Belang van ruismodel bij tijdreeksmodellering

NHV-werkgroep Tijdreeksanalyse

Discussiemiddag
1 oktober 2015

Paul Baggelaar

Algemene vorm tijdreeksmodel

deterministische component

$$Y_t = \boxed{f(P,X,t)} + \boxed{N_t}$$

stochastische component (ruis)

ruismodel

Autoregressieve deel

'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

Nut van ruismodel

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

2. Kan **objectieve** kwantificering onzekerheden uitspraken mogelijk maken

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

Principe ruismodellering

Breng stochastische component terug tot verschijnsel dat:

- minimaal is
- bekende waarschijnlijkheidswetten volgt

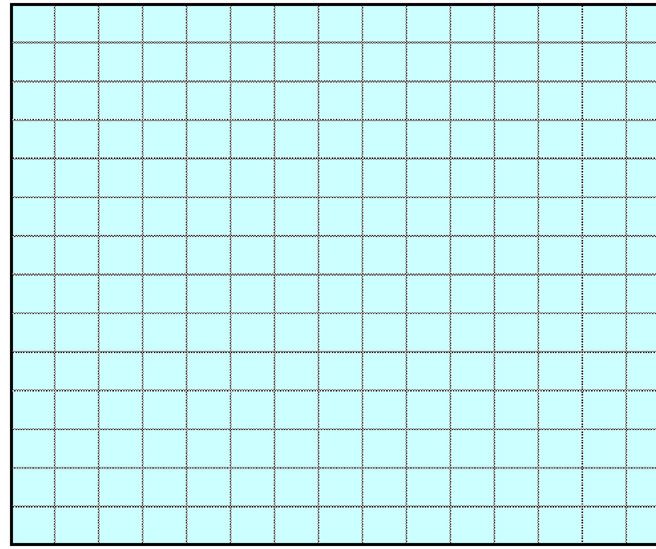
Daardoor kunnen op basis van het model uitspraken worden gedaan met:

- minimale en
- kwantificeerbare onzekerheden

De kracht van statistiek - 1

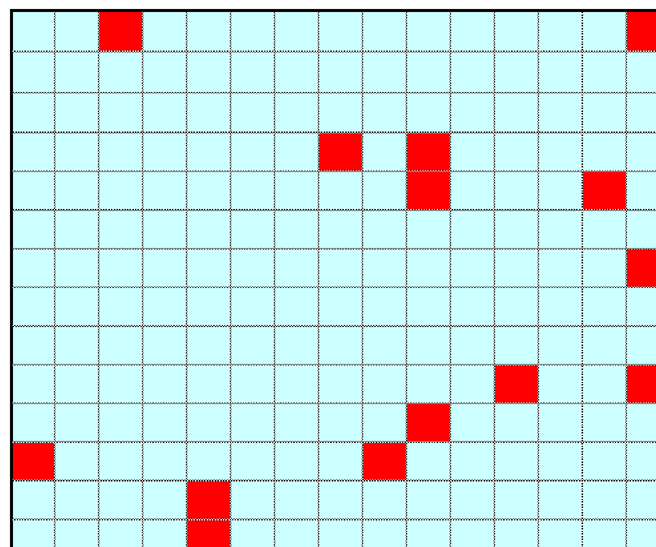
Populatie

?



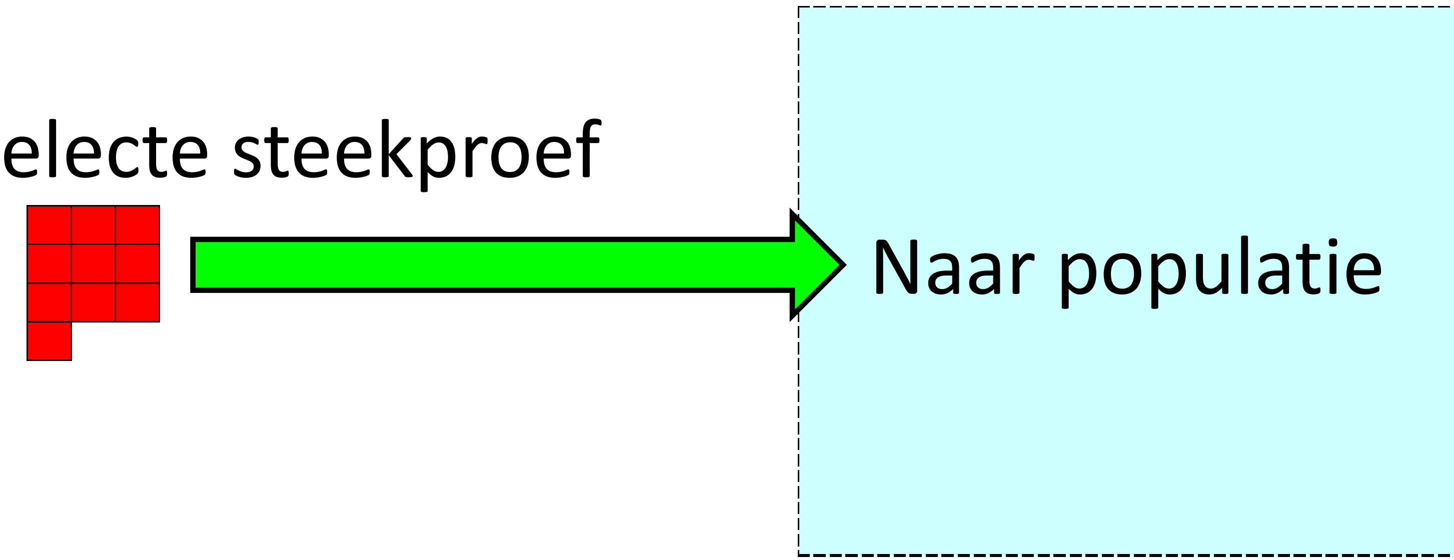
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135
136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160	161	162	163	164	165
166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190	191	192	193	194	195
196	197	198	199	200	201	202	203	204	205	206	207	208	209	210

