

beoordeling

January 16, 2021

1 Beoordeling Tijdreeksmodellen

Auteurs: O.N. Ebbens (Artesia), D. Brakenhoff (Artesia)

Een veelgestelde vraag is hoe je een geoptimaliseerd tijdreeksmodel moet beoordelen. Dit notebook laat zien welke methoden hiervoor gebruikt kunnen worden aan de hand van voorbeelden. De eenvoudigste methode voor de beoordeling van een tijdreeksmodel is een visuele beoordeling. Ook kunnen er diverse statistieken berekend worden voor een geoptimaliseerd tijdreeksmodel. Aan de hand van een stijghoogtereeks uit de praktijk worden deze methodes toegelicht. Voor de berekende statistieken uitgelegd wat de getallen betekenen en hoe deze gebruikt kunnen worden.

Een veelgebruikt criterium voor modelbeoordeling is de ‘fit’, die vaak wordt uitgedrukt met de afkorting EVP, wat staat voor het “Explained Variance Percentage”. Hiermee wordt uitgedrukt hoe goed de simulatie op de metingen lijkt. Het notebook laat zien in hoeverre deze ‘fit’ een goede indicatie is voor de voorspellende waarde van het model.

Het notebook behandelt de meest gebruikte statistische toetsen. Speciale aandacht wordt besteed aan het kwantificeren van de betrouwbaarheid van de modelparameters. Welke voorwaarden zijn daarvoor vereist om dit te kunnen doen? Hoe kan je het model aanpassen zodat je aan de vereiste statistische voorwaarden voldoet?

%% DB: Ik heb wat tekst aanpassingen gedaan. Introductie kan nog wat duidelijker als de inhoud van het notebook compleet is denk ik. %%

%% voorbeelden toevoegen van: - Model met hoge EVP met grote onzekerheid in parameter (plaatje hieronder maken in notebook?). Hiermee laten zien waarom 2x standaarddeviatie belangrijk is. - Modellen met hoge correlatie tussen 2 parameters (eentje aanpassen en vastzetten dan zien dat de fit goed blijft?). Makkelijkst met een recharge factor. - Model wat door alle diagnostische toetsen heenkomt.

Nog doen - waar dubbele procent-tekens staan (%%) moet nog iets gebeuren - Meer referenties opnemen, oa statistieken fit paper van Raoul - model nemen wat door alle toetsen heenkomt (met gelijke tijdstappen), mogelijk basic model van bpastas - Kijk naar woorden als vaak en regelmatig en vermijdt deze - Model met goede evp en slechte fit openmen - hoofdstuk toevoegen -> hydrologische beoordeling: - bij een onttrekking gws omlaag - verdampingsfactor - geheugen - ruimtelijke beoordeling (resultaten t.o.v. andere buizen in de omgeving), pastas artikel verwijzing - frequency of exceedance

%% Een tijdreeksmodel kan op een aantal onderdelen beoordeelt worden. In dit notebook komen de volgende onderwerpen aan bod:

1.0.1 Inhoudsopgave

- Maken tijdreeksmodel
- Visuele beoordeling
- Fit statistieken
- **Validatie** %% wordt onderdeel van fit statistiek. Kan je een model valideren, heet het validatie/verificatie/..?
- Diagnostische toetsen
- Correlatie modelparameters
- Hydrologische beoordeling
- **Geheugen van de responsefunctie** %% onderdeel van hydrologische beoordeling

1.1 Beoordeling

Voor het gebruiken of afkeuren van een tijdreeksmodel is het belangrijk om vast te stellen wat het doel van het tijdreeksmodel is. Op basis van dit doel kunnen de criteria worden vastgesteld voor de hierboven genoemde onderdelen.

Python version: 3.8.5 (default, Sep 3 2020, 21:29:08) [MSC v.1916 64 bit (AMD64)]

