

De bouwstenen van een tijdreeksmodel

R.A. Collenteur

February 26, 2021

1 Voorbewerking

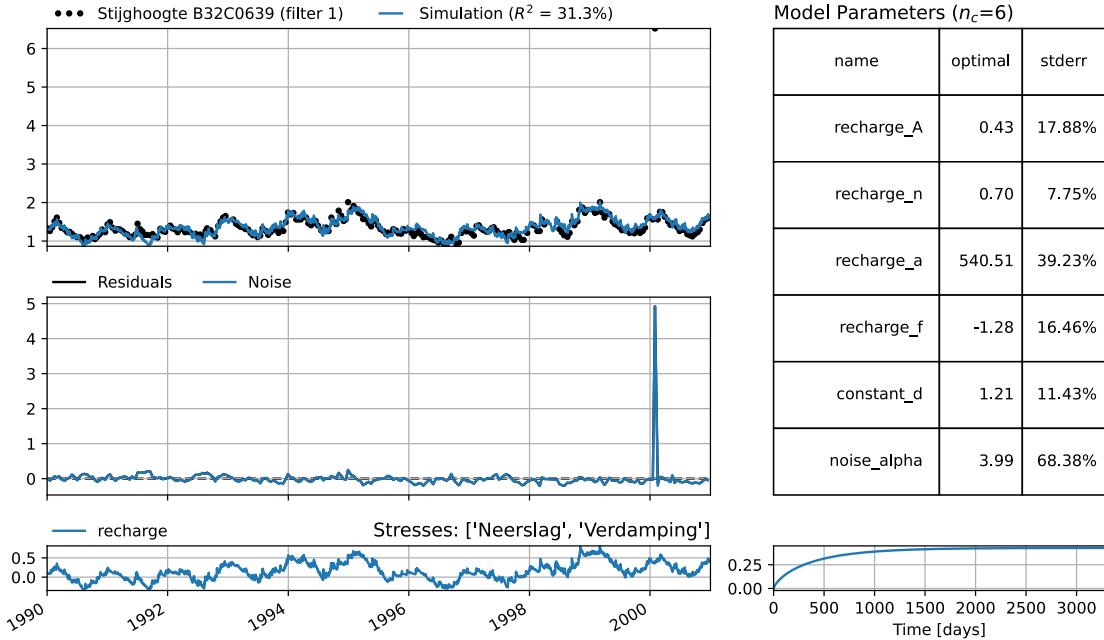
Voor tijdreeksanalyse zijn grondwaterstanden en verklarende reeksen nodig. Beide komen in verschillende lengtes, frequenties en kwaliteit. In veel gevallen is een voorbewerking van de reeksen nodig. In dit notebook wordt geïllustreerd welke bewerkingen je kan doen en wanneer deze nodig zijn. Onderstaande voorbeelden worden uitgewerkt:

Inhoudsopgave

- Opschonen stijghoogte reeks
- Vereenvoudigen stijghoogte reeks
- Equidistant maken verklarende reeks
- Voorgeschiedenis verklarende reeks
- Eenheden afstemmen

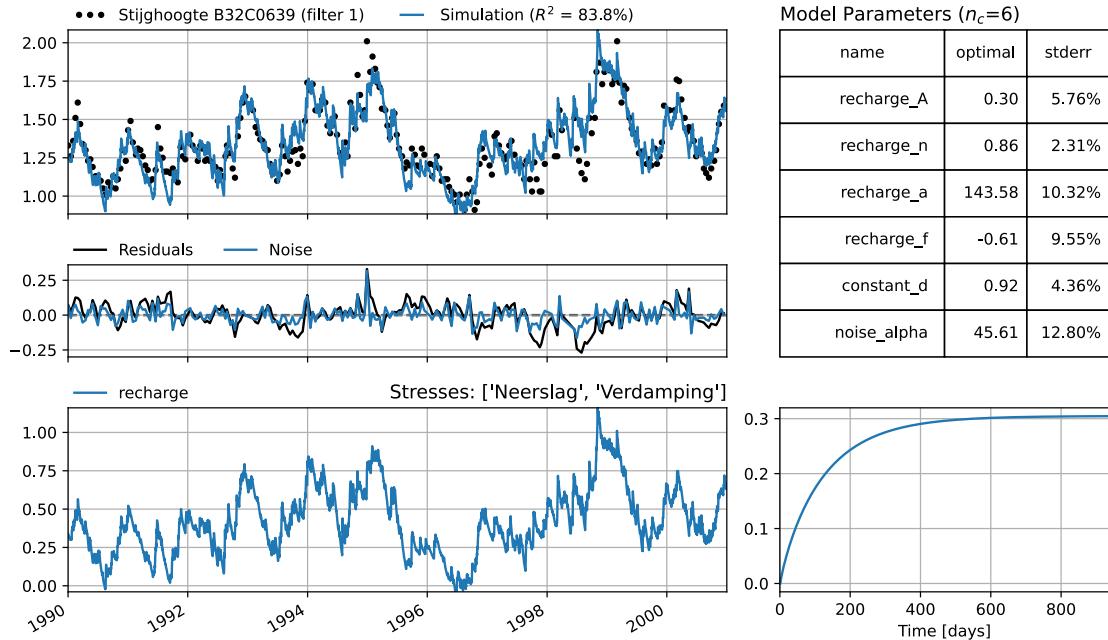
1.1 Opschonen stijghoogte reeks

Het doel van tijdreeksanalyse is een grondwaterstands of stijghoogte reeks te verklaren uit invloerreeksen. Door middel van automatische optimalisatie probeert de software de reeks zo dicht mogelijk te benaderen. Wanneer er fouten in de reeks aanwezig zijn kan dat grote gevolgen hebben voor de kwaliteit van het model. Onderstaand voorbeeld geeft een model dat is geoptimaliseerd zonder een duidelijke uitbijter te verwijderen.