3
53
15
105
55
185
150
70
200
174
160
147
166
74



De kracht van statistiek - 2

Van aselecte steekproef

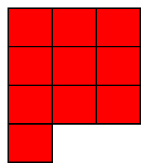


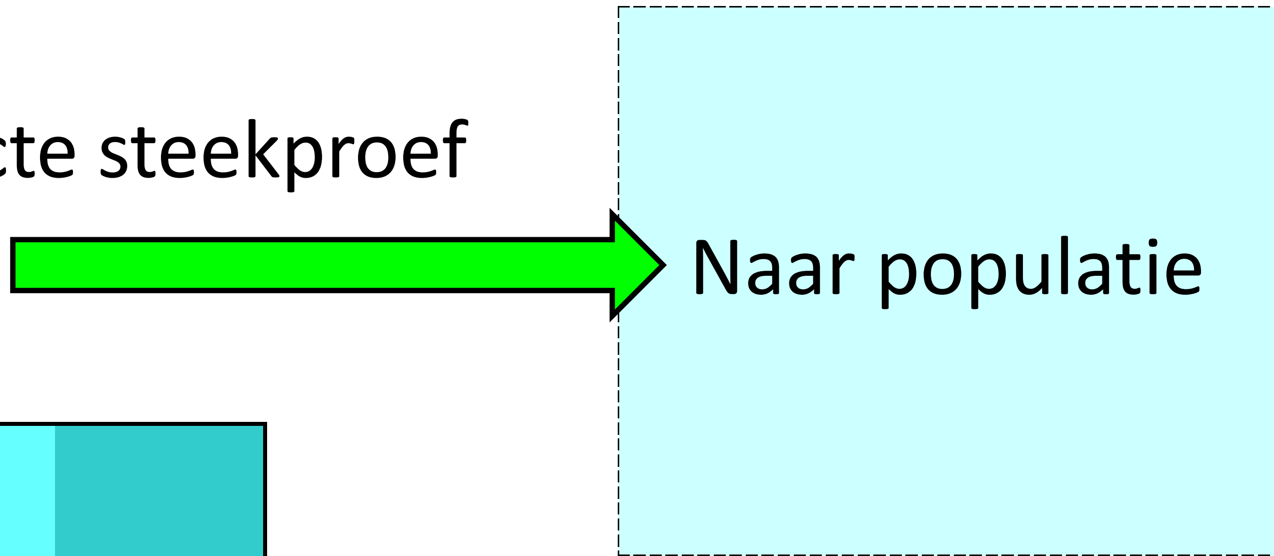
Onverdachte / objectieve uitspraken over populatie, met:

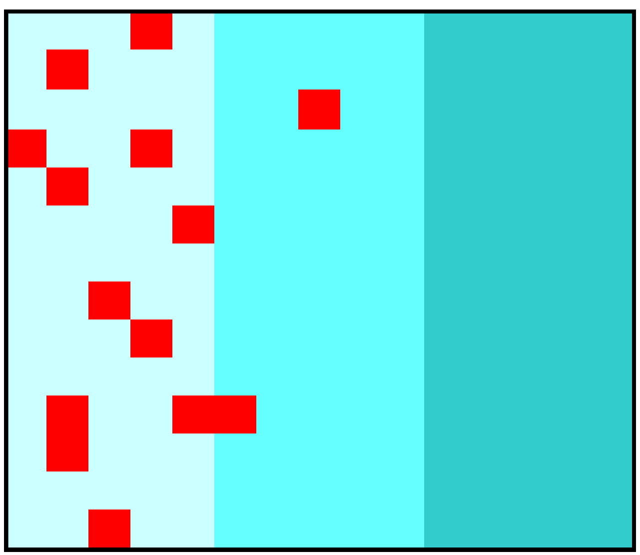
- statistische significanties
- betrouwbaarheden

Een valkuil van statistiek

Niet-aselecte steekproef







Objectieve kwantificering

- statistische significanties
- betrouwbaarheden

is **niet** meer mogelijk

Eenvoudig tijdreeksmodel

onderzoeksvariabele

intercept

modelresidu

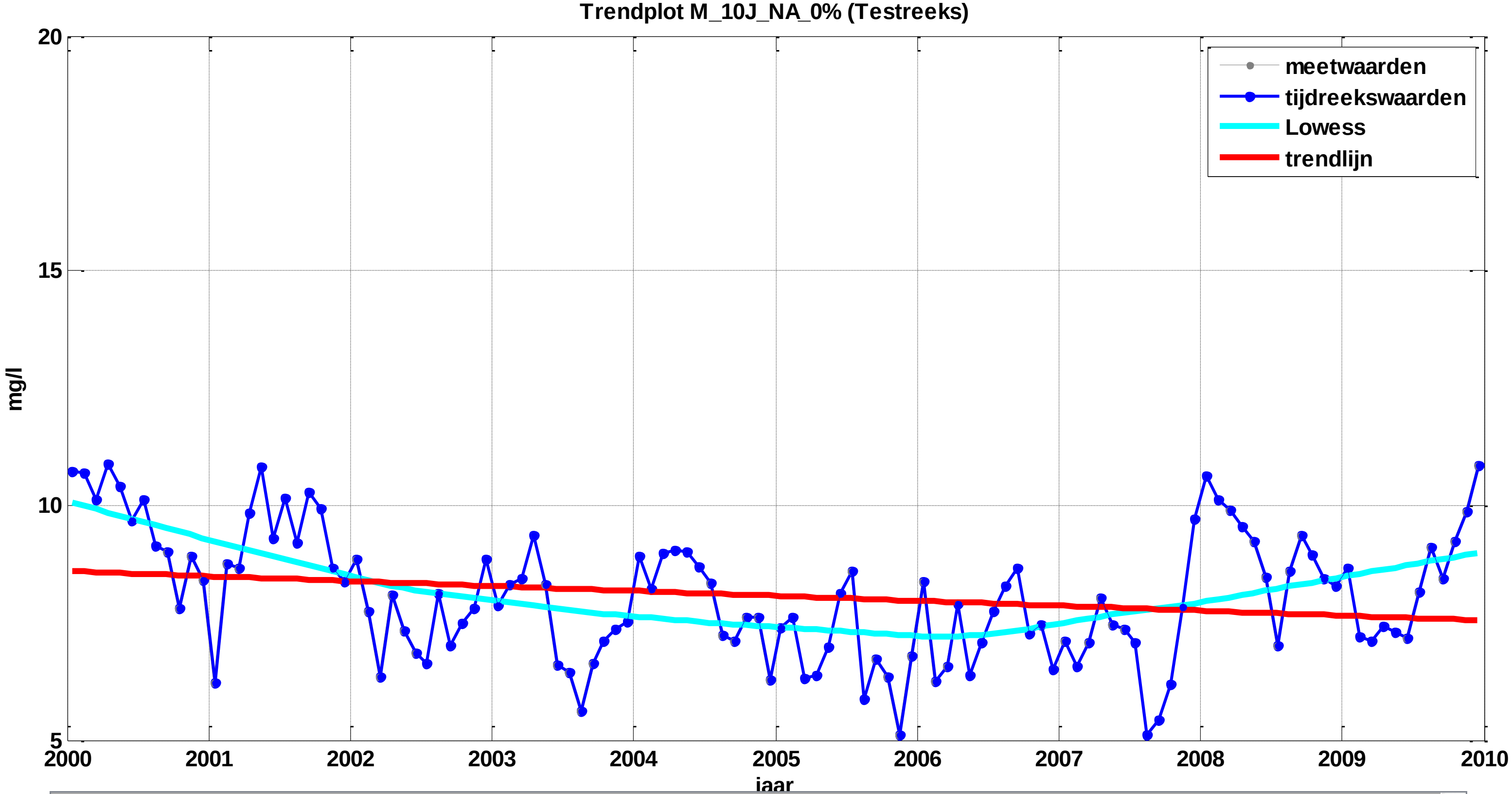
$$Y_t = b_0 + b_1 . X_t + a_t$$

tijdsindex

helling

tijd

Voorbeeld lineaire regressie



Resultaat LR-toets is
Oordeel = kleine trend
De geschatte trend = -0.10504 eenheden per jaar.