Numpy version: 1.19.2

Scipy version: 1.5.2

Pandas version: 1.2.0

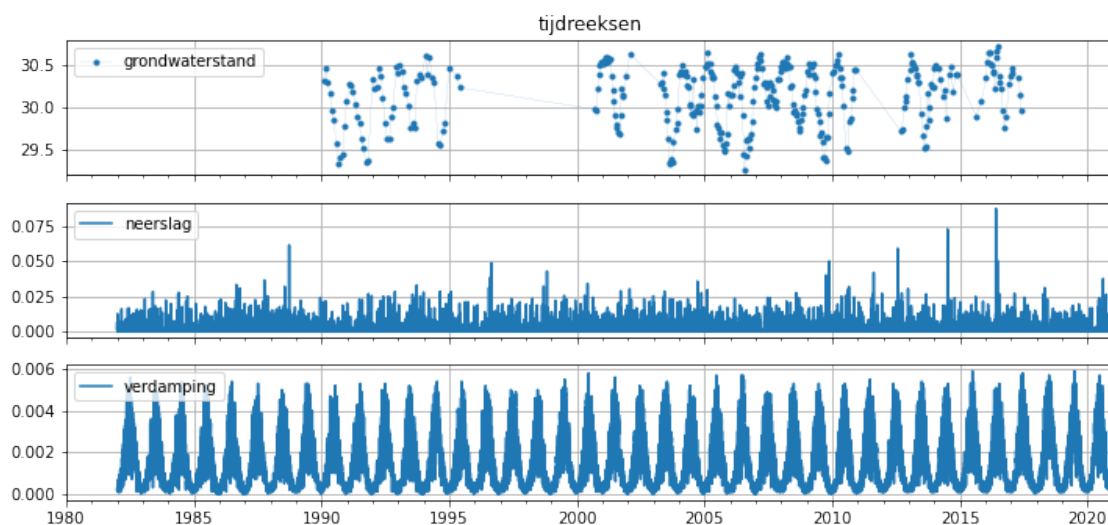
Pastas version: 0.17.0b

Matplotlib version: 3.3.2

numba version: 0.51.2

1.2 Maken tijdreeksmodel

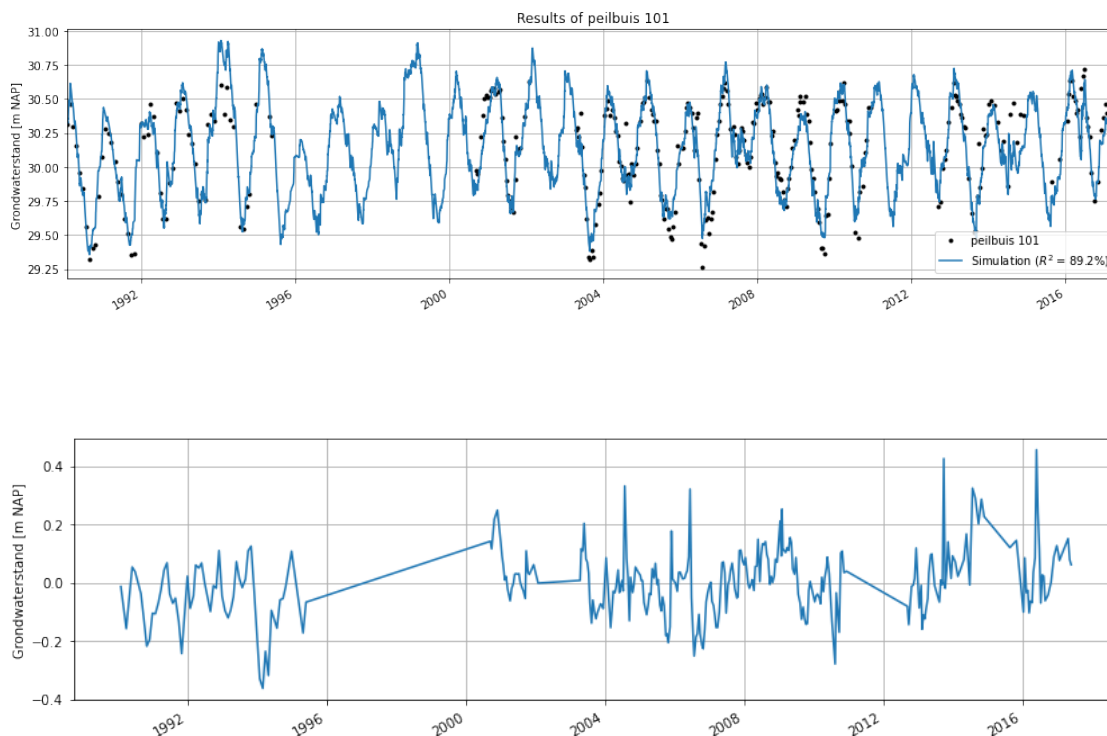
In dit notebook maken we gebruik van een simpel tijdreeksmodel om de grondwaterstand te simuleren met neerslag en verdamping als verklarende variabelen.



1.3 Visuele beoordeling

De eerste check om de betrouwbaarheid van het model in te schatten is vaak een visuele vergelijking van de modelsimulatie met de metingen. In onderstaande grafiek is deze vergelijking gemaakt. De modelresultaten en de metingen komen behoorlijk goed overeen. Ook is duidelijk dat de verschillen tussen het model en de metingen vooral ontstaan bij de extremen, de dalen in de zomer en de pieken in de winter.

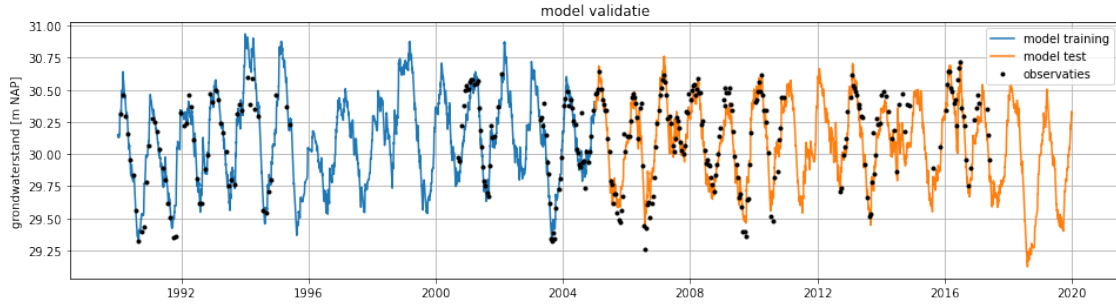
%%residuen erbij laten zien en een voorbeeld slechte fit (met hoge evp)%%



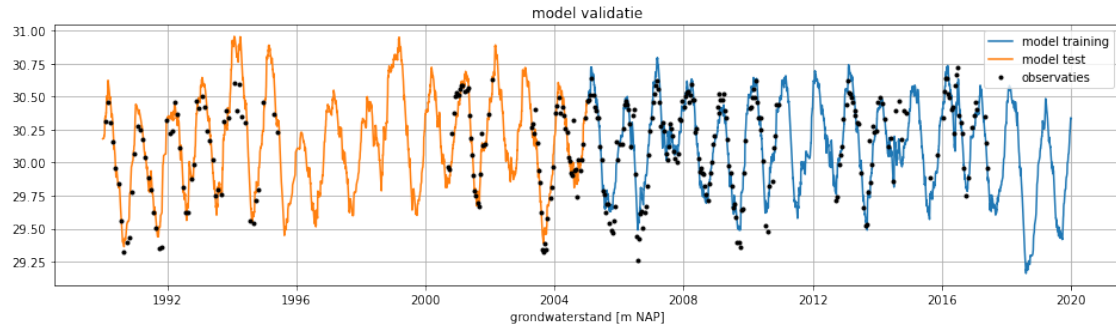
1.4 Validatie

Wanneer er voldoende metingen zijn kan ervoor worden gekozen om een validatie uit te voeren. Hierbij wordt de tijdreeks met gemeten grondwaterstanden gesplitst in een trainingsperiode en een testperiode. Het tijdreeksmodel wordt gefit op de metingen in de trainingsperiode en er wordt een modelsimulatie gemaakt voor de testperiode. Het vergelijken van de gesimuleerde grondwaterstanden met de metingen in de testperiode geeft een goede indicatie hoe goed het model presteert.

%%voorbeeld toevoegen model wat wel goed fit maar slechte validatie (model van raoul met trend)%%



De trainings- en testperiode kunnen zelf worden gekozen. De testperiode kan ook voor de trainingsperiode liggen, zie voorbeeld hieronder. Let er wel op dat er van de verklarende tijdreeksen, in dit geval neerslag en verdamping, metingen beschikbaar moeten zijn voor zowel de test- als de trainingsperiode.



1.5 Fit statistieken

Naast de visuele vergelijking van gemodelleerde en gemeten grondwaterstanden bestaan er ook verschillende statistieken die iets zeggen over de fit tussen gemodelleerde en gemeten grondwaterstanden. %%Referenties naar fit statistieken opnemen%% Een aantal veelgebruikte fit statistieken zijn:

- **som van de kwadratische fout / sum of the squares of the error** $SSE = \sum_{i=1}^N (x_i - y_i)^2$
- **gemiddelde absolute fout / mean absolute error** $MAE = \sum_{i=1}^N \frac{|x_i - y_i|}{N}$
- **wortel van de kwadratisch gemiddelde fout / root mean squared error** $RMSE = \sqrt{\sum_{i=1}^N \frac{(x_i - y_i)^2}{N}}$
- **verklaarde variantie / explained variance percentage** $EVP = \frac{\sigma_h^2 - \sigma_r^2}{\sigma_h^2} * 100$
- **determinatiecoëfficiënt / r-squared** $R^2 = 1 - \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (x_i - \bar{x})^2}$

N is het aantal metingen. N_{param} is het aantal vrije parameters. x_i en y_i zijn respectievelijk de gemeten en de gemodelleerde grondwaterstand op tijdstip i . σ_h^2 en σ_r^2 zijn respectievelijk de variantie in de gemeten grondwaterstanden en de residuën.