De resultaten van dit model zijn niet erg bemoedigend. De verklaarde variantie bedraagt slechts 31%. Dit komt mede door de grote afwijking bij de uitbijter, maar ook de standaard-fouten van de parameters zijn relatief groot en de responsijd van de neerslag en verdamping lijkt voor nederlandse begrippen aan de hoge kant.

We gaan nu kijken wat de resultaten worden als we de uitbijter verwijderen.



In bovenstaande figuur is te zien dat het model veel beter presteert met de opgeschoonde reeks. Het blijkt dat de uitbijter (ook als is het er maar één) een grote invloed heeft op het resultaat. Naast een betere fit (84%), zijn ook de standaardfouten kleiner en is de respons korter geworden.

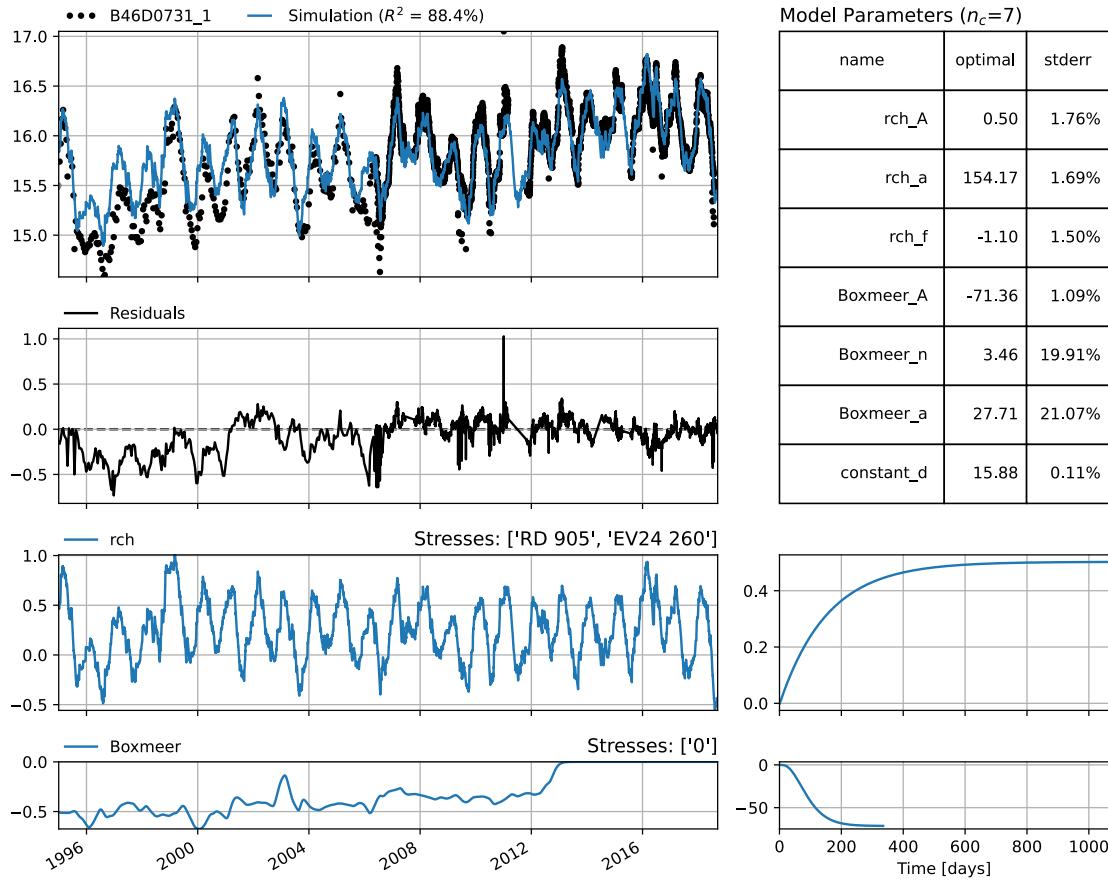
1.2 Vereenvoudigen stijghoogte reeks

Sommige reeksen bevatten zeer hoogfrequente data, met bijvoorbeeld een tijdstap van een dag of nog korter. Afhankelijk van het doel van de tijdreeksanalyse en de beschikbare invoer, kan het gewenst zijn om de reeksen te aggregeren of uit te dunnen. Het eerste voorbeeld betreft een veelvoorkomende situatie: er is gedurende lange tijd ongeveer tweewekelijks gemeten en sinds enkele jaren op een hoge frequentie. Wanneer we alle metingen gebruiken is het mogelijk dat de nadruk bij de optimalisatie komt te liggen op de periode waar hoogfrequent is gemeten, simpelweg omdat er meer metingen zijn. In het onderstaande voorbeeld is dat het geval.

```
Fit report B52B0389           Fit Statistics
=====
nfev    29                      EVP      88.45
nobs   4086                     R2       0.88
noise  False                    RMSE     0.13
tmin   1995-01-01 00:00:00      AIC      nan
tmax   2018-09-14 00:00:00      BIC      nan
freq   D                        Obj      33.42
warmup 3650 days 00:00:00      ---
solver LeastSquares            Interpolated No

Parameters (7 were optimized)
=====
              optimal    stderr    initial   vary
rch_A        0.502702  ±1.76%    0.204425  True
rch_a        154.165362  ±1.69%   10.000000  True
rch_f        -1.097611  ±1.50%   -1.000000  True
Boxmeer_A    -71.360214  ±1.09%  -497.068840  True
Boxmeer_n     3.459625  ±19.91%   1.000000  True
Boxmeer_a     27.714898  ±21.07%  10.000000  True
constant_d   15.883111  ±0.11%   15.907175  True

Parameter correlations |rho| > 0.5
=====
rch_A      rch_a      0.74
          rch_f      0.53
          constant_d -0.71
rch_f      constant_d -0.96
Boxmeer_n Boxmeer_a  -0.98
```