Significantie en betrouwbaarheid

Significantie:

$$p = \text{Kans}\left[t_{97,5\%;n-2} > \frac{|b_1| \cdot \sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2}}{s_a}\right]$$

Student-t-waarde

Standafw modelresidu

95%-betr.interval:

$$b_1 \pm t_{97,5\%;n-2} \cdot \frac{s_a}{\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2}}$$

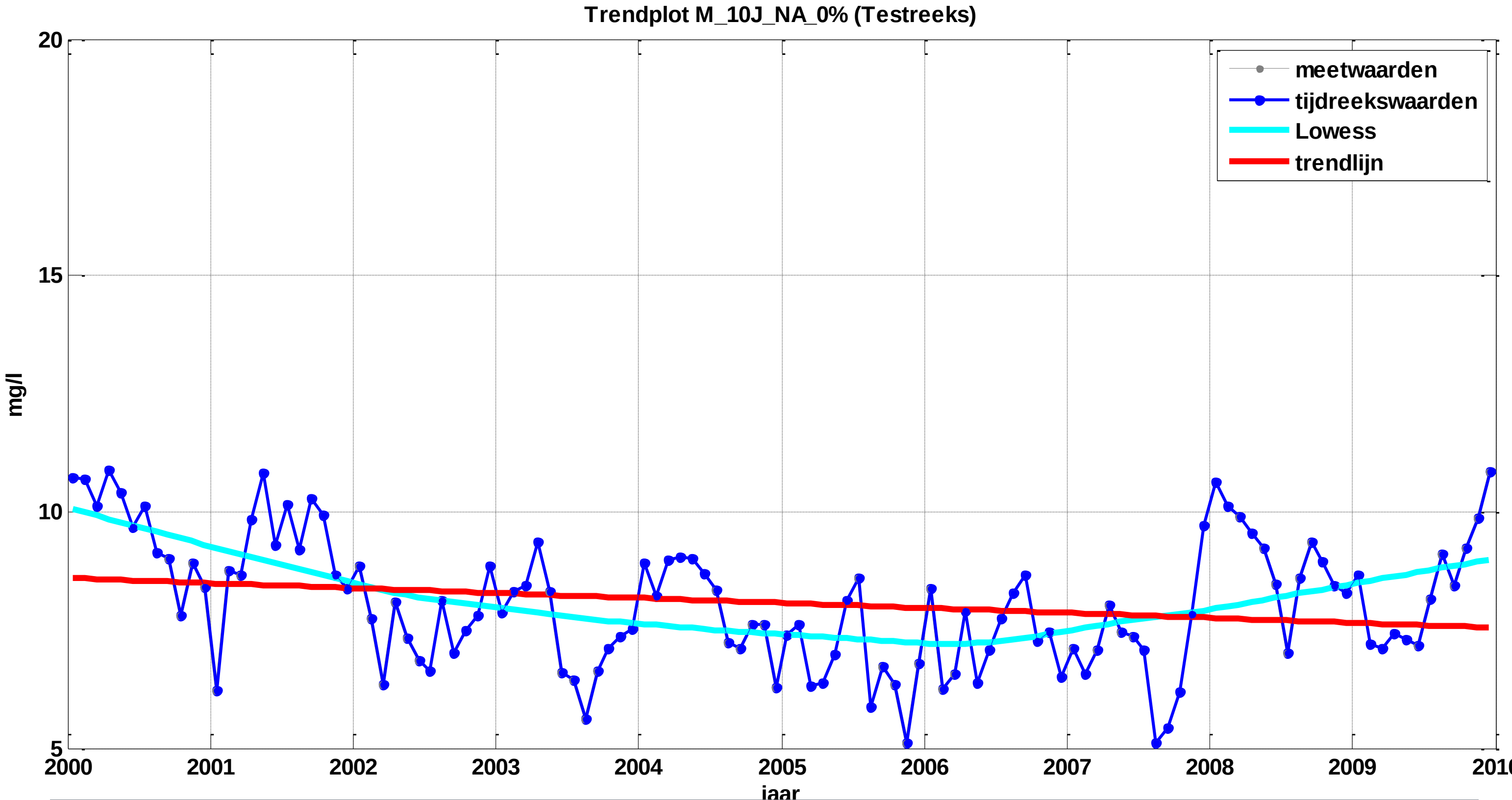
Significantie en betrouwbaarheid

Berekening veronderstelt Gaussiaanse witte ruis

Modelresiduën zijn dan:

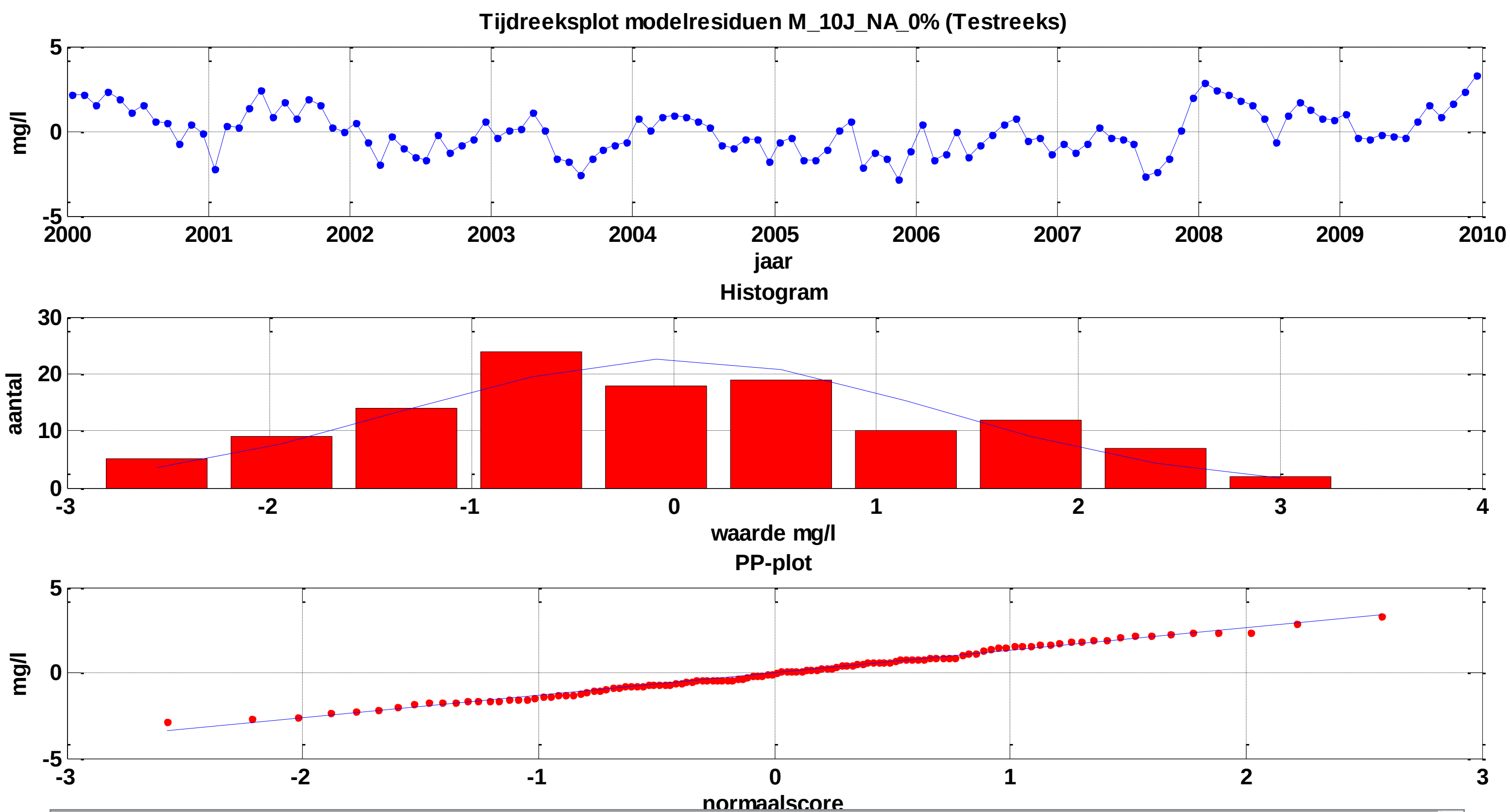
- afkomstig uit zelfde normale kansverdeling
- gemiddeld nul
- onafhankelijk van elkaar
- onafhankelijk van modelwaarden

