kalibratie_notebook

January 25, 2021

Inhoudsopgave

0.1 Kalibratie van een tijdreeksmodel

Onder kalibratie van een tijdreeksmodel verstaan we het vinden van de modelparameters zodat de (met het tijdreeksmodel) gesimuleerde stijghoogtes zo goed mogelijk overeenkomen met de gemeten stijghoogtes. De gevonden parameters worden de optimale parameters genoemd. De meest gebruikte methode om de optimale parameters te vinden is de minimalisatie van de de som van de kwadraten van de verschillen tussen de gemeten stijghoogtes en de gemodelleerde stijghoogtes, ook wel de kleinste kwadraten methode genoemd. De som van de kwadraten is een niet-lineaire functie van de modelparameters. Om de optimale modelparameters te vinden wordt gebruikt gemaakt van een zoekmethode.

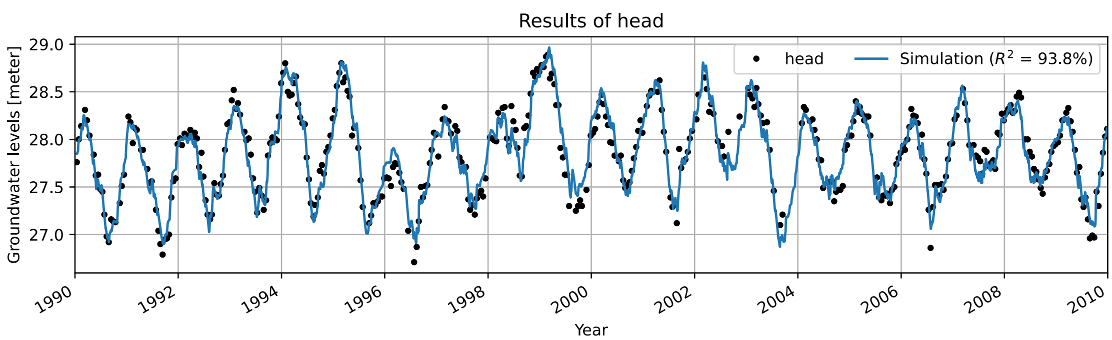
Er bestaan een aantal verschillende zoekmethoden die gebruikt kunnen worden om de optimale modelparameters te vinden. Iedere zoekmethode is gebaseerd op een ander zoek algoritme, maar het idee is altijd hetzelfde: het model wordt een aantal keer gedraaid met verschillende waarden van de modelparameters. Gebaseerd op de verschillende modeluitkomsten (en de daaruit voorvloeiende som van de kwadraten) bepaald het zoekalgoritme een volgende mogelijke set modelparameters die een kleinere som van de kwadraten zou kunnen opleveren. Het zoekalgoritme gaat door totdat het besluit dat de beste parameters gevonden zijn. De efficientie van het zoekalgoritme wordt bepaald door het aantal keer dat het model gedraaid moet worden om de optimale modelparameters te vinden. Het is vanzelfsprekend aan te bevelen om de initiele modelparameters zo dicht mogelijk bij de optimale modelparameters te kiezen. Daartoe heeft de meeste software geprobeerd om slimme keuzes te maken voor de initiele waarden van de modelparameters. Soms heeft de modelleur specifieke kennis van het systeem waardoor betere initiele waarden gekozen kunnen worden.