Vooral in en begin van de reeks zijn de gesimuleerde waarden structureel hoger dan de metingen, terwijl in de tweede helft van de reeks de simulatie gemiddeld beter klopt. Het onderstaande voorbeeld gaat uit van dezelfde invoer, alleen is nu de stijghoogtereeks geresampled zodat er gedurende de hele periode tweewekelijks metingen zijn.

```
Fit report B52B0389           Fit Statistics
=====
nfev      33                  EVP          87.63
nobs     561                  R2           0.88
noise    False                 RMSE         0.16
tmin    1995-01-01 00:00:00   AIC          nan
tmax    2018-09-11 00:00:00   BIC          nan
freq     D                    Obj          7.24
warmup   3650 days 00:00:00
solver   LeastSquares         Interpolated No
```

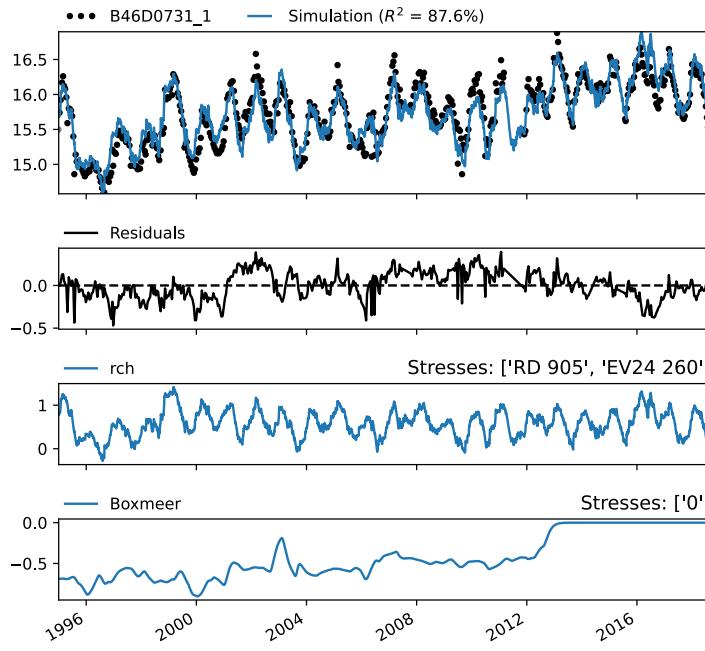
Parameters (7 were optimized)

```
=====
          optimal    stderr    initial   vary
rch_A       0.726368  ±3.29%   0.204425  True
```

rch_a	209.068284	$\pm 3.91\%$	10.000000	True
rch_f	-0.904330	$\pm 3.75\%$	-1.000000	True
Boxmeer_A	-95.933577	$\pm 1.94\%$	-497.068840	True
Boxmeer_n	2.326110	$\pm 22.04\%$	1.000000	True
Boxmeer_a	37.039832	$\pm 25.00\%$	10.000000	True
constant_d	15.580374	$\pm 0.32\%$	15.651168	True

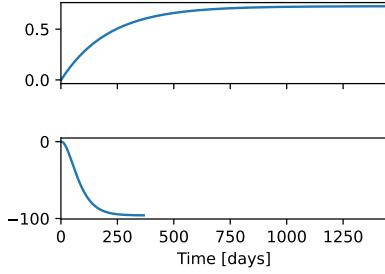
Parameter correlations $|\rho| > 0.5$

rch_A	rch_a	0.70
	constant_d	-0.60
rch_f	constant_d	-0.90
Boxmeer_n	Boxmeer_a	-0.96