%% openstaande vragen: - waarom zit de correctie voor aantal parameters bij de berekening van r-squared? -> issue pastas - verschil verklaarde variantie/r-squared/NSE, hoe op te nemen in deze uitleg? %%

```
Mean absolute error (mae)           : 0.085
Root mean squared error (rmse)      : 0.114
Sum of the squares of the error (sse) : 4.174
Explained variance percentage (evp)  : 89.2%
R-squared (rsq)                     : 0.892
```

1.5.1 model validatie

Wanneer een model validatie is uitgevoerd kunnen de gesimuleerde en gemeten grondwaterstanden ook worden vergeleken met behulp van dezelfde statistieken. In principe wordt de R-squared enkel gebruikt om de “goodness of fit” te berekenen bij een model kalibratie. Wanneer wordt berekend hoe goed de voorspellingen van het model zijn wordt gebruik gemaakt van de Nash-Sutcliffe model efficiency coefficient. Deze is equivalent aan de R-squared values alleen worden nu niet het aantal vrije modelparameters meegenomen.

%% Misschien wat onduidelijk het verschil tussen r-squared en NSE, komt hiervandaan <https://stats.stackexchange.com/questions/185898/difference-between-nash-sutcliffe-efficiency-and-coefficient-of-determination> %%

Nash-Sutcliffe model efficiency $NSE = 1 - \frac{\sum_{i=1}^N (x_i - y_{i,sim})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}$ $y_{i,sim}$ is de met het model voorspelde grondwaterstand.

fit statistieken validatie periode 1990 - 2005

```
-----
Mean absolute error (mae)           : 0.088
Root mean squared error (rmse)      : 0.119
Sum of the squares of the error (sse) : 1.719
Explained variance percentage (evp)  : 89.7%
Nash-Sutcliffe model efficiency coefficient (nse) : 0.887
```

fit statistieken validatie periode 2005 - 2020

```
-----
Mean absolute error (mae)           : 0.099
Root mean squared error (rmse)      : 0.126
Sum of the squares of the error (sse) : 3.173
Explained variance percentage (evp)  : 87.5%
Nash-Sutcliffe model efficiency coefficient (nse) : 0.862
```

1.5.2 overzicht fit statistieken

parameter	waarde bij optimale fit	minimum	maximum
sse	0	0	∞
mae	0	0	∞
rmse	0	0	∞

parameter	waarde bij optimale fit	minimum	maximum
evp	100%	0	100
R^2	1	0	1
nse	1	$-\infty$	1

1.5.3 Overfitting

Wanneer er meer verklarende variabelen, en bijbehorende modelparameters, aan een tijdreeksmodel worden toegevoegd is de fit over het algemeen beter. Dit is zelfs het geval als de grondwaterstanden niet worden beïnvloed door de verklarende variabele. Immers, hoe meer vrijheidsgraden (parameters) het model heeft des te makkelijker het model op de metingen kan worden gefit. Maar als zo'n verklarende variabele (die er eigenlijk niet toe doet) wordt toegevoegd verbetert de voorspellende waarde van het model niet. Het fenomeen waarbij meer parameters worden toegevoegd aan het model dan gerechtvaardigd op basis van de metingen heet overfitting.

Het is dus niet mogelijk om op basis van enkel de fit statistieken aan te geven of het model verbetert na het toevoegen van een verklarende variabele. Twee statistieken die wel rekening houden met overfitting zijn:

- **aic**: Akaike information criterion
- **bic**: Bayesian information criterion

Zowel de **aic** als de **bic** zijn een maat voor de relatieve hoeveelheid informatie die verloren is gegaan in het model. Het zijn dus relatieve waarde en kunnen het best gebruikt worden om verschillende modellen voor dezelfde meetreeks met elkaar te vergelijken. Wanneer de **aic** en **bic** lager worden na het toevoegen van een verklarende variabele betekent dit dat het model verbetert (er gaat relatief minder informatie verloren). Als de **aic** en **bic** hoger worden dan heeft het toevoegen van de parameters niet bijgedragen aan een verbetering van het model.

In het voorbeeld hieronder is het eerste model vereenvoudigd door de neerslag en de verdamping samen te nemen als een verklarende reeks.

Eerst zijn de **evp**, **aic** en **bic** van het eerste voorbeeld model gepresenteerd. In dat model zijn de neerslag en de verdamping als twee losse verklarende variabelen opgegeven waarbij voor beide invloeden 3 parameters worden toegevoegd aan het model.

%% aic en bic moeten op de noise worden uitgerekend mits de noise ongecorrleerd is %%

```

Explained variance percentage (evp) : 89.25%
R-squared (rsq)                     : 0.89
Akaike information criterion (aic)   : 13.14
Bayesian information criterion (bic) : 43.36

```

Het is kan worden vereenvoudigd door de grondwateraanvulling (het verschil tussen neerslag en verdamping) als verklarende reeks toe te passen. Als de statistieken opnieuw berekend worden voor dit model is duidelijk dat de fit iets minder goed is geworden (van 89% naar 87%) maar de **aic** en **bic** zijn lager. Als er een model geselecteerd moet worden, dan gaat de voorkeur uit naar het model met de laagste **aic**. Met andere woorden, de extra parameters in het eerste model leveren dus niet voldoende verbetering op in de fit om toevoeging van de parameters te ondersteunen. De **aic** valt daarom hoger uit ten opzichte van het vereenvoudigde model.

%% We hebben ook voorbeelden waarbij een aanzienlijke verbetering in fit na toevoegen van een invloed, de AIC alsnog hoger uitvalt..., dus wat moeten we daarmee? Zie bijvoorbeeld het model met en zonder rivier in “adding_rivers.ipynb”.%%

```
Explained variance percentage : 87.18%
R-squared (rsq)              : 0.87
Akaike information criterion  : 6.79
Bayesian information criterion : 25.68
```

	Value
Statistic	
rmse	0.113815
rmsn	0.090891
sse	4.171305
mae	0.084914
nse	0.892459
evp	89.249388
rsq	0.892551
bic	43.364761
aic	13.143542

1.6 Betrouwbaarheid modelparameters

Naast de beoordeling van de model fit is het ook van belang om te kijken naar de modelparameters. Externe invloeden worden gemodelleerd met responsfuncties die ieder één of meerdere parameters bevatten. Daarnaast kunnen in tijdreeksmodellen ook nog een constante en een noisemodel worden toegevoegd. In het voorbeeld hieronder zijn de parameters weergegeven van een tijdreeksmodel met recharge als externe invloed.

	initial	optimal	pmin	pmax	stderr
gw_aanvulling_A	210.475594	480.819216	0.00001	21047.559369	36.513677
gw_aanvulling_n	1.000000	1.098409	0.10000	100.000000	0.038177
gw_aanvulling_a	10.000000	102.749876	0.01000	10000.000000	11.843657
constant_d	30.124737	29.888102	NaN	NaN	0.022960
noise_alpha	17.000000	45.193179	0.00001	5000	6.885089

Per parameters zijn de volgende waarden weergegeven: - **initial**: de initiële waarde - **optimal**: de geoptimaliseerde waarde - **pmin** en **pmax**: de bandbreedte waarbinnen de parameters worden geoptimaliseerd - **stderr**: de standaarddeviatie van de geoptimaliseerde parameter

De berekende standaarddeviaties zijn alleen betrouwbaar als aan een aantal voorwaarden wordt voldaan. Deze voorwaarden zijn hieronder uitgewerkt. Daaronder wordt behandeld hoe je de standaarddeviatie kan gebruiken om meer inzicht te krijgen in de modelresultaten.

vuistregel:

Als de standaarddeviaties van de parameters worden gebruikt, is een diagnostische check van het ruismodel nodig.