Voorbeeld lineaire regressie



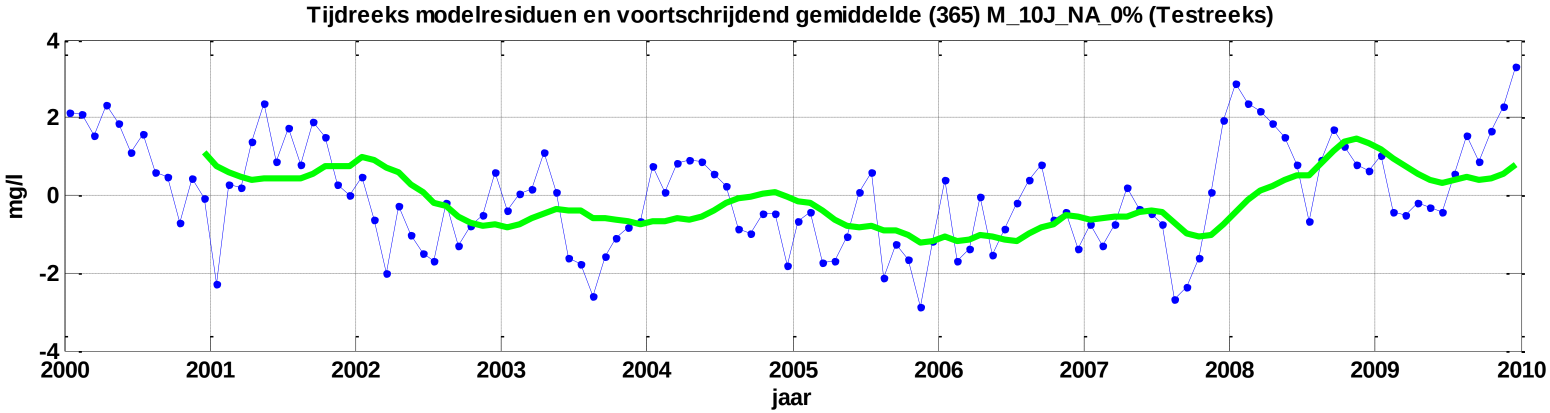
Resultaat LR-toets is
Oordeel = kleine trend
De geschatte trend = -0.10504 eenheden per jaar.

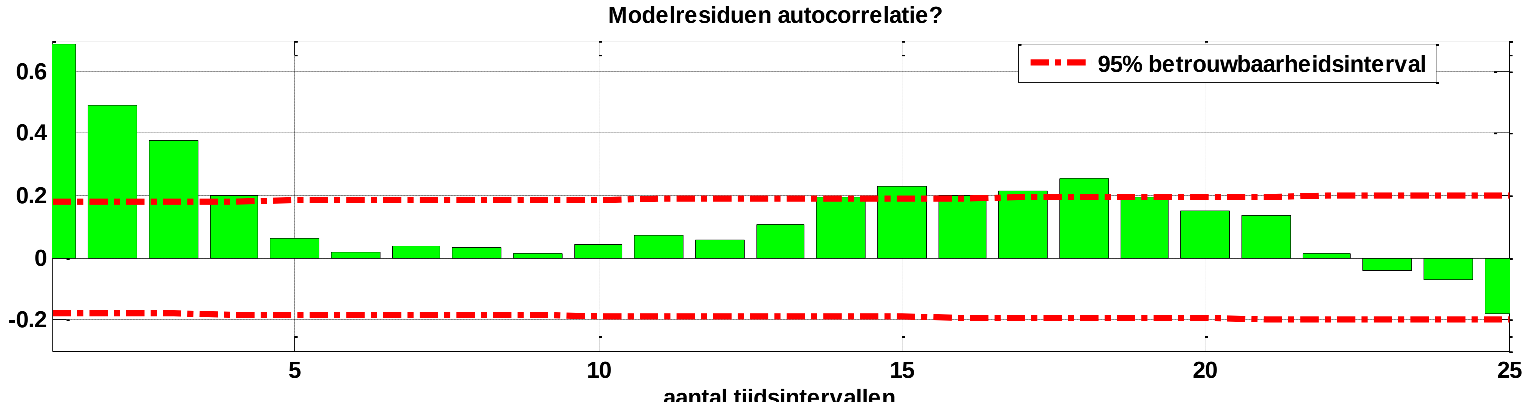
Modelresiduën normaal verdeeld?



Resultaat Lilliefors-toets op normaliteit (#120)is
p-waarde = 50.0%.
We gaan uit van niet-normaliteit bij p < 5%.

Modelresiduën geen autocorrelatie?





Resultaat Portmanteau-toets op autocorrelatie van residuen (#120) is
Qm-waarde = 126.6525, met 15 vrijheidsgraden.
Met 95% betrouwbaarheid gaan we uit van autocorrelatie als Qm > 24.9958