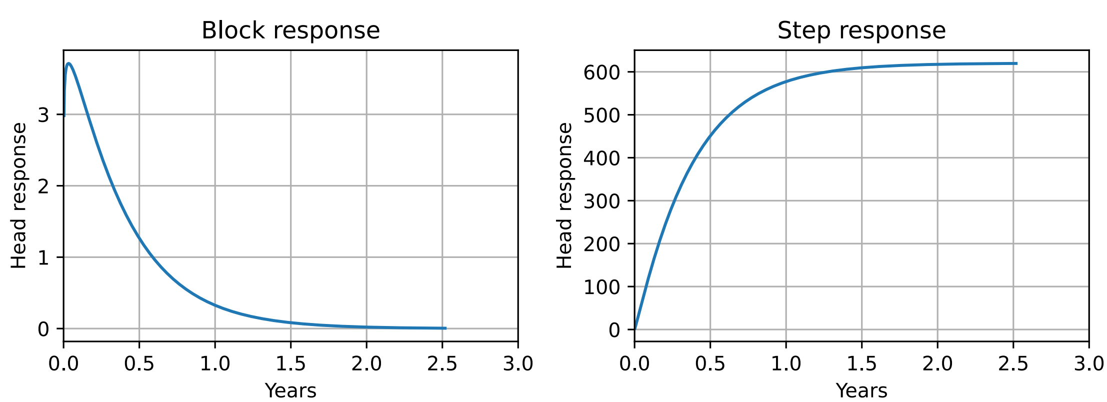
De meest gebruikte zoekalgoritmes zoeken de optimale modelparameters door de modelparameters steeds een beetje aan te passen zodat de som van de kwadraten steeds een beetje kleiner wordt. Op deze manier leidt er een direct pad van de initiele waarden naar de optimale waarden en bij elk stapje langs het pad is de som van de kwadraten kleiner. Dit kan tot gevolg hebben dat het zoekalgoritme in een lokaal minimum belandt. Het is namelijk goed mogelijk dat langs het pad van de initiele parameter waarden naar de optimale parameter waarden de kwadraten som een tijdje toeneemt voordat het weer afneemt. Er bestaan geavanceerde zoekalgoritmes die het model ook draaien met modelparameters die juist een grotere kwadraten som geven met als doel om uiteindelijk het globale minimum te vinden. Dit kost vanzelfsprekend (veel) meer rekentijd.

In dit Notebook wordt ingegaan op twee aspecten van modelcalibratie. Het eerste aspect is de vraag hoe lang een reeks moet zijn om goed te kunnen kalibreren. De tweede vraag is wanneer er een ruismodel gebruikt moet worden.

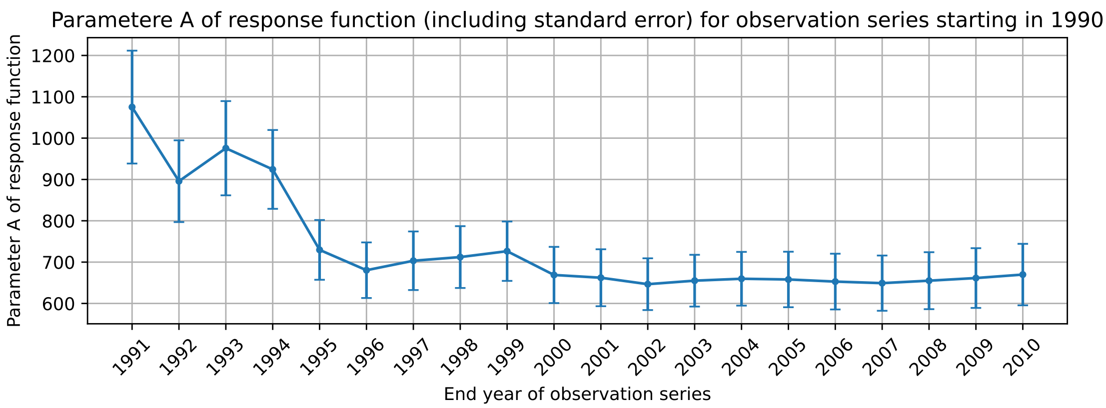
0.2 Hoe lang moet een tijdreeks zijn om goed te kunnen kalibreren?

