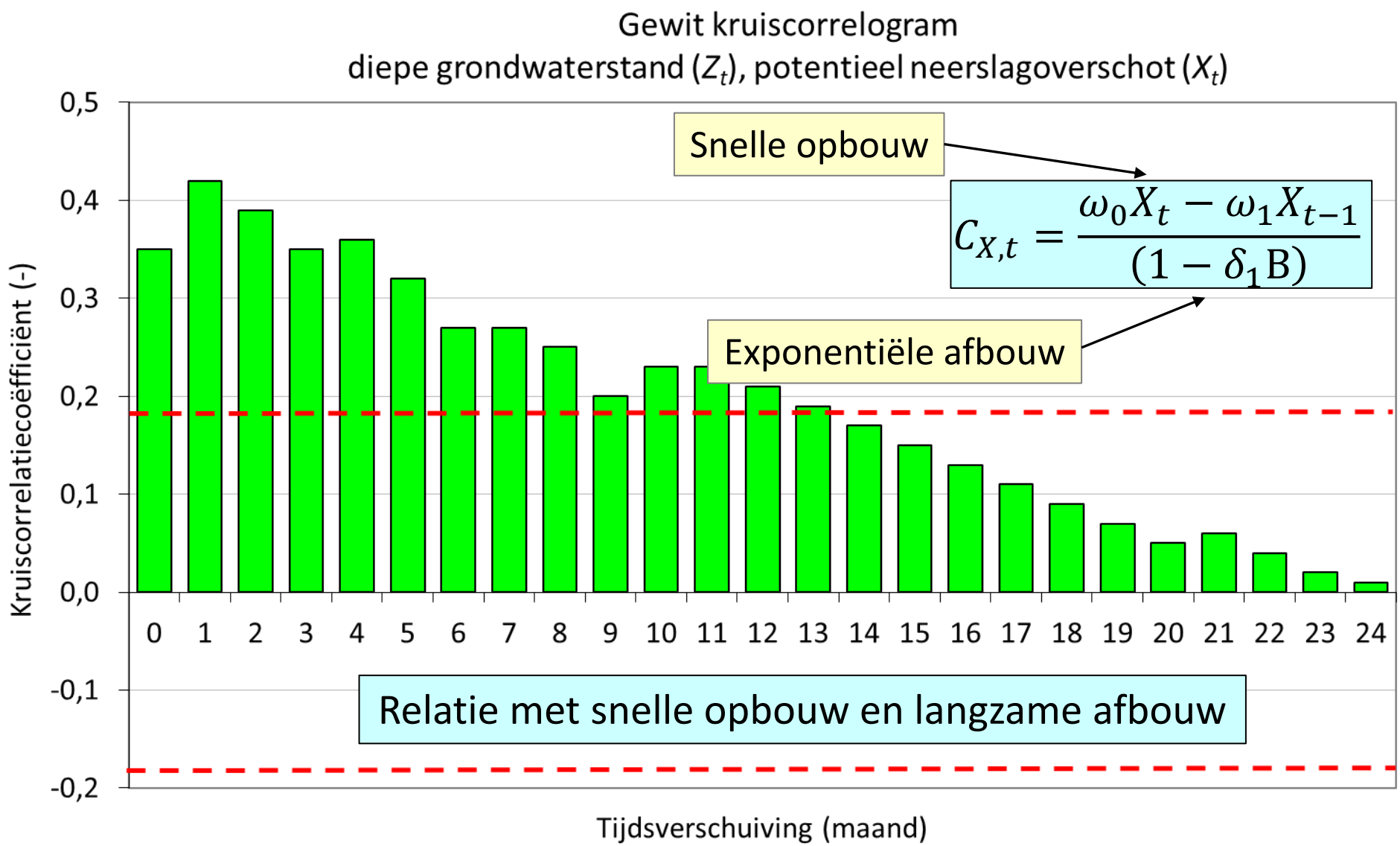
Houd van de data, wantrouw je model

en dus beslist niet andersom!

Vragen?




Gewit kruiscorrelogram
freatische grondwaterstand (Z_t), potentieel neerslagoverschot (X_t)
Langzame opbouw
$$C_{X,t} = \frac{\omega_0 X_t - \omega_1 X_{t-1} - \cdots - \omega_5 X_{t-5}}{(1 - \delta_1 \mathrm{B})}$$
Exponentiële afbouw
Relatie met langzame opbouw en langzame afbouw
Kruiscorrelatiecoëfficiënt (-)
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
-0,2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Tijdsverschuiving (maand)


Gewit kruiscorrelogram
diepe grondwaterstand (Z_t), potentieel neerslagoverschot (X_t)
Kruiscorrelatiecoëfficiënt (-)
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
-0,2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Tijdsverschuiving (maand)
Snelle opbouw
$$C_{X,t} = \frac{\omega_0 X_t - \omega_1 X_{t-1}}{(1 - \delta_1 \mathrm{B})}$$
Exponentiële afbouw
Relatie met snelle opbouw en langzame afbouw