1.6.1 Model diagnostiek

De berekende standaarddeviaties zijn betrouwbaar als aan een aantal voorwaarden wordt voldaan. De twee voorwaarden voor de met het ruismodel berekende ruis zijn dat het: 1. zich gedraagt als witte ruis met een gemiddelde van nul. 2. niet gecorreleerd is aan een verklarende tijdreeks.

Er is sprake van witte ruis wanneer: 1. er geen significante autocorrelatie in de ruis aanwezig is 2. de ruis homoscedastisch is 3. de ruis voldoet aan een, bij voorkeur normale, statistische verdeling.

Het is aan te bevelen om te checken of aan deze voorwaarden wordt voldaan wanneer gebruik wordt gemaakt van de standaarddeviaties van de parameters. De voorwaarden voor de ruis kunnen gecheckt worden via een aantal statistische toetsen. Deze test worden meestal uitgevoerd met een nulhypothese (H_0) en een alternatieve hypothese (H_A):

- H_0 : de ruis is onafhankelijk, homoscedastisch of normaal verdeeld
- H_A : de ruis is niet onafhankelijk, homoscedastisch, of normaal verdeeld

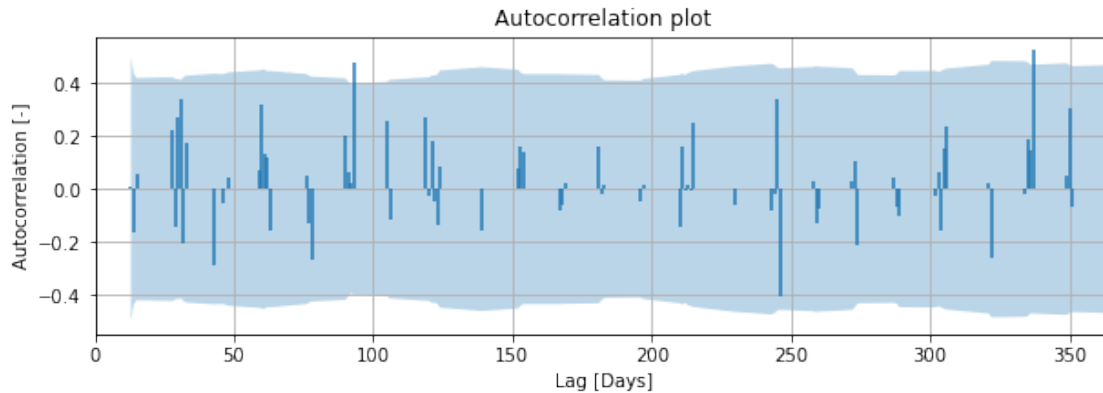
Voor een hypothese toets wordt een bepaalde statistische waarde berekend (bijv. Q_{test}), die vervolgens wordt vergeleken met een theoretische waarde op basis van een bepaalde verdeling (bijv. $\chi^2_{\alpha,h}$) die afhankelijk is van een significantieniveau (bijv. $\alpha = 0.05$) en soms afhankelijk is van het aantal vrijheidsgraden h . Het resultaat is het wel of niet verwerpen van de nulhypothese. Dit resultaat kan nooit gebruikt worden om de alternatieve hypothese te accepteren. Wanneer bij de test voor autocorrelatie de nulhypothese niet verworpen wordt kunnen we concluderen dat er geen significante autocorrelatie aanwezig is in de ruis maar we kunnen niet concluderen dat er geen autocorrelatie is.

In deze paragraaf worden de 3 voorwaarden voor witte ruis en de daarbij horende statistische toetsen verder toegelicht.

%% Nog wat vragen: - waarom twee testen voor normale verdeling en autocorelatie? Wat als de ene wel door de test komt en de ander niet? - Klopt het dat dit alle voorwaarden zijn voor witte ruis? - klopt het wel wat ik zeg over het accepteren van de nulhypothese maar het niet verwerpen van de alternatieve hypothese? %%

Autocorrelatie Het kan gebeuren dat een tijdreeks gecorreleerd is met een in de tijd verschoven versie van zichzelf. Als dat het geval is spreken we van autocorrelatie. De autocorrelatie kan worden weergegeven in een autocorrelatiediagram zoals hieronder. Hierin is op de x-as de verschuiving in de tijd weergegeven (lag) en op de y-as de hierbij berekende correlatie. De autocorrelatie wordt weergegeven tot een maximale verschuiving van 1 jaar (365 dagen). De blauwe band in de grafiek geeft het 95% betrouwbaarheidsinterval weer. Wanneer 95% van de berekende autocorrelaties binnen dit interval liggen kunnen we concluderen dat er geen significante autocorrelatie aanwezig is in de reeks.

%% Vragen: - Als het betrouwbaarheidsinterval wordt berekend op basis van de autocorrelaties en je vervolgens gaat toetsen of de autocorrelaties binnen dit betrouwbaarheidsinterval liggen dan komt deze altijd door de toets toch? Of maak ik nu een denkfout? Waar wordt de blauwe band mee berekend? - dit model komt niet door de Runs test, enig idee waarom niet? %%



Naas de visuele check op autocorrelatie bestaan er ook statische toetsen om te checken of er autocorrelatie aanwezig is. Deze zijn in onderstaande tabel weergegeven. De eerste twee methodes **Ljung_Box** en **Durbin-Watson** zijn alleen geschikt voor equidistante tijdreeksen. In tijdreeksen van grondwaterstanden missen vaak metingen en dus zijn deze niet equidistant. Voor het checken van autocorrelatie worden daarom meestal de **Stoffer-Toloi** en de **Runs Test** gebruikt die wel kunnen omgaan met niet equidistante reeksen.

Naam	Beschrijving	geschikt voor niet equidistante reeksen
Ljung-Box test	Deze test checkt of de autocorrelatie in een tijdreeks significant verschilt van nul	Nee
Durbin-Watson test	Deze test checkt de autocorrelatie bij een verschuiving van één tijdstap.	Nee
Stoffer-Toloi test	Deze test is vergelijkbaar met de Ljung-Box test, maar aangepast voor missende waarde	Ja
Runs test	Deze test checkt of de waarden in een tijdreeks willekeurig zijn zonder uit te gaan van een bepaalde kansverdeling.	Ja

%% Hoe zit het met Stoffer-Toloi? Hoeveel data mag er missen? Kan je ook bij een frequentie waar soms iets mist toch deze test doen?%%