Beschouw een reeks stijghoogtes van 20 jaar waarbij de stijhoogte ongeveer twee keer per maand gemeten is. Er is een tijdreeksmodel gemaakt waarbij de regen en potentiële verdamping de verklarende reeksen zijn en de geschaalde Gamma functie gebruikt wordt als response functie. De hele reeks wordt gebruikt om het model te kalibreren. Het model fit de data goed, zoals in onderstaande figuur te zien is. De blokresponse en stapresponse functies worden ook getoond. Het geheugen van de blokresponse functie is iets meer dan 2.5 jaar. Na een regenbui op de eerste dag gaat de stijghoogte vrij snel omhoog, maar het duurt ongeveerd 2.5 jaar voordat de stijghoogte weer terug is op het niveau voor de regenbui. De piek van de blokresponse functie is ongeveer 3.8, hetgeen betekent dat als het 1 mm regent, dat dan de stijghoogte uiteindelijk 3.8 mm omhoog gaat. Dat komt overeen met een effectieve porositeit (freatische berging) van 1 / 3.8 = 0.26, wat een redelijke waarde is. De stapresponse gaat uiteindelijk naar een waarde iets boven 600. Dat betekent dat als het vanaf de eerste dag continue 1 mm/d regent, dat dan de stijghoogte uiteindelijk iets meer dan 600 mm omhoog gaat. Dit is gelijk aan de parameter A van Gamma response functie (zie Notebook over modelstructuur).





In het model hierboven wordt 20 jaar gebruikt om het model te kalibreren. Wat nu als er een kortere reeks gebruikt wordt? Wordt dan dezelfde responsefunctie gevonden? Om dit te onderzoeken wordt het model gekalibreerd op 1 jaar data (1990-1991), twee jaar data (1990-1992), etc. tot 20 jaar data (1990-2010). Voor elke kalibratie wordt de waarde van de parameter A opgeslagen (de uiteindelijke hoogte van de stapreponse). In onderstaande figuur wordt de gekalibreerde waarde van A, inclusief 95% betrouwbaarbeidsinterval, geplot tegen de lengte van de kalibratieperiode. Uit de figuur blijkt dat als de kalibratieperiode lang genoeg is, dat er steeds ongeveer hetzelfde uitkomt. Voor dit geval is een klaibratieperiode van 5 jaar (1990-1995) genoeg. Dat is, voor dit geval, ongeveer twee keer het geheugen van de blokresponse.



Intuitief lijkt het te kloppen dat de kalibratieperiode een aantal keer het geheugen van het systeem moet zijn. Het is echter moeilijk om een vaste regel te geven voor de benodigde lengte van de kalibratieperiode. Van der Spek en Bakker (2017) hebben onderzocht wat de minimale lengte van een kalibratieperiode is waarna de onzekerheid van voorspelde stijghoogtes niet verder afneemt. Zij onderzochten 18 reeksen met responsetijden die varieerden van 60 tot 1200 dagen, maar konden geen duidelijke relatie vinden tussen het geheugen van het systeem (de responsetijd) en de lengte van de kalibratieperiode. Voor de onderzochte reeksen wat het gemiddelde 95% voorspellingsinterval ongveer 20% van het verschil tussen de hoogst en laagst gemeten stijghoogte bij een kalibratieperiode van 5 jaar, maar waren stuk beter voor kalibratieperiodes van 10 jaar, hoewel het wel uitmaakte welk 10 jaar gebruikt werd voor kalibratie.

0.3 Wanneer moet een ruismodel gebruikt worden?

Een ruismodel kan gebruikt worden om te proberen er voor te zorgen dat de overgebleven verschillen tussen het model en de metingen aan een aantal statistische toetsen voldoen zodat er statistische uitspraken gedaan kunnen worden met het model (bijvoorbeeld: wat is de kans dat de stijghoogte boven een bepaalde grenswaarde uitkomt).

De verschillen tussen de gemeten stijghoogte en de gemodelleerde stijhoogte (de residuën) is vrijwel altijd gecorreleerd in de tijd. Simpel gezegd: als het model vandaag iets hoger is dan de gemeten stijghoogte, dan is de kans groot dat het model volgende week ook iets hoger is dan de gemeten stijghoogte. Dat is op zich prima te verklaren. Het kan bijvoorbeeld zo zijn dat er een regen reeks gebruikt wordt van een weerstation dat 20 km verderop ligt. Als het nu bij het regenstation wel geregend heeft maar bij de peilbuis niet, dan zal het model waarschijnlijk een te hoge stijghoogte simuleren. En aangezien het een tijdje duurt voordat de stijhoogte weer gezakt is na een regenbui, zal het model waarschijnlijk voor een aantal dagen/weken/maanden te hoge waarden geven. Gelukkig is de kans groot dat het een tijdje later bij de peilbuis juist wel regent terwijl het bij het regenstation niet regent, zodat de effecten elkaar weer een beetje opheffen.