Uitgebreide lineaire regressie

$$Y_t = b_0 + b_1 \cdot X_t + N_t$$

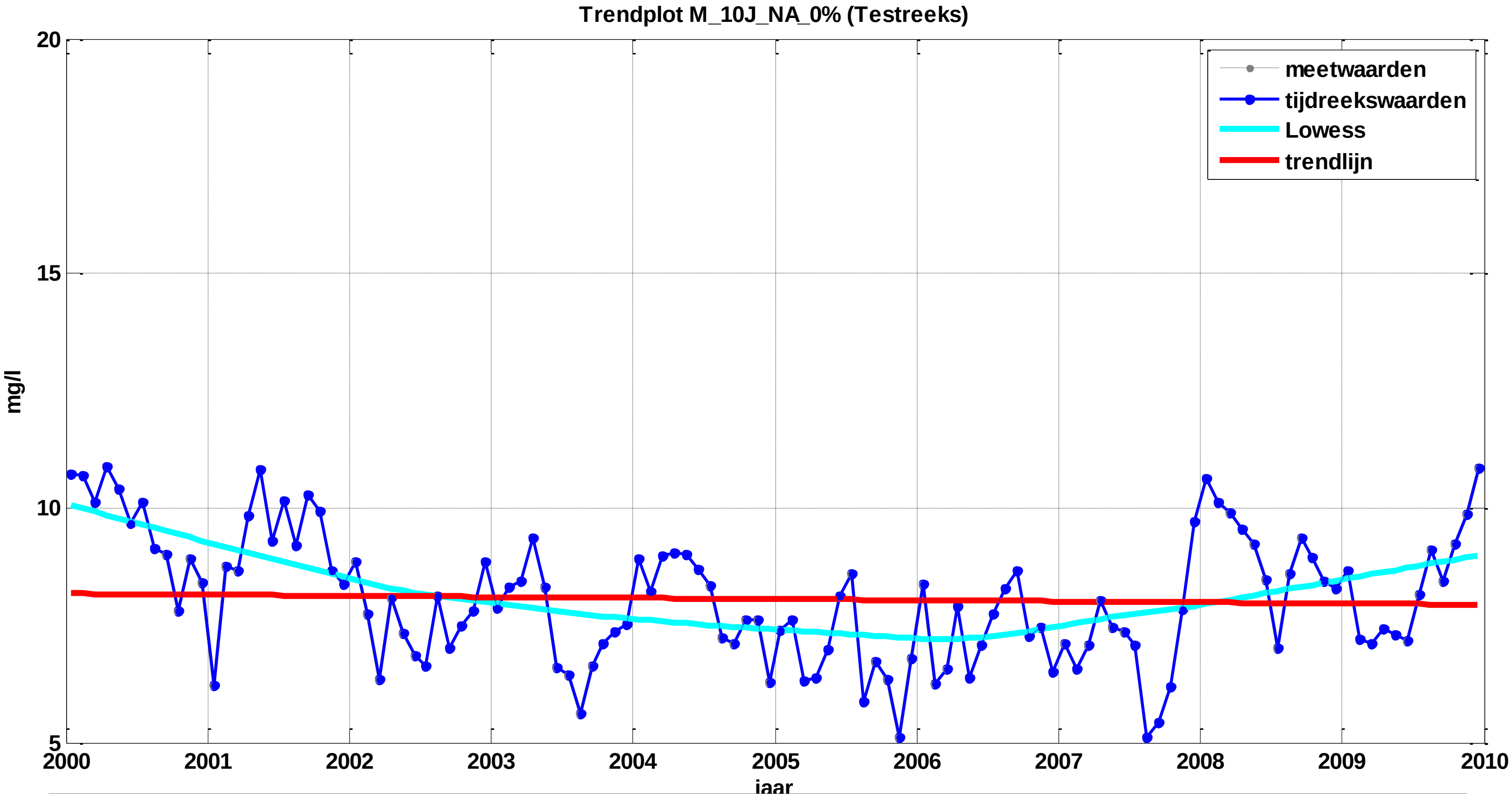
ruis

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

autoregressieve modelparameter

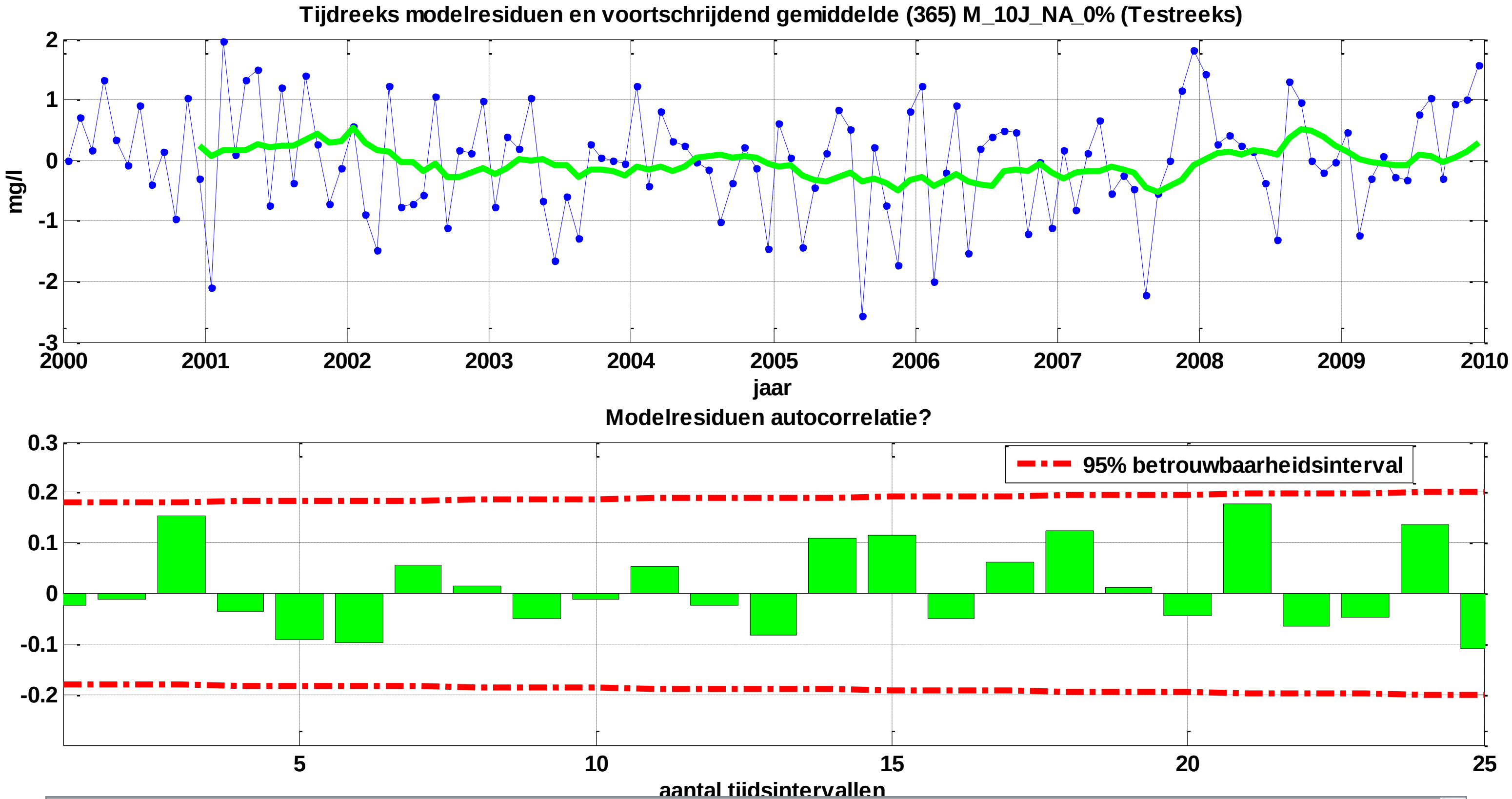
modelresidu

Voorbeeld uitgebreide lineaire regressie



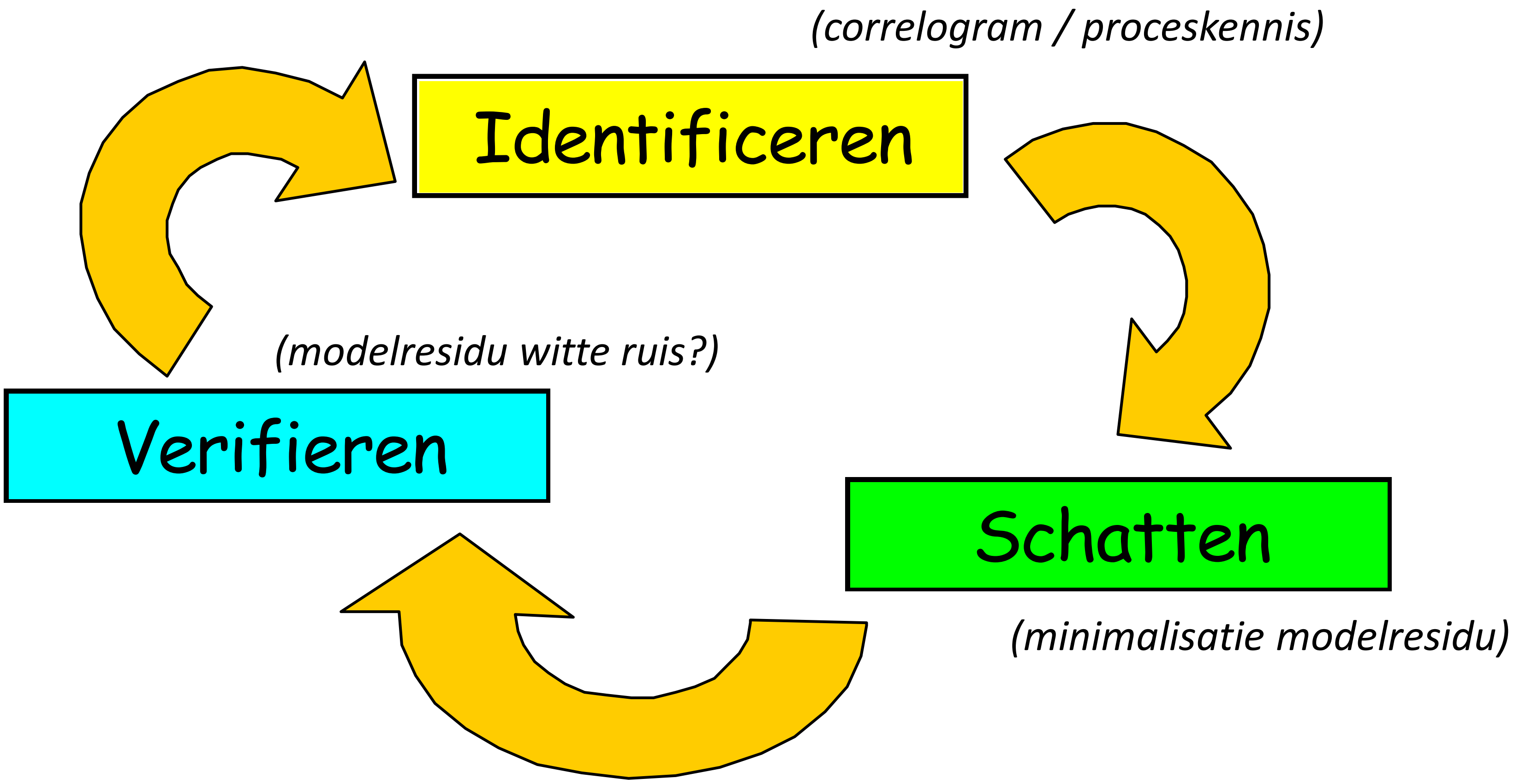
Resultaat LRa-toets is
Oordeel = geen trend
De geschatte trend = -0.023608 eenheden per jaar.

Modelresiduën geen autocorrelatie?