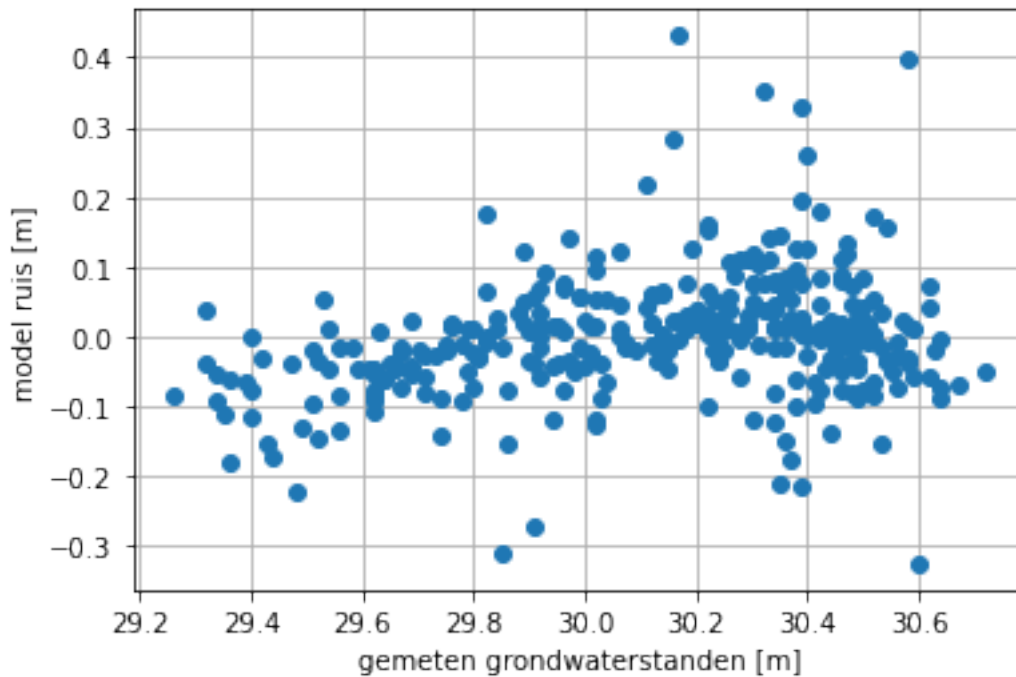
Stoffer-Toloi test:
Nullhypothese verworpen

Runs test:
Nullhypothese verworpen

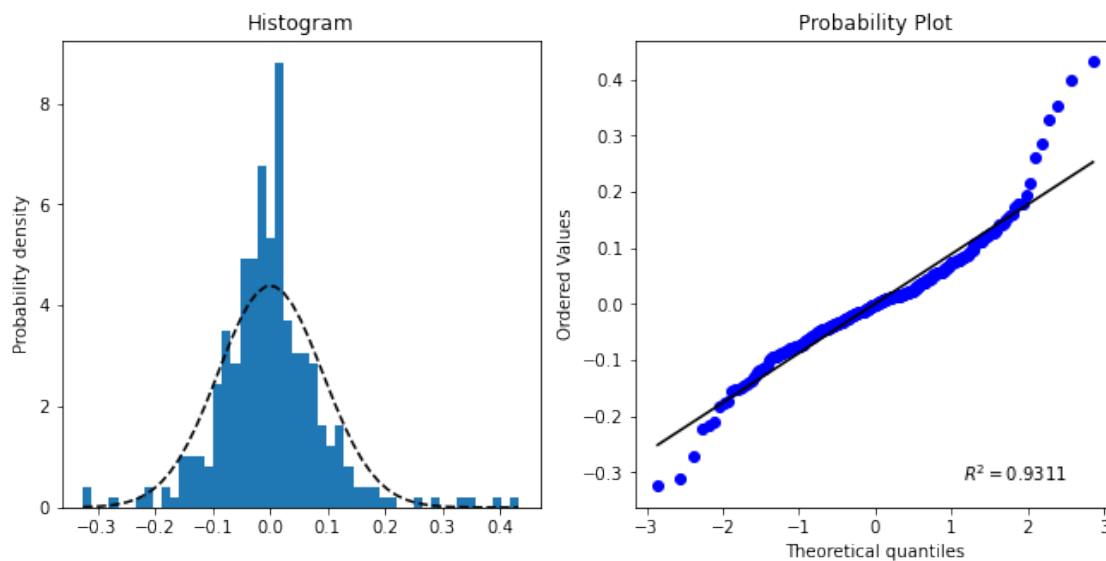
Homoscedasticiteit Een reeks is homoscedastisch als alle waarden dezelfde eindige variantie hebben. Bij ons tijdreeksmodel betekent dit dat de ruis onafhankelijk is van de gemeten grondwa-

terstanden. Dit is hieronder visueel weergegeven door deze waarden tegen elkaar uit te zetten in een grafiek.

Wanneer sprake is van heteroscedasticiteit ...



Normale verdeling Een veelgebruikte aanname is dat de ruïs normaal verdeeld is. In principe is het ook mogelijk dat er een andere kansverdeling bij de ruïs hoort. De verdeling van de ruïs kan weergegeven worden met een histogram in combinatie met een kansverdeling.



Het is een beruchte uitdaging om te bepalen of een tijdreeks normaal verdeeld is. Er bestaan wel testen maar deze zijn niet altijd betrouwbaar. Aan de ene kant geldt dat als er weinig metingen beschikbaar zijn, de afwijkingen van een normale verdeling erg sterk moeten zijn voordat de test ze kan detecteren. Aan de andere kant geldt dat bij een groot aantal metingen, hele kleine (irrelevante) afwijkingen van de normale verdeling al tot significante uitkomsten leiden. Het is daarom lastig objectief vast te stellen of ruis normaal verdeeld is.

Naam	Beschrijving	Geschikt voor niet equidistante tijdreeksen?
Shapiro-Wilk	Deze test checkt of de data uit een normale verdeling is getrokken	Onbekend
D'Agostino	Deze test checkt of de ruis voldoet aan een normale verdeling (H0 hypothesis).	Onbekend

%% vragen: - werken deze testen voor niet equidistante tijdreeksen? - waarom zijn de p-waarden zo klein bij kleine tijdstappen? Tijdstap van 70 dagen lukt het wel, gevoelsmatig wel weer heel groot. %%

Shapiro-Wilk test:

Reject the Null-hypothesis. p = 0.0

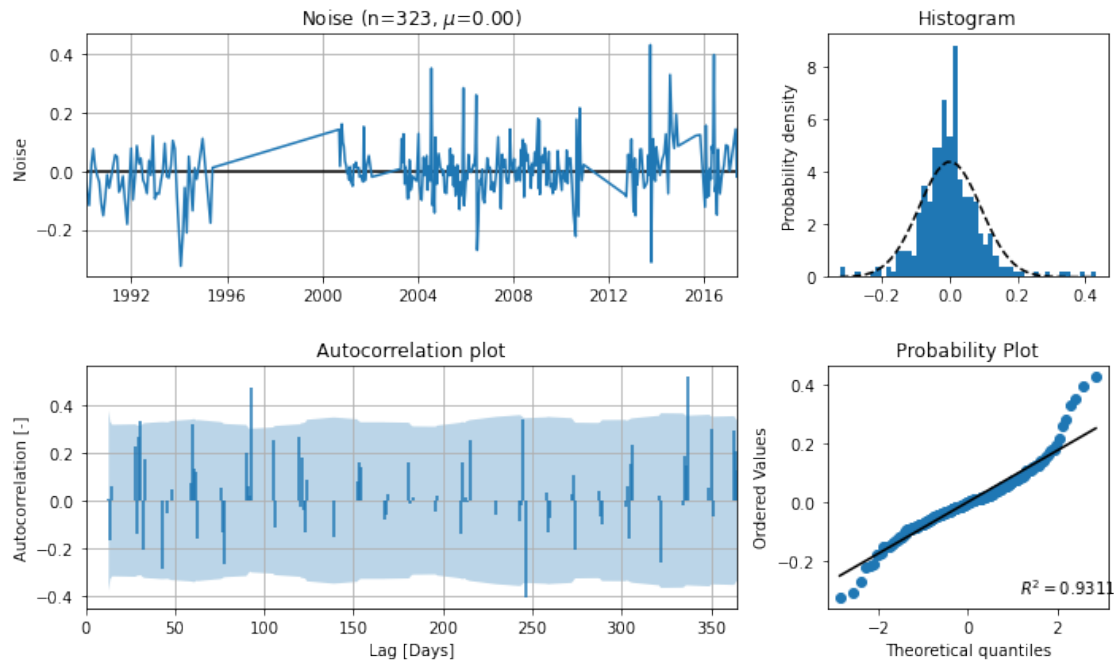
D'Agostino test:

Reject the Null-hypothesis. p = 0.0

Overzicht visuele en diagnostische toetsen

	Checks	Statistic	P-value	Reject H0
Shapirooo	Normality	0.95	0.00	True
D'Agostino	Normality	34.11	0.00	True
Runs test	Autocorr.	-2.62	0.01	True
Stoffer-Toloi	Autocorr.	11.55	0.64	False

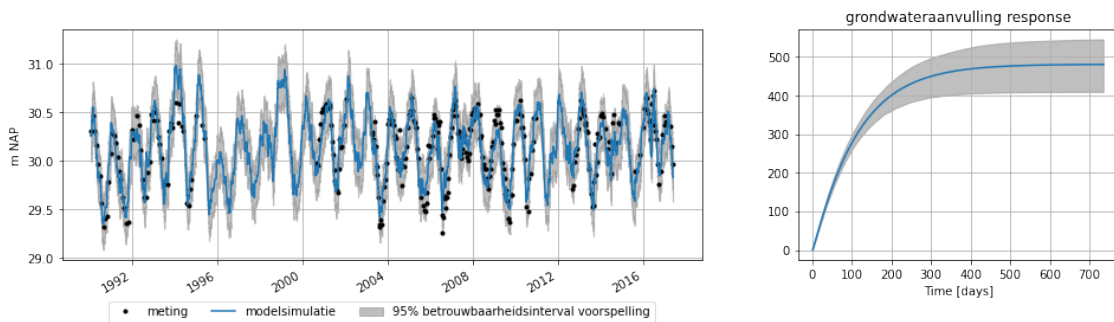
```
c:\users\oebbe\02_python\pastas\pastas\stats\tests.py:616: UserWarning:
FixedFormatter should only be used together with FixedLocator
  ax.set_xticklabels(ax.get_xticklabels(), rotation=0, ha="center")
```



1.6.2 Standaarddeviatie parameters

Wanneer wordt voldaan aan bovengenoemde voorwaarden kan de standaarddeviatie van de modelparameters worden gebruikt om de betrouwbaarheid van het model in te schatten. Zoals bijvoorbeeld hieronder bij het plotten van de bandbreedte van een modelsimulatie en de bandbreedte van een responsfunctie.

metingen binnen betrouwbaarheidsinterval: 95.4%



De standaarddeviatie kan ook worden gebruikt om de betrouwbaarheid van een model te toetsen. Bijvoorbeeld door te kijken naar de grootte van de standaarddeviatie t.o.v. de geoptimaliseerde parameterwaarde. Hieronder testen we of de absolute, geoptimaliseerde parameterwaarde groter

is dan 2x de standaarddeviatie. In dit geval komen alle parameters behalve de `step_d` parameter door onze test. Dit betekent dat de invloed van de step-trend een grote onzekerheid kent.

%%Is deze check wel interessant? En we mogen deze ook helemaal niet doen volgende onze eigen analyse want de 3 voorwaarden voor witte ruis zijn niet getest%%

```
gw_aanvulling_A    True
gw_aanvulling_n    True
gw_aanvulling_a    True
constant_d         True
noise_alpha        True
dtype: bool
```

1.7 Parameter correlaties

Ook is het van belang om te kijken naar de parameter correlaties. Wanneer twee modelparameters sterk gecorreleerd zijn betekent dit dat bijna dezelfde modelresultaten kunnen worden verkregen met andere waarden voor deze twee parameters.

Hieronder zijn de parameter correlaties gepresenteerd voor het vereenvoudigde model met grondwateraanvulling als verklarende reeks.

```
<pandas.io.formats.style.Styler at 0x215c0b8e580>
```

In het geval van het model met grondwateraanvulling als invloed wordt de grondwaterstand berekend als de som van de bijdrage van de recharge en de constante. Een lagere constante kan relatief eenvoudig gecorrigeerd worden door een grotere bijdrage van de recharge. In dit geval is de grote correlatie dan ook niet erg problematisch voor de bruikbaarheid van het model. Grote parameter correlaties kunnen wel problematisch zijn als je de bijdragen van verschillen verklarende variabelen los van elkaar wilt bekijken terwijl de parameters van de verklarende variabelen sterk gecorreleerd zijn. Een voorbeeld hiervan wordt hieronder uitgewerkt.

%% dit voorbeeld nog uitwerken met onttrekkingsreeksen die gecorreleerd zijn %%

voorbeeld problematische parameter correlaties Je maakt een grondwatermodel met als verklarende reeksen recharge en oppervlaktewaterpeilen. Het model presteert goed maar de parameters van beide verklarende variabelen zijn sterk gecorreleerd. Het model is daardoor niet geschikt om de invloed van recharge en oppervlaktewater van elkaar te scheiden. Dus kan het model niet gebruikt worden om het effect van een stijging in het oppervlaktewaterpeil te berekenen. Het is immers niet duidelijk of de stijging van de grondwaterstand wordt veroorzaakt door toegenomen recharge of door een stijging van het oppervlaktewaterpeil.

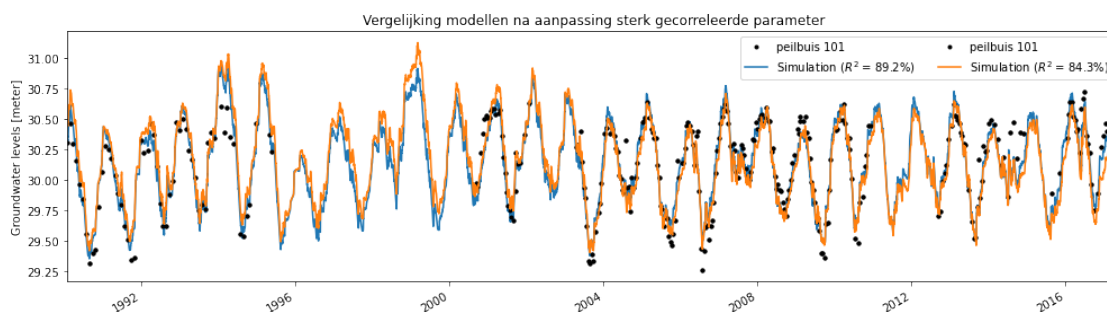
Het model kan nog wel bruikbaar zijn voor andere doeleinden. Als er bijvoorbeeld een grondwateronttrekking is gestart in de omgeving kan het effect daarvan mogelijk wel met het model bepaald worden. Voor het bepalen van dat effect is het immers niet van belang om de invloeden van recharge en oppervlaktewater van elkaar te scheiden.

voorbeeld parameter correlaties Hieronder is de correlatiematrix weergegeven van de parameters in het model met neerslag en verdamping als losse verklarende variabelen waarbij alleen correlaties hoger dan 0.5 of kleiner dan -0.5 zijn weergegeven.