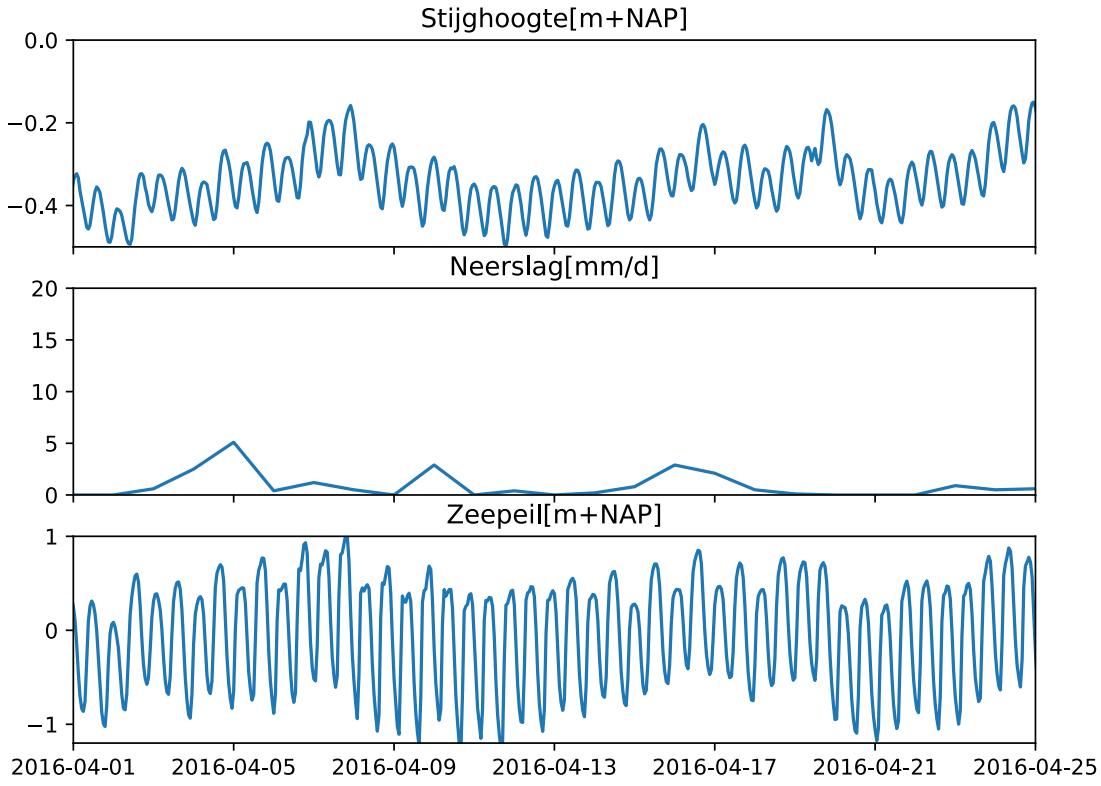


Model Parameters ($n_c=7$)

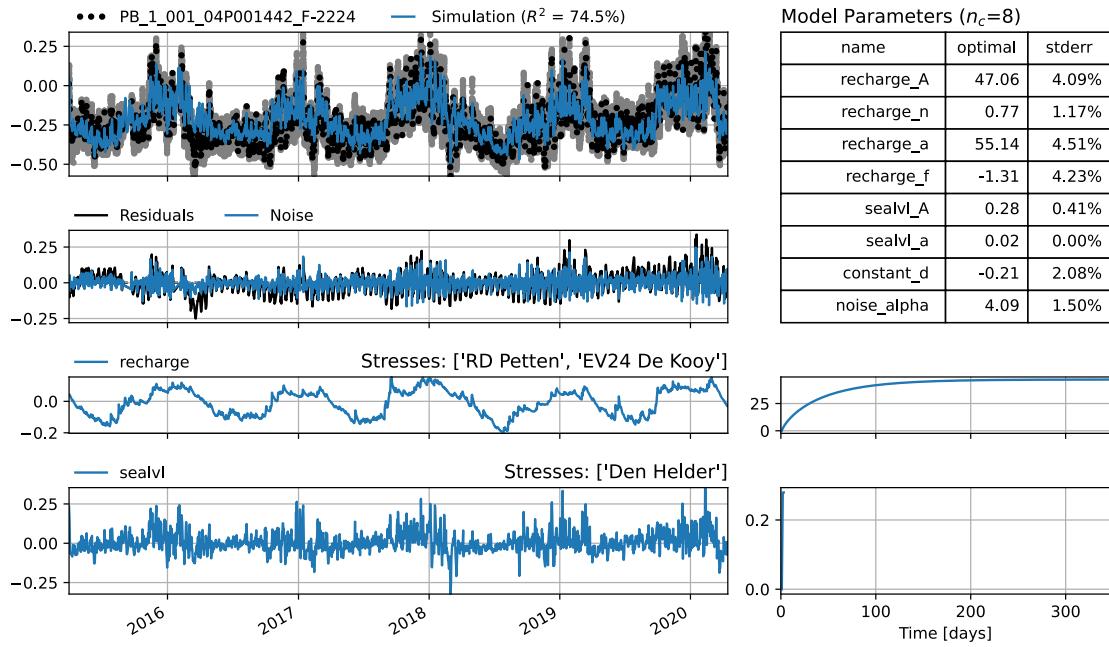
name	optimal	stderr
rch_A	0.73	3.29%
rch_a	209.07	3.91%
rch_f	-0.90	3.75%
Boxmeer_A	-95.93	1.94%
Boxmeer_n	2.33	22.04%
Boxmeer_a	37.04	25.00%
constant_d	15.58	0.32%