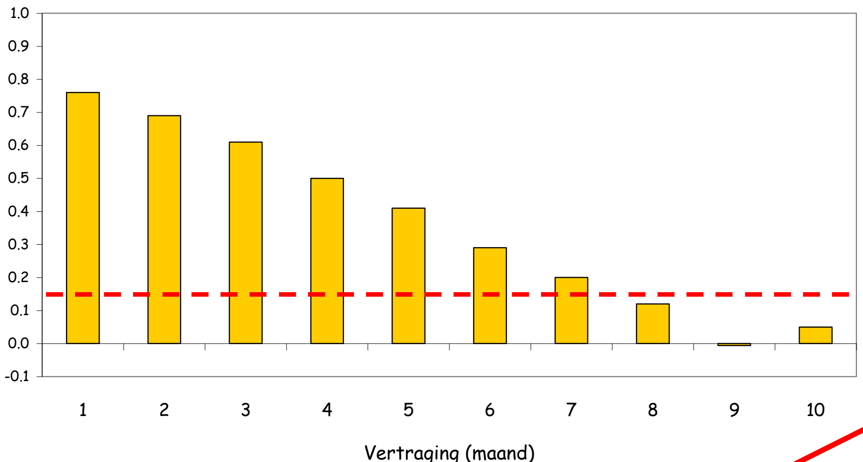
Resultaat Portmanteau-toets op autocorrelatie van residuen (#120) is
Qm-waarde = 11.2214, met 14 vrijheidsgraden.
Met 95% betrouwbaarheid gaan we uit van autocorrelatie als Qm > 23.6848

Modelleringscyclus

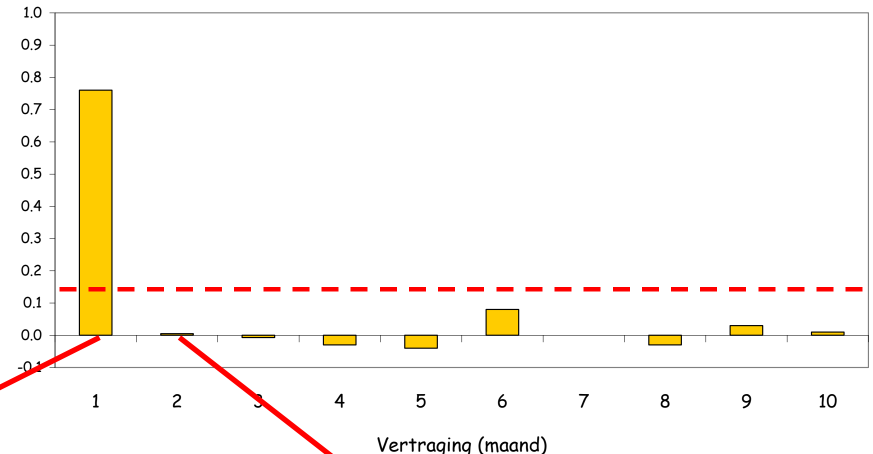


Identificatie ruismodel

autocorrelogram



partieel autocorrelogram



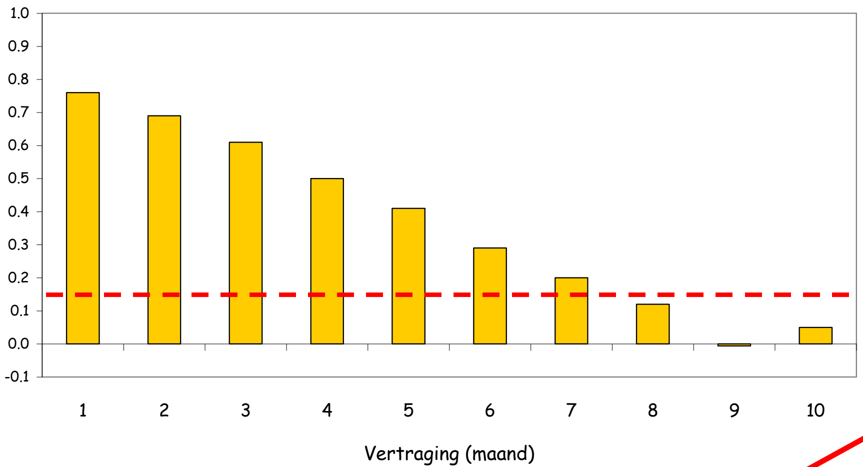
$$\boxed{N_t = \phi_{11} N_{t-1} + a_t}$$