De correlatie in de residuën van een tijdreeksmodel zijn op zich geen probleem, behalve als de gebruiker het model wil gebruiken om statistische uitspraken te doen. Als er statistische uitspraken gedaan moeten worden, dan moeten de residuën aan bepaalde statistische voorwaarwaarden voldoen. Eén van de belangrijkste voorwaarden is dat de residuën niet gecorreleerd zijn. Een manier om te trachten dit te bewerkstelligen is om een ruismodel te gebruiken. Het doel van het ruismodel is om de gecorreleerde residuën te transformeren naar ruis die ongecorreleerd is. De eenvoudigste manier om te trachten dit te doen is door het gebruik van een AR1 ruismodel. In dit model is het residu $\varepsilon(t)$ op tijstip t gelijk aan een factor ρ keer het residu op tijdstip $t - \Delta t$ plus een random (ongecorreleerde) fout $n(t)$, of in formule vorm

$$\varepsilon(t) = \rho\varepsilon(t - \Delta t) + n(t)$$

Als de tijd Δt tussen twee residuën varieert, dan kan de factor ρ een functie zijn van Δt die afneemt naarmate de tijdstap Δt tussen twee metingen groter wordt. De formule is dan

$$\varepsilon(t) = \mathrm{e}^{-\Delta t/\alpha}\varepsilon(t - \Delta t) + n(t)$$

waarbij α een parameter is die aangeeft hoe snel de correlatie tussen residuën afneemt als δt toeneemt. De correlatie is verwaarloosbaar als $\Delta t \approx 3\alpha$ (want $\mathrm{e}^{-3} \approx 0.05$).

In onderstaande voorbeeld worden synthetische reeksen gegeneerd en geanalyseerd om te laten zien dat bij tijdreeksmodellen zonder ruismodel de parameters wel goed geschat worden, maar de betrouwbaarheidsintervallen onderschat worden.

Voor het maken van de synthetische reeks wordt de dagelijkse grondwateraanvulling gelijk gesteld aan de gemeten regen min de gemeten potentiele verdamping van het vorige voorbeeld (dat is dus echte regen en verdampings data). De response functie van de grondwateraanvulling is een exponentiele functie met $A = 600$ en $a = 150$ d. Het basisniveau is $d = 25$ m. Een synthetische stijghoogte reeks wordt gesimuleerd op dezelfde tijdstippen waarop gemeten was in het vorige voorbeeld. Deze gegenereerde synthetische reeks bevat geen fouten en een tijdreeksmodel met een exponentiele response functie kan de parameters vrijwel exact terugvinden.

Om de werking van het ruismodel te illustreren, wordt er een residu (een fout) opgeteld bij de synthetische stijghoogte reeks. Het residu op tijdstip t, $\epsilon(t)$, is gecorrleerd met het residu bij de vorige meeting op tijdstip $t - \Delta t$ volgens de formule die hierboven al gegeven was. De ruis $n(t)$ wordt getrokken uit een normale verdeling met gemiddelde nul en standaard afwijking σ_n. Hoe groter de correlatie tussen de residuën is, hoe groter de de standaardafwijking van de residuën. De standaard afwijking σ_r van de residuën kan berekend worden uit de standaardafwijking σ_n van de ruis als