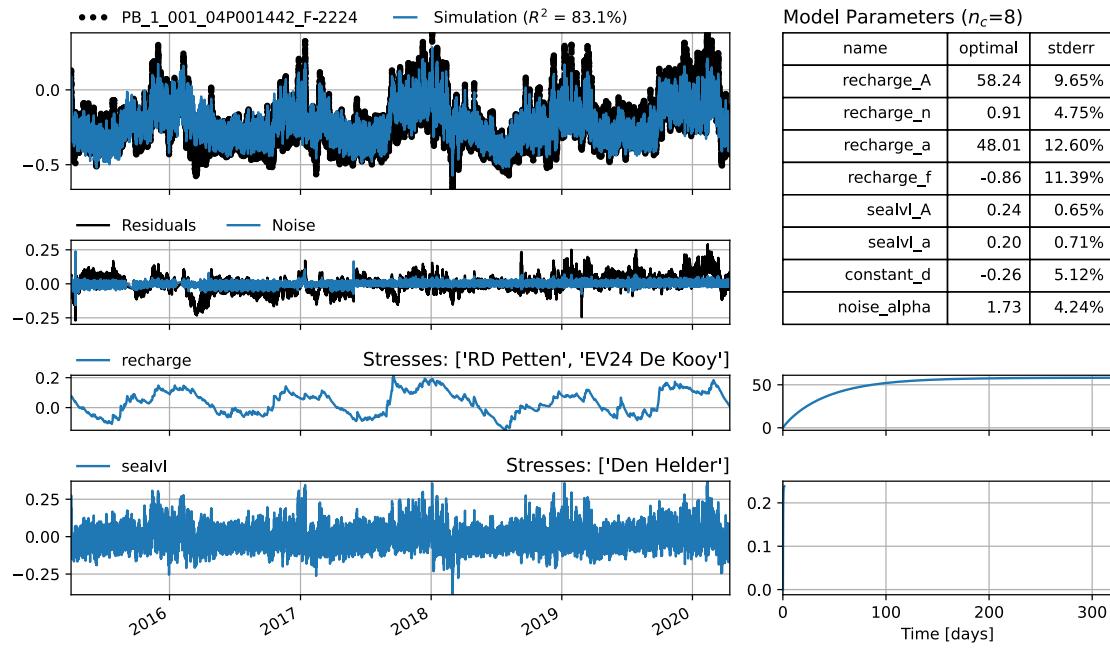
Na het filteren van de hoogfrequente data is te zien dat de residuen nu evenwichtiger zijn verdeeld over de reeks. De fit in de beginperiode is beter geworden ten koste van de fit aan het eind. Dit heeft ook geleid tot een ander model; ook de responsen van de bijdragen zijn gewijzigd. Er zijn ook situaties waarin het juist belangrijk is om een hoge frequentie te handhaven. In het volgende voorbeeld wordt dit geïllustreerd. Het gaat om een peilfilter dat onder invloed staat van het getij. In onderstaande figuur is goed te zien dat de stijghoogte wordt beïnvloed door het getij.



Als eerste poging maken we een model met een tijdsinterval van een dag. Onderstaande figuur toont het resultaat:



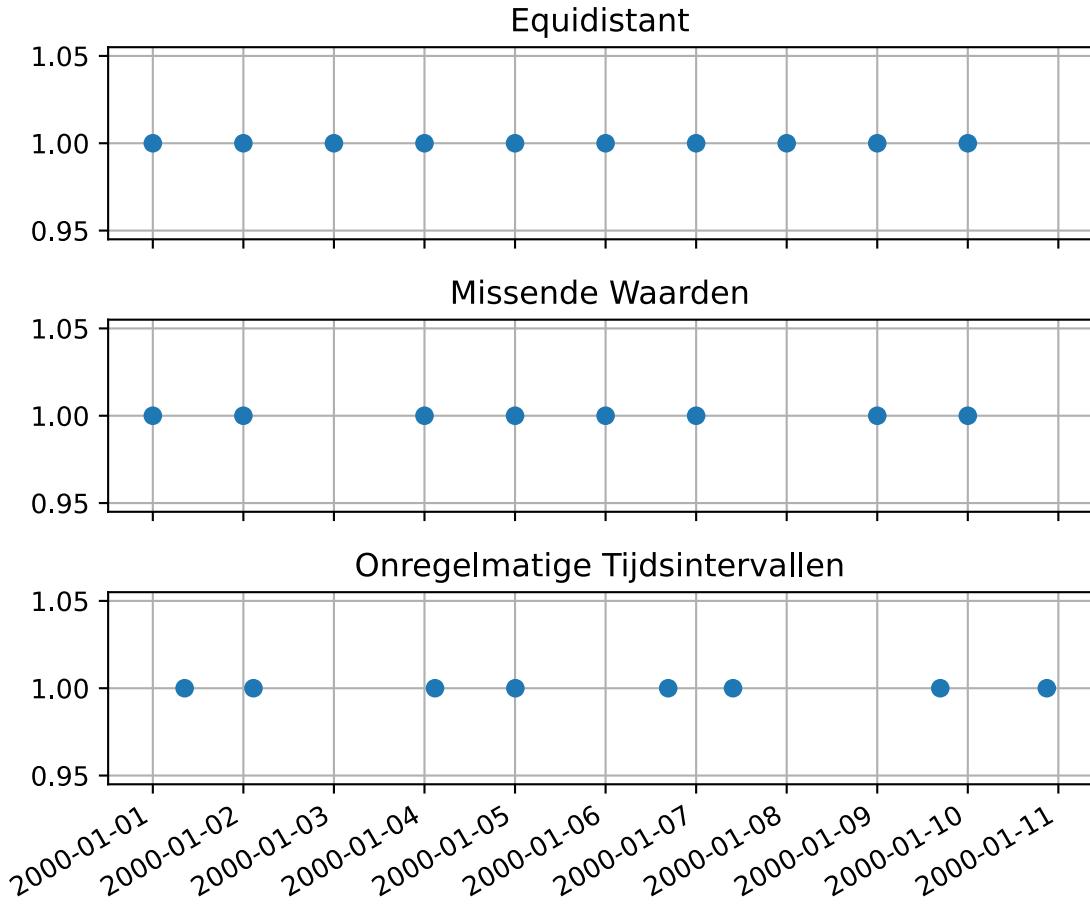
Het is duidelijk te zien dat de dagelijkse fluctuatie niet wordt meegenomen. Er is ook een duidelijk patroon in de residuen te zien. We proberen nu een simulatie op uurwaarde.