$$\boxed{N_t = \phi_{21} N_{t-1} + \phi_{22} N_{t-2} + a_t}$$

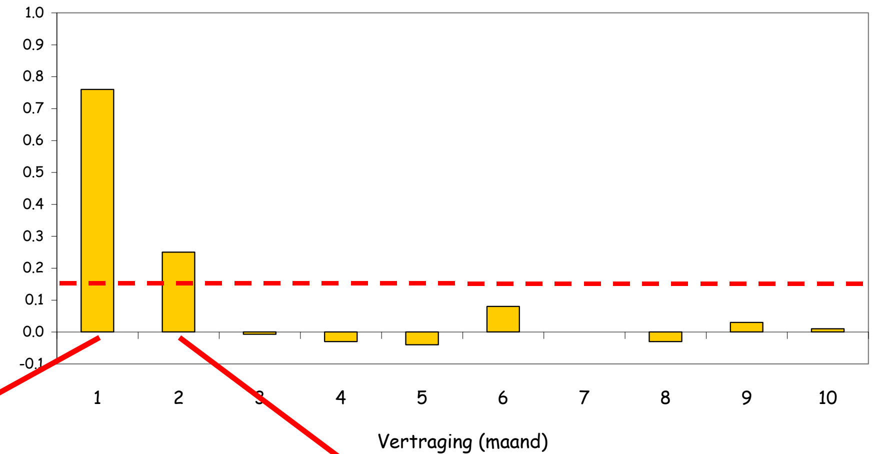
Keuze ruismodel: $\boxed{N_t = \phi_1 N_{t-1} + a_t}$

Identificatie ruismodel

autocorrelogram



partieel autocorrelogram



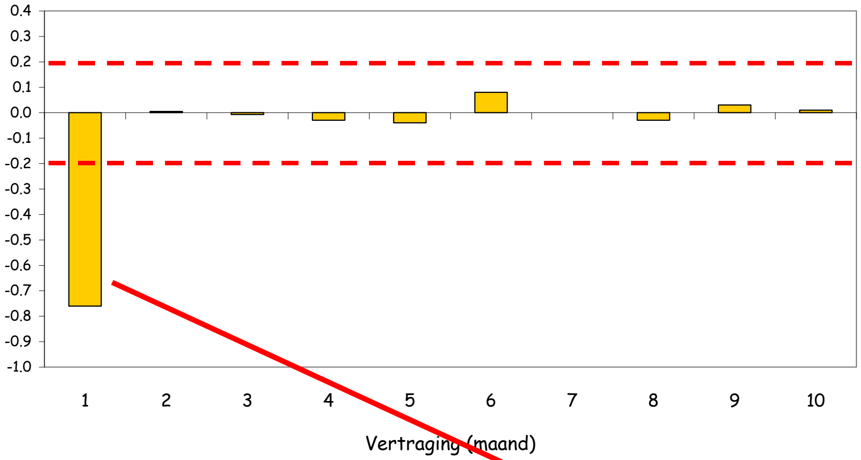
$$N_t = \phi_{11} N_{t-1} + a_t$$

$$N_t = \phi_{21} N_{t-1} + \phi_{22} N_{t-2} + a_t$$

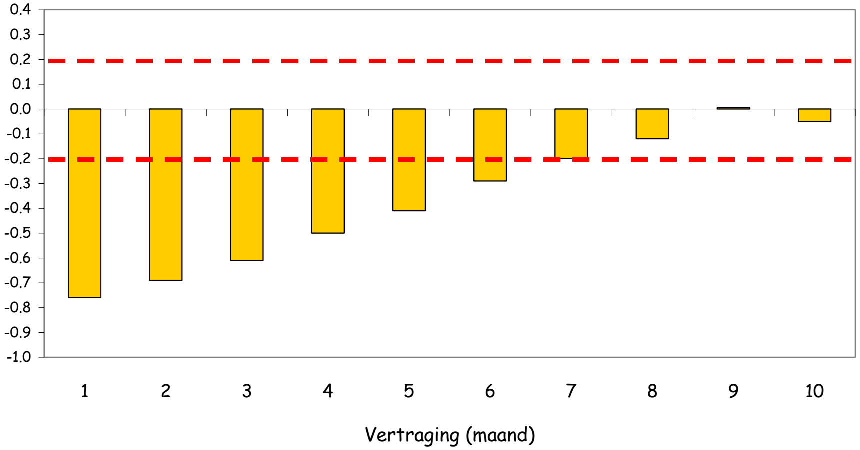
Keuze ruismodel: $N_t = \phi_1 N_{t-1} + \phi_2 N_{t-2} + a_t$

Identificatie ruismodel

autocorrelogram



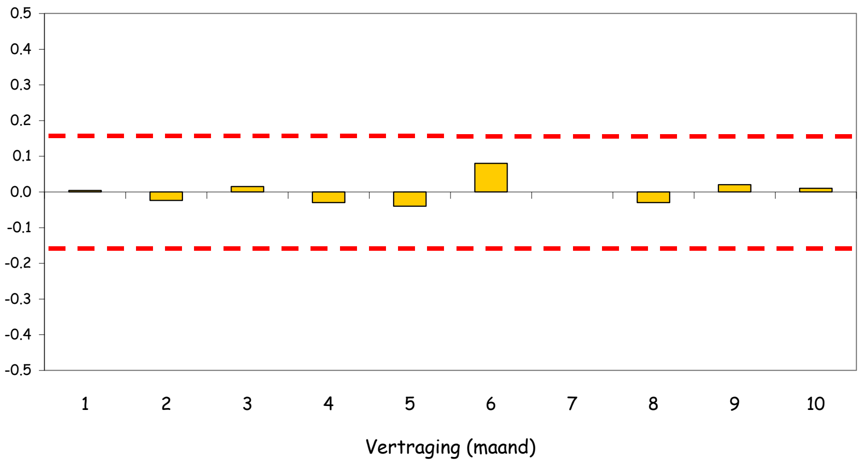
partieel autocorrelogram



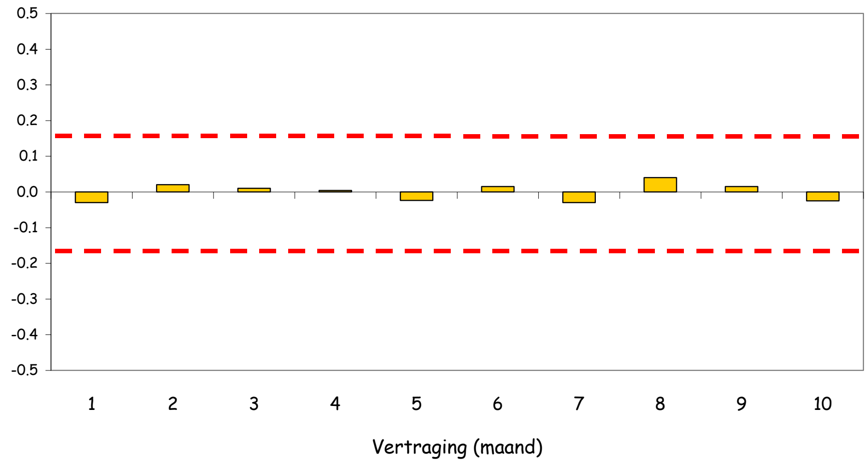
Keuze ruismodel: $\boxed{N_t = a_t - \theta_1 a_{t-1}}$

Verificatie ruismodel
Onafhankelijkheid modelresiduën?