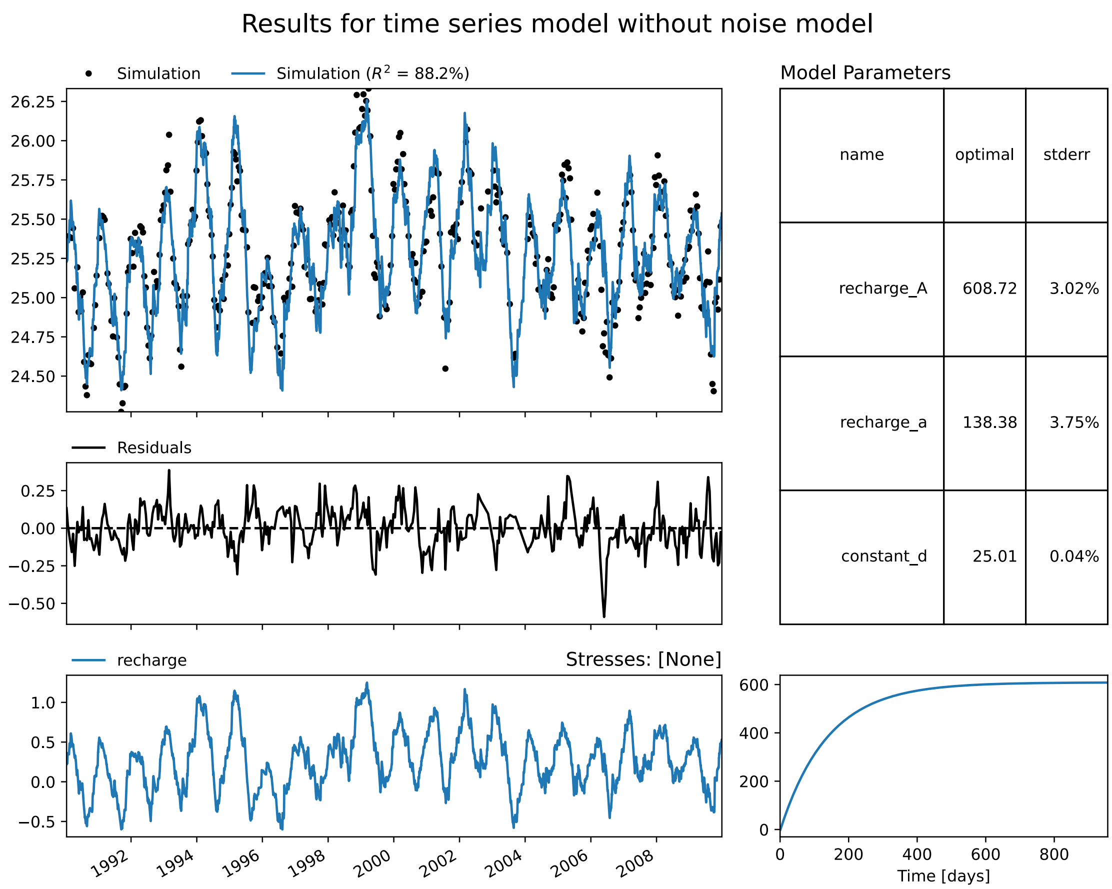
$$\sigma_r = \sigma_n / \sqrt{1 - \mathrm{e}^{-2\Delta t/\alpha}}$$