Dit model presteert veel beter, omdat de informatie die in het zeepeil aanwezig is beter gebruikt kan worden om de grondwaterstand te verklaren.

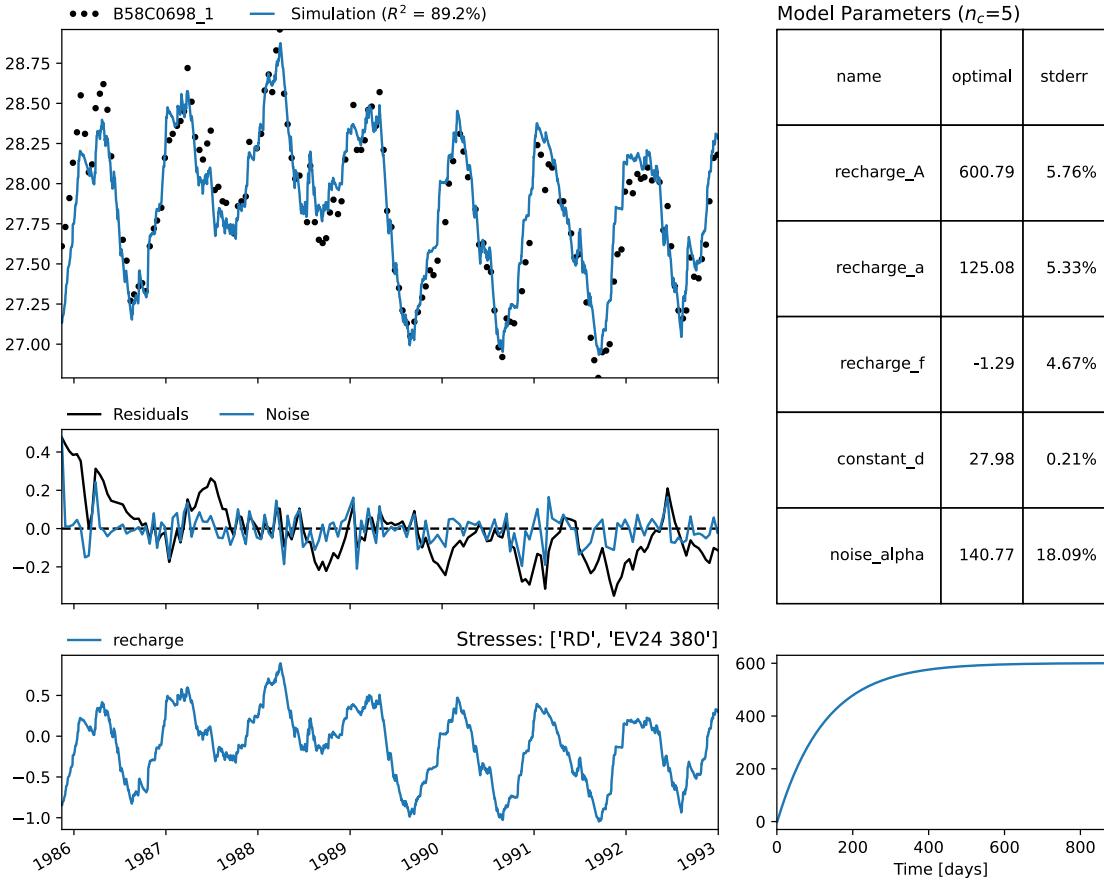
1.3 Equidistant maken verklarende reeks

Voor de meeste tijdreeksanalyseprogramma's is het nodig dat de verklarende reeksen gelijke tijdstappen hebben, ofwel equidistant zijn. In sommige gevallen worden ze zo aangeleverd. Als dat niet zo is, is het nodig om de reeksen equidistant te maken. De manier waarop dit gebeurt, hangt af van het type verklarende reeks. Rivierpeilen kunnen bijvoorbeeld geïnterpoleerd worden afhankelijk van de grootte van de tijdstap en de dynamiek van de reeks. Bij onttrekkingssreeksen moet men zich afvragen of een missende waarde betekent dat de onttrekking uit stond, of dat er gegevens ontbreken omdat er geen waarneming heeft plaatsgevonden. Het opvullen van de data introduceert onzekerheden, die niet automatisch worden meegenomen in de analyse. Men kan er ook voor kiezen de tijdstap van de tijdreeksanalyse te vergroten, wanneer dat beter bij de resolutie van de beschikbare gegevens past. In onderstaande figuur worden de verschillende type reeksen geillustreerd:

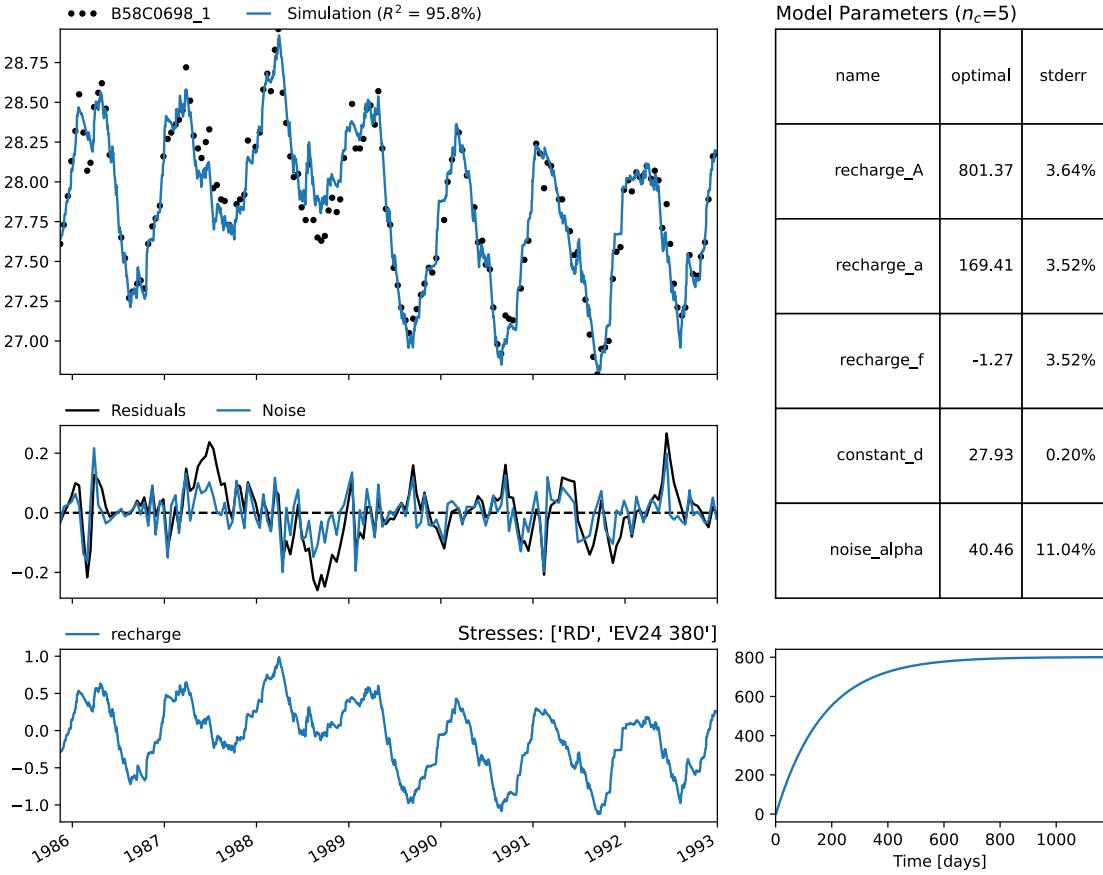


1.4 Voorgeschiedenis verklarende reeks

Elk hydrologisch systeem heeft een zeker geheugen. Dit betekent dat de stijghoogte afhangt van wat er een bepaalde periode ervoor is gebeurd. Daarom kent elk model een zekere inloopperiode. Bij het fitten van een grondwaterstand is het belangrijk dat er voldoende informatie over verklarende reeksen aanwezig is vóór de eerste waarneming waarop gekalibreerd wordt. Wanneer dat niet het geval is, zal er een schatting gemaakt moeten worden, óf de betreffende waarnemingen zullen niet in de optimalisatie meegenomen moeten worden. In onderstaand voorbeeld is de neerslag bekend gedurende een periode van 100 dagen vóór de eerste stijghoogtemeting. Daarvóór wordt aangenomen dat de neerslag 0 is.



Het is duidelijk te zien dat simulatie in het begin van de reeks lager ligt dan de metingen. Dit wordt gecompenseerd door een te hoge ligging aan het eind van de reeks. In onderstaand voorbeeld nemen we hetzelfde model, maar nu gaan we uit van de gemiddelde neerslag tijdens de inloopperiode.

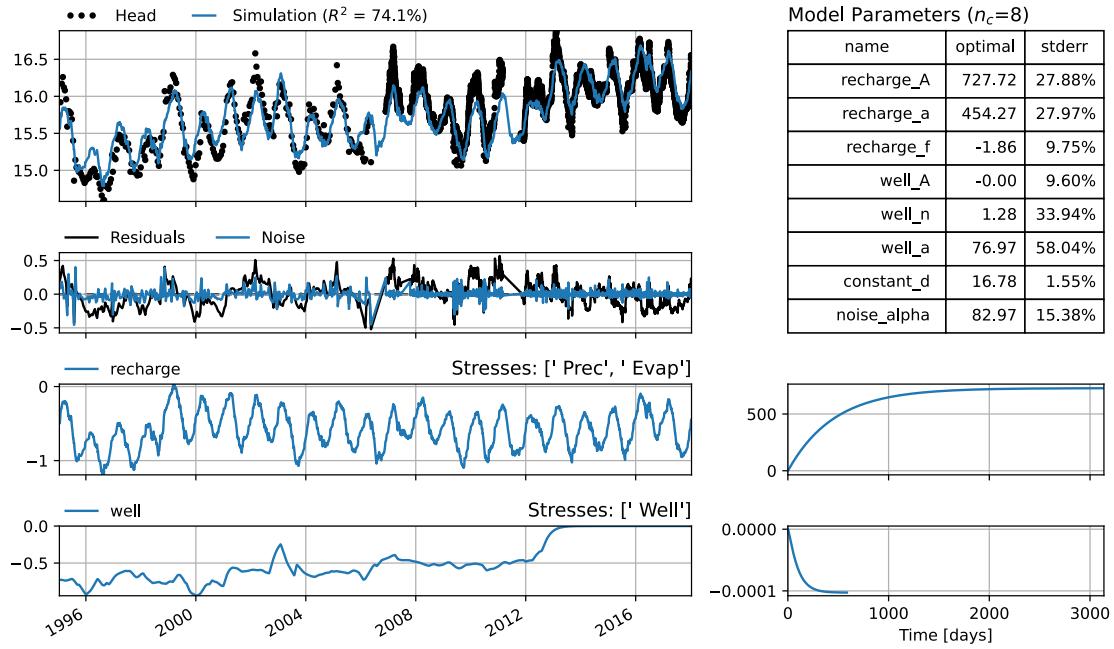


Nu het gemiddelde wordt genomen vóór dat er neerslagdata aanwezig is, is er geen trend in de residuen meer aanwezig. De modelfit is ook verbeterd, en de respons van de neerslag ziet er ook significant anders uit. De gevonden responstijd van de neerslag is veel groter dan 100 dagen. Daarom is het van belang neerslag voldoende neerslaggevens te hebben of een realistische schatting, indien deze niet aanwezig zijn.