<pandas.io.formats.style.Styler at 0x215bf21b1c0>

Parameters met een sterke positieve of negatieve correlatie gevan aan dat een aanpassing in een parameter gecompenseerd kan worden met aanpassingen aan de gecorreleerde parameters terwijl de fit van het model vergelijkbaar blijft. Er zijn dus meerdere parameter combinaties die een model met een goede fit opleveren.

De parameter `verdamping_A` is bijvoorbeeld sterk negatief gecorreleerd met `verdamping_a`. In het voorbeeld hieronder wordt parameter `verdamping_A` aangepast, door het met een factor 5 te vermenigvuldigen. Als het model opnieuw geoptimaliseerd wordt terwijl de waarde voor `verdamping_A` vast staat, levert dat een model met een vergelijkbare fit op. De aanpassing van deze `verdamping_A` is gecompenseerd via de andere sterk gecorreleerde parameters.



De optimale parameters zijn hieronder weergegeven voor het oorspronkelijke model en het model waarbij `verdamping_A` is aangepast en vastgezet. Het is duidelijk te zien dat de aanpassing in `verdamping_A` is gecompenseerd door de overige parameters aan te passen. De grootste veranderingen zijn in de parameters die sterk gecorreleerd waren met `verdamping_A`; de waarde van `verdamping_a` is een factor 10 groter en `verdamping_n` is ongeveer gehalveerd.

	oorspronkelijk	aanpassing param	factor verandering
<code>neerslag_A</code>	403.979858	380.356276	0.941523
<code>neerslag_n</code>	0.991377	1.066711	1.075989
<code>neerslag_a</code>	115.178247	95.456967	0.828776
<code>verdamping_A</code>	-451.100840	-2255.504199	5.000000
<code>verdamping_n</code>	1.347121	0.872727	0.647846
<code>verdamping_a</code>	69.540394	729.949842	10.496775
<code>constant_d</code>	30.003412	32.941363	1.097921
<code>noise_alpha</code>	39.705090	70.118762	1.765989

1.8 Geheugen

%%RESPONSEFUNCTIE ERUIT OM HET MINDER PASTAS SPECIFIEK TE MAKEN%%

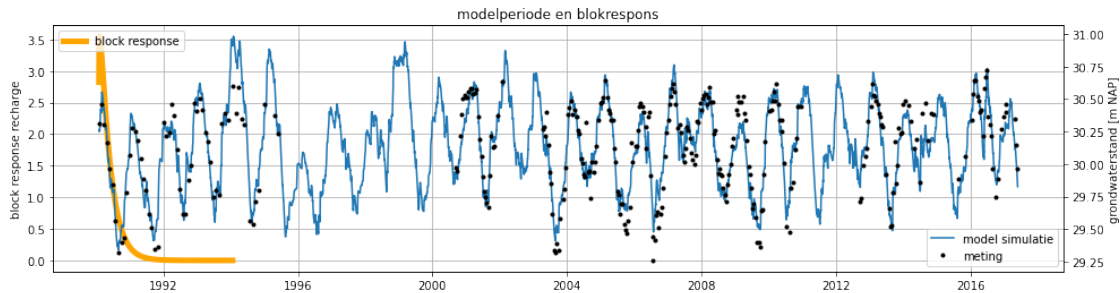
DB: leuk om hier ook een voorbeeldje te zien met (te) lange respons?

Hieronder zijn van twee modellen de block response en de modelresultaten in één figuur geplot. Bij

het eerste model is de lengte van het tijdreeksmodel een aantal maal groter dan het geheugen van de responsefunctie. Dit model voldoet daarmee aan de eis. Bij het tweede model is het geheugen van de responsefunctie langer dan de duur van het tijdreeksmodel, dit model voldoet niet aan de eis.

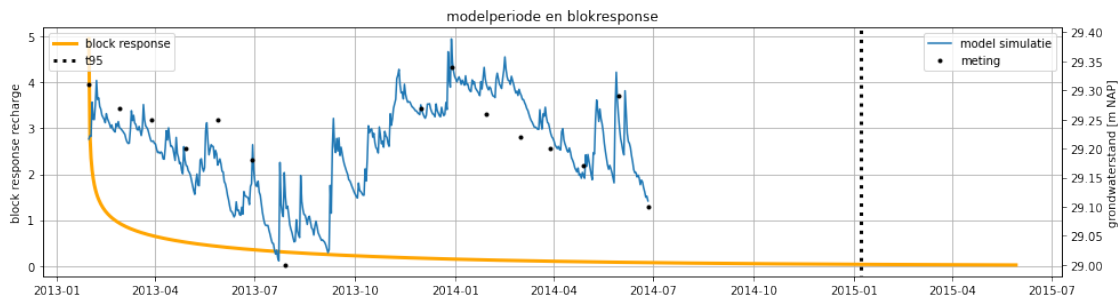
voorbeeld geheugen 1

<matplotlib.legend.Legend at 0x215c232cd60>



voorbeeld geheugen 2

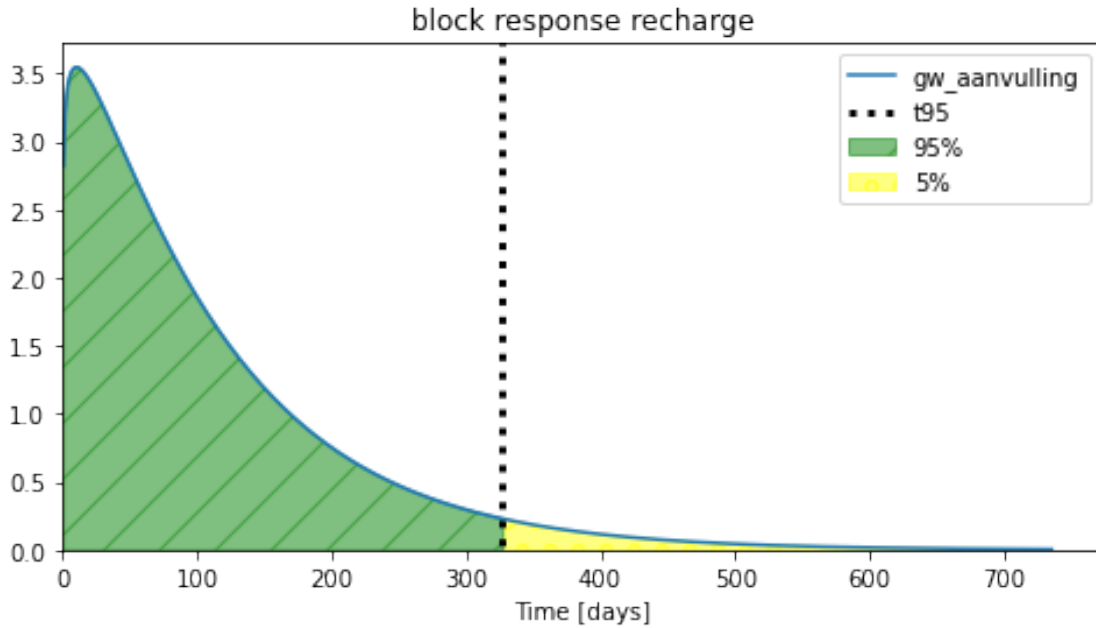
<matplotlib.legend.Legend at 0x215c23439d0>