autocorrelogram

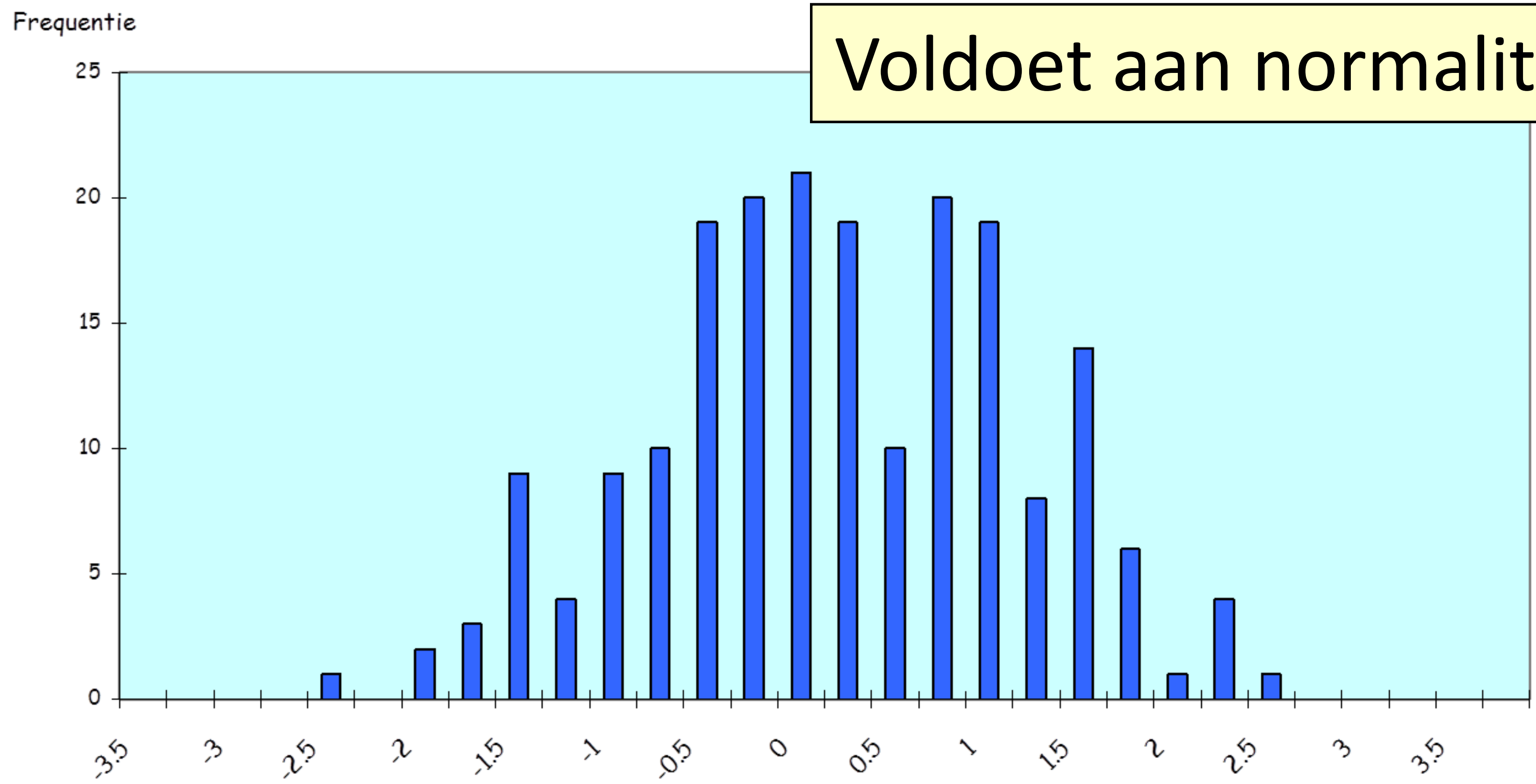


partieel autocorrelogram



Geen ‘structuur’ meer aanwezig

Verificatie ruismodel
Normaliteit modelresiduën?



Voldoet aan normaliteit

Bij niet-normaliteit => transformeren uitvoerreeks

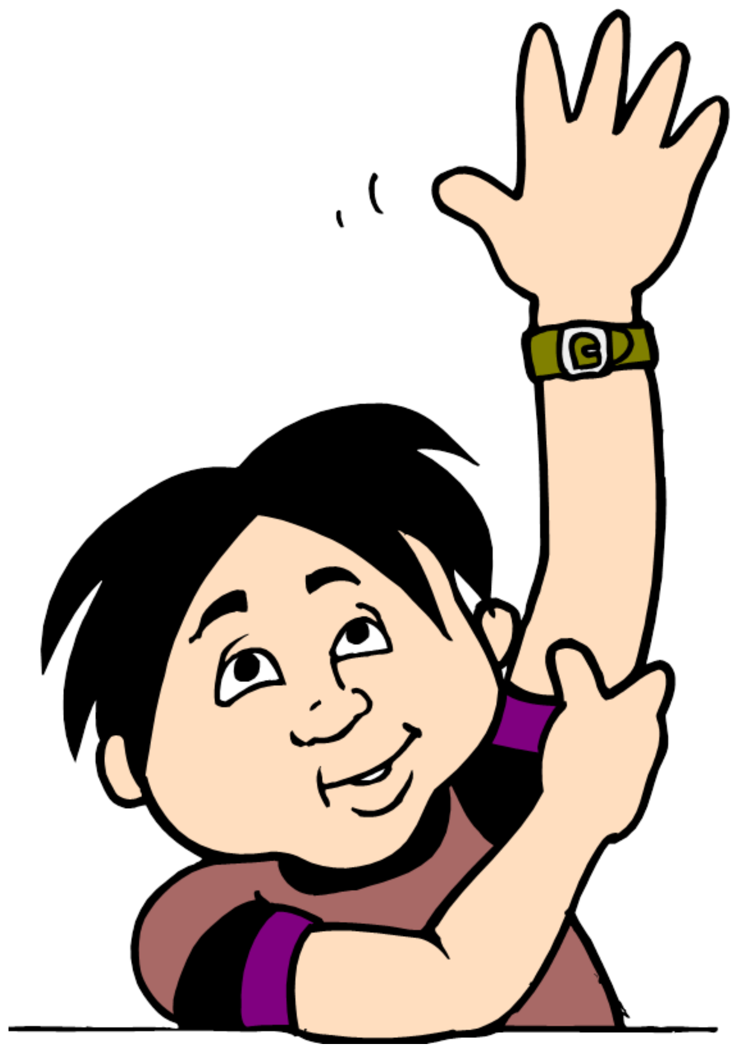
Opties bij geen Gaussiaanse witte ruis

1. Niet melden en onzekerheden kwantificeren alsof er wél sprake is van Gaussiaanse witte ruis
2. Wel melden, maar stellen dat het weinig uitmaakt voor de onzekerheden
3. Wel melden en stellen dat de onzekerheden van uitspraken daardoor niet meer objectief kwantificeerbaar zijn (ze zijn ‘zacht’)

Opties bij wit, maar niet Gaussiaans

1. Transformeer uitvoerreeks
2. Ga uit van theoretische kansverdeling waar modelresiduën best aan voldoen
3. Bootstrappen empirische kansverdeling van modelresiduën?
4. Bootstrappen empirische kansverdeling van modelresiduën, maar vastgesteld met modelvalidatie?
   - onafhankelijke validatieset lastig te vinden
   - kan realistischer beeld geven, omdat dan ook modelonzekerheid wordt verdisconteerd

Vragen?



Doelstellingen tijdreeksanalyse

1. Systeeminzicht

2. Voorspellen uitvoer

3. Systeemregeling