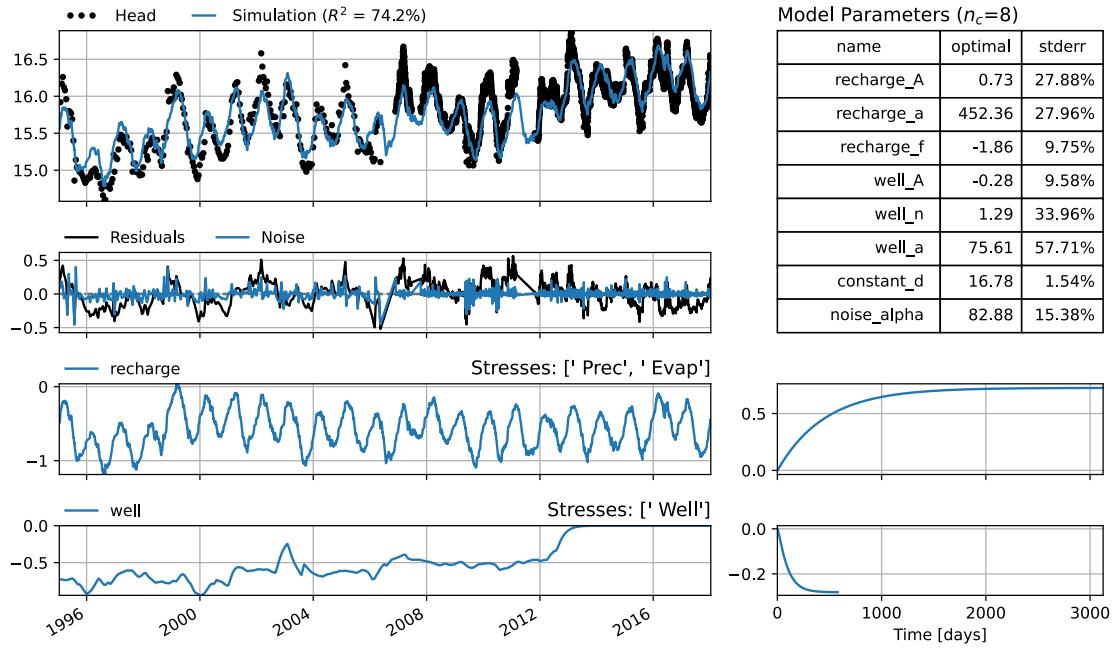
1.5 Eenheden afstemmen

Voor de interpretatie van de resultaten (en soms ook voor het succes van de kalibratie) is het gewenst om de eenheden van de verschillende reeksen goed te kiezen en eventueel op elkaar af te stemmen.

In onderstaand voorbeeld proberen we de invloed van een onttrekking vast te stellen. De invoer is aangeleverd op dagbasis in standaardeenheden. Neerslag en verdamping worden geleverd in meters per dag, en de onttrekking in kubieke meters per dag. We zijn geïnteresseerd in het totaal-effect van de onttrekking ter hoogte van de waarnemingsput. Eerst doen we een tijdreeksanalyse met de eenheden zoals we die zijn aangeleverd:



Het effect van de put (de “gain”) is gelijk aan de parameter `well_A` in de tabel. Deze heeft de eenheid van stijghoogte gedeeld door de eenheid van de onttrekking. In dit geval is dat dus een verlaging (in meter) bij een onttrekking van één kubieke meter per dag. Vanzelfsprekend is dat een heel klein getal, en in de tabel wordt dit zelfs afgerond naar 0.00. De gain van de recharge stelt de stijghoogte-verhoging voor die zou ontstaan als het continu één meter per dag zou regenen (737 meter). Evenzo is de verlaging door één meter verdamping per dag gelijk aan $1.86 \times 737 = 1370$ meter. Deze eenheden zijn niet erg praktisch voor de interpretatie. Daarom doen we deze tijdreeksanalyse nog eens opnieuw, maar dan veranderen we de eenheden van verdamping en neerslag naar mm/dag, en de onttrekking in miljoen kubieke meters per jaar.



Bovenstaande figuur toont de resultaten met de gewijzigde eenheden. Wat opvalt is dat ook na omrekening de getallen iets verschillen. Dit zijn afwijkingen die ontstaan bij de optimalisatie van de parameters, die nu iets anders verloopt omdat de verhoudingen tussen de invoerreeksen verschillen. De gain bij één millimeter neerslag per dag is nu 0.72 meter. De verlaging ten gevolge van een onttrekking van één miljoen kuub per jaar is 0.28 meter. Om de eindverhoging bij stopzetting van de winning te schatten moet deze waarde vermenigvuldigd worden met de werkelijke wincapaciteit. (in miljoen kuub per jaar). Wanneer het een winning van 2 miljoen kuub per jaar betreft, schatten we de verhoging op 56 centimeter.