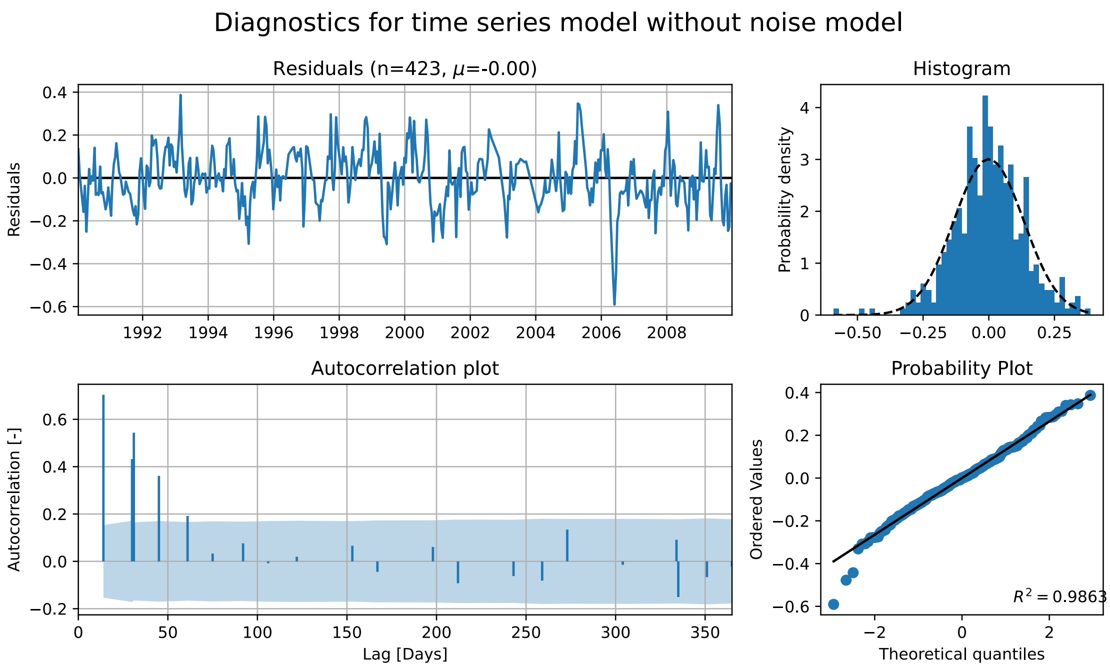
De residuën worden gegenereerd met $\sigma_n = 0.1$ m, $\alpha = 50$ d. De standaardafwijking van de residuën, σ_r, voor een tijdstap van 14 dagen is dan gelijk aan

```
sigma_r equals 0.15 m
```

De synthetische stijghoogte reeks bevat nu een gecorreleerde fout, net als in de werkelijkheid. We gaan nu eerst een tijdreeksmodel maken van de synthetische reeks zonder ruismodel. Het tijdreeksmodel blijkt een goede fit te geven en de geschatte parameters komen goed overeen met de parameters waarmee de synthetische reeks gemaakt wa. De diagnostische test laat echter zien dat er een autocorrelatie in de residuën zit.

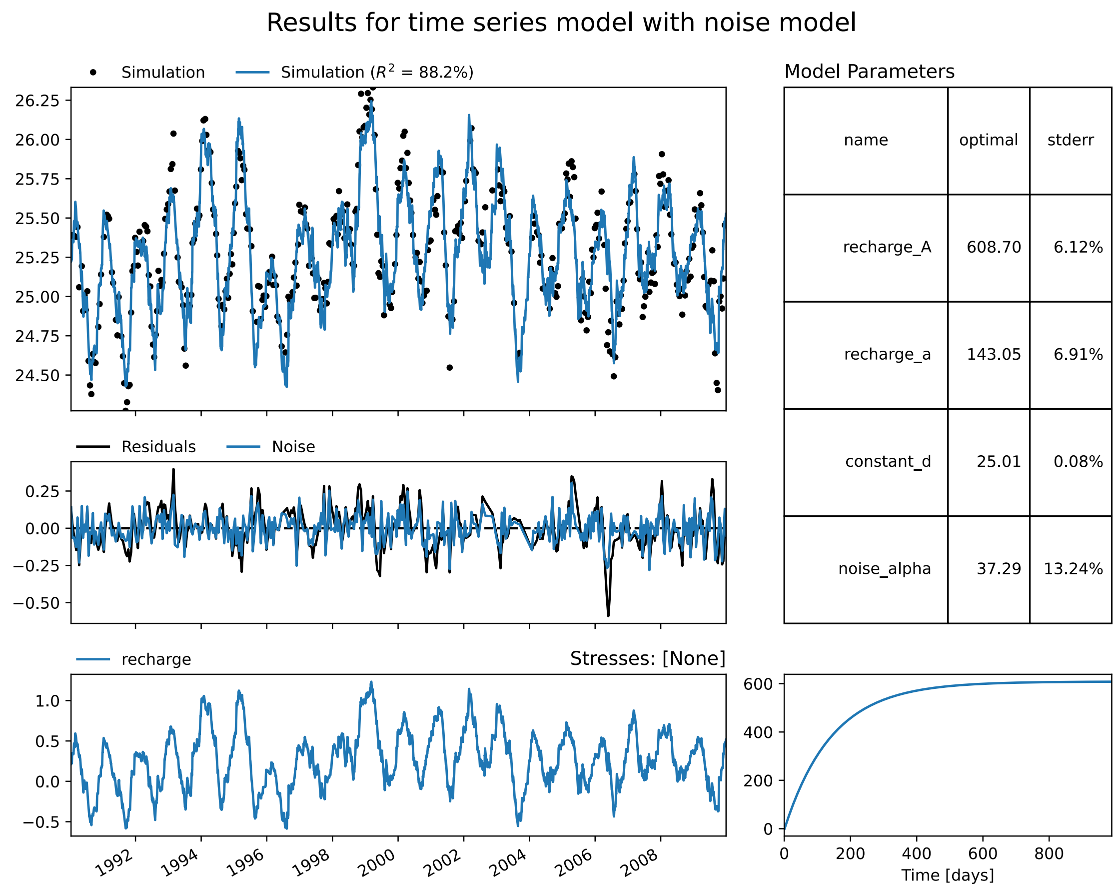
```
/Users/mark/git/pastas/pastas/stats/tests.py:616: UserWarning: FixedFormatter
should only be used together with FixedLocator
  ax.set_xticklabels(ax.get_xticklabels(), rotation=0, ha="center")
```