1.8.1 bepalen geheugen van de responsfunctie

Het bepalen van het geheugen van de responsfunctie is niet altijd even vanzelfsprekend. Regelmatig worden exponentiële responsefuncties gebruikt met een horizontale asymptoot op $y=0$, de respons zal dus nooit gelijk zijn aan 0. Het is daarom niet mogelijk om het geheugen te bepalen door te kijken wanneer de respons 0 is. Om deze reden wordt als maat voor het geheugen vaak de tijd genomen die het duurt voordat 95% van de respons is uitgewerkt. Hieronder is dit visueel weergegeven.

geheugen (t_{95}) -> 327 dagen



1.9 Fit report pastas

%% DB: Pastas specifiek ...?

Een groot deel van de hierboven besproken tools voor het beoordelen van een model worden in pastas automatisch getoond na het optimaliseren van een tijdreeksmodel met `ml.solve()`. Dit zogenaamde fit rapport geeft een overzicht van de resultaten van de optimalisatie. Het rapport bestaat uit 4 onderdelen: - Algemene modelinstellingen (linksboven) - Fit statistieken (rechtsboven) - Parameters (midden) - Parameter correlaties (onder)

Hieronder is een fit rapport weergegeven, daaronder is per onderdeel aangegeven welke resultaten worden getoond.

```
Fit report peilbuis 101                                Fit Statistics
=====
nfev      41                      EVP                      89.24
nobs     323                      R2                       0.89
noise    True                     RMSE                     0.11
tmin     1990-01-30 00:00:00      AIC                      13.14
tmax     2017-05-30 00:00:00      BIC                      43.36
freq      D                       Obj                      1.34
warmup   3650 days 00:00:00      ---
solver   LeastSquares             Interpolated           No

Parameters (8 were optimized)
=====
              optimal  stderr      initial  vary
```


neerslag_A	402.443505	±11.31%	230.604464	True
neerslag_n	0.992814	±4.86%	1.000000	True
neerslag_a	114.123625	±18.89%	10.000000	True
verdamping_A	-449.965596	±15.74%	-2255.504199	True
verdamping_n	1.352186	±18.60%	1.000000	True
verdamping_a	69.086711	±33.59%	10.000000	True
constant_d	30.004957	±0.51%	30.124737	True
noise_alpha	39.761295	±15.22%	17.000000	True

Parameter correlations $|\rho| > 0.5$

```
=====
neerslag_A  neerslag_a    0.75
              constant_d  -0.67
neerslag_n  neerslag_a   -0.74
neerslag_a  constant_d   -0.52
verdamping_A verdamping_n  0.92
              verdamping_a -0.96
              constant_d   -0.78
verdamping_n verdamping_a -0.98
              constant_d   -0.71
verdamping_a constant_d    0.74
```

Algemene modelinstellingen Hier worden een aantal instellingen en resultaten genoemd van het model. Dit zijn: - **nfev**: Het aantal model simulaties dat is uitgevoerd door het optimalisatie-algoritme voordat de optimale parameters zijn gevonden. Wanneer de optimale oplossing niet gevonden kan worden en het maximaal aantal functie-evaluaties is bereikt wordt een waarschuwing gegeven. - **nobs**: Het aantal metingen dat is gebruikt om de model simulatie te fitten op de metingen. In ons geval is dat het aantal metingen in **oseries1**. - **noise**: parameters om aan te geven of het noise model is gebruikt bij de optimalisatie. Meer informatie over het noisemodel is hier te vinden [%%REFERENTIE%%](#) - **tmin** and **tmax**: de start en de eindtijd van de metingen die gebruikt worden om het model te optimaliseren. Deze kunnen worden opgegeven in de **ml.solve()** functie. Standaard wordt de start- en eindtijd van de **oseries** gebruikt. In dit geval is aangegeven dat als starttijd 1 januari 1990 gebruikt moet worden. - **freq**: de frequentie van het tijdreeksmodel. Standaard wordt een dagelijkse frequentie gebruikt [%%KLOPT DIT???%%](#). Dit betekent dat de gesimuleerde en gemeten waarde op dagbasis worden vergeleken indien er metingen beschikbaar zijn op een dag. - **warmup**: De inlooptijd van het model. De grondwaterstanden op een bepaalde dag hangen af van de externe invloeden (stresses) over een bepaalde periode voor deze dag. We weten de lengte van deze periode niet. Standaard wordt een inlooptijd van 10 jaar min de starttijd **tmin** van het model gebruikt. - **solver**: Het optimalisatie-algoritme dat wordt gebruikt om de gesimuleerde grondwaterstanden te fitten op de gemeten grondwaterstanden. Met de standaard instellingen wordt gebruikt gemaakt van de [least_squares](#) solver uit de [scipy](#) package.

Fit statistics Rechtsboven in het model rapport Deze statistieken geven informatie over de fit tussen gesimuleerde en gemeten grondwaterstanden. De volgende fit parameters worden getoond: - **EVP**: Explained variance ook wel bekend als de Nash Sutcliffe coëfficiënt. Een waarde van 100% betekent dat de gemodelleerde en gesimuleerde grondwaterstanden volledige overeenkomen. In de praktijk komt dat niet voor. Meestal wordt een fit hoger dan 60-80% acceptabel geacht om de modelresultaten verder te gebruiken. - **R2**: the R^2 waarde wordt ook wel de determinatiecoëfficiënt genoemd. Een waarde van 1 staat voor een optimale fit. - **RMSE**: de root mean squared error. Een

waarde van 0 staat voor een optimale fit. - AIC: the Aikake Information Criterion. - BIC: the Bayesian Information Criterion.

Parameters Dit deel geeft de parameter waarde weer die zijn gevonden met het optimalisatie algoritme. Behalve de optimale waarde wordt ook de standaard deviatie en de initiële waarde weergegeven. In de laatste kolom is aangegeven of de parameter wel **vary=True** of niet **vary=False** is meegenomen in de optimalisatie.

Parameter correlations Deze sectie geeft de correlaties tussen de parameters aan. Alleen de correlaties met een absolute waarde hoger dan 0.5 zijn weergegeven. Wanneer twee parameters een correlatie van 1 of -1 hebben betekent dit dat vergelijkbare modelresultaten bereikt kunnen worden met een verschillende set van de twee parameters.

%% Ter informatie, statistieken over andere pastas modellen %%

1.10 Statistieken van alle pastas voorbeeld modellen

```
<pandas.io.formats.style.Styler at 0x215c2152670>
```

1.11 Discussie

1.12 ## Referenties