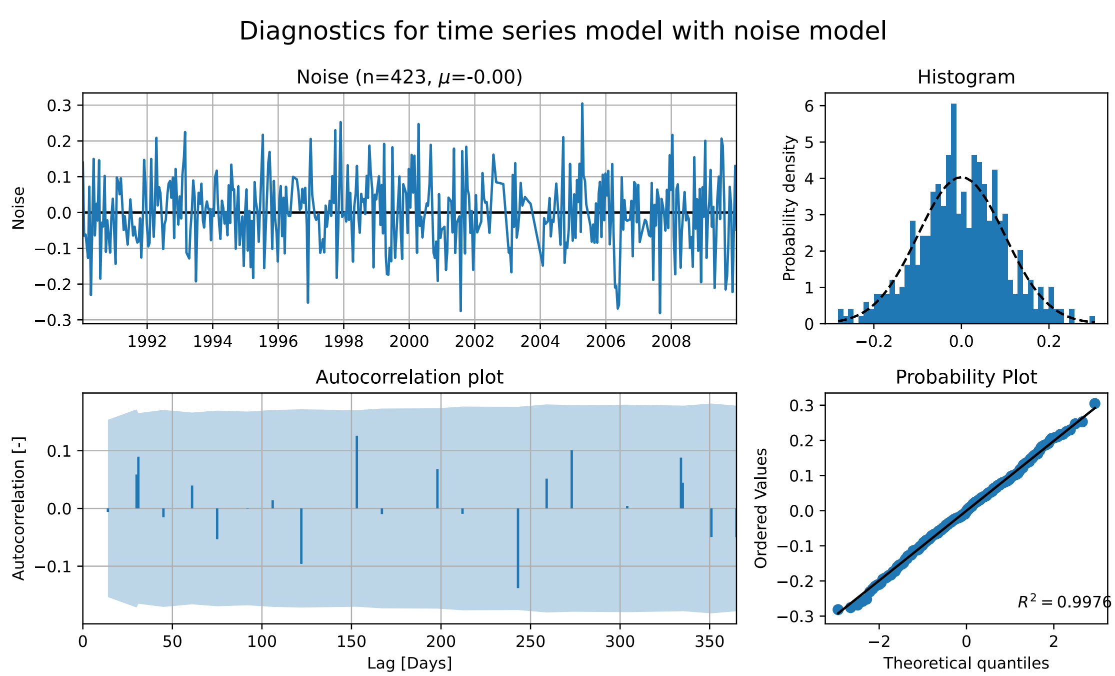




Als we nu de wederom een tijdreeksmodel maken van de synthetische stijghoogte reeks, maar dit keer wel een ruismodel toevoegen, dan is de fit wederom goed en komen de geschatte parameters wederom goed overeen met de opgegeven parameters, maar nu is de autocorrelatie van de noise binnen het geaccepteerde niveau. Het ruismodel heeft zijn werk goed gedaan. De fit van tijdreeksmodel met een ruismodel is altijd lager dan de fit van een tijdreeksmodel met ruismodel. Het verschil is hier echter klein.

```
/Users/mark/git/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")
```


Results for time series model with noise model
Simulation
Simulation (R^2 = 88.2%)
Model Parameters
name | optimal | stderr
recharge_A | 608.70 | 6.12%
recharge_a | 143.05 | 6.91%
constant_d | 25.01 | 0.08%
noise_alpha | 37.29 | 13.24%
Residuals
Noise
recharge
Stresses: [None]
Time [days]


Diagnostics for time series model with noise model
Noise (n=423, μ=-0.00)
Noise
Histogram
Probability density
Autocorrelation plot
Autocorrelation [-]
Lag [Days]
Probability Plot
Ordered Values
Theoretical quantiles
$R^2 = 0.9976$

Vervolgens gaan we een experiment doen. We creëeren 100 keer een synthetische reeks en gaan daar 100 keer een tijdreeksmodel van maken. Daar de synthetische reeks elke keer andere (gecorreleerde) redisuën heeft, zijn de optimale parameters elke keer een beetje anders. Het experiment is twee keer gedaan. De eerste keer is er 100 keer een tijdreeksmodel gemaakt zonder een ruismodel en vervolgens is er 100 keer een tijdreeksmodel gemaakt met een ruismodel.

```
0, 20, 40, 60, 80,
```

```
Results without a noise model
Mean value of parameter A: 603.3330970429524
Mean value of parameter a: 151.41435163520478
Mean value of parameter d: 24.997355612478383
Mean value of parameter d: 85.16714672614877
```

```
0, 20, 40, 60, 80,
```

```
Results with a noise model
Mean value of parameter A: 600.6379520027002
Mean value of parameter a: 150.38185608575677
Mean value of parameter d: 24.99853419247898
Mean value of parameter alpha: 45.867493779336
Mean value of evp: 85.11528481603865
```

Het blijkt dat zowel het tijdreeksmodel zonder het ruismodel als het tijdreeksmodel met het ruismodel gemiddeld een goede schatting geeft van de parameters. Dus waarom moeten we dan een ruismodel gebruiken? De reden is dat de geschatte onzekerheid van de parameters te klein is als er geen ruismodel gebruikt wordt. De statistische manier waarop de onzekerheid geschat wordt gaat er namelijk van uit dat de overgebleven ruis niet gecorreleerd is. Als de ruis echter toch gecorreleerd is, dan is de geschatte onzekerheid te klein. Dit kan gemakkelijk nagegaan worden met het hierboven uitgevoerde experiment. Tijdens het experiment zijn de parameters 100 keer geschat inclusief een schatting van de standaard fout van de parameters. Onder enkele statistische aannames (inclusief de aannames van normaliteit en ongecorreleerde ruis), is het 95% betrouwbaarheidsinterval tussen min twee keer de stanaard fout tot plus twee keer de standaar fout (eigenlijk is het 1.96 i.p.v. 2, maar laten we dan hier maar afronden). Er kan dus eenvoudig nagegaan worden hoe vaak de echte waarde van de parameters (waarmee de synthetische reeks gegenereerd is) ligt binnen het 95% betrouwbaarheidsinterval van de geschatte parameters. Aangezien we het experiment 100 keer gedaan hebben, zou dit ongeveer 95 keer het geval moeten zijn. Zoals te zien is hieronder, geeft het tijdreeksmodel zonder ruismodel slechts zo'n 60-65 keer een betrouwbaarheidsinterval dat de echte waarde bevat. Het tijdreeksmodel met een ruismodel doet het veel beter: meer dan 95 keer liggen de echte parameters van de responsen functie binnen het 95% betrouwbaarheidsinterval.

```
results without noise model
number of times true A within estimated confidence interval: 67
number of times true a within estimated confidence interval: 72
number of times true d within estimated confidence interval: 53
```

```
results without noise model
number of times true A within estimated confidence interval: 96
number of times true a within estimated confidence interval: 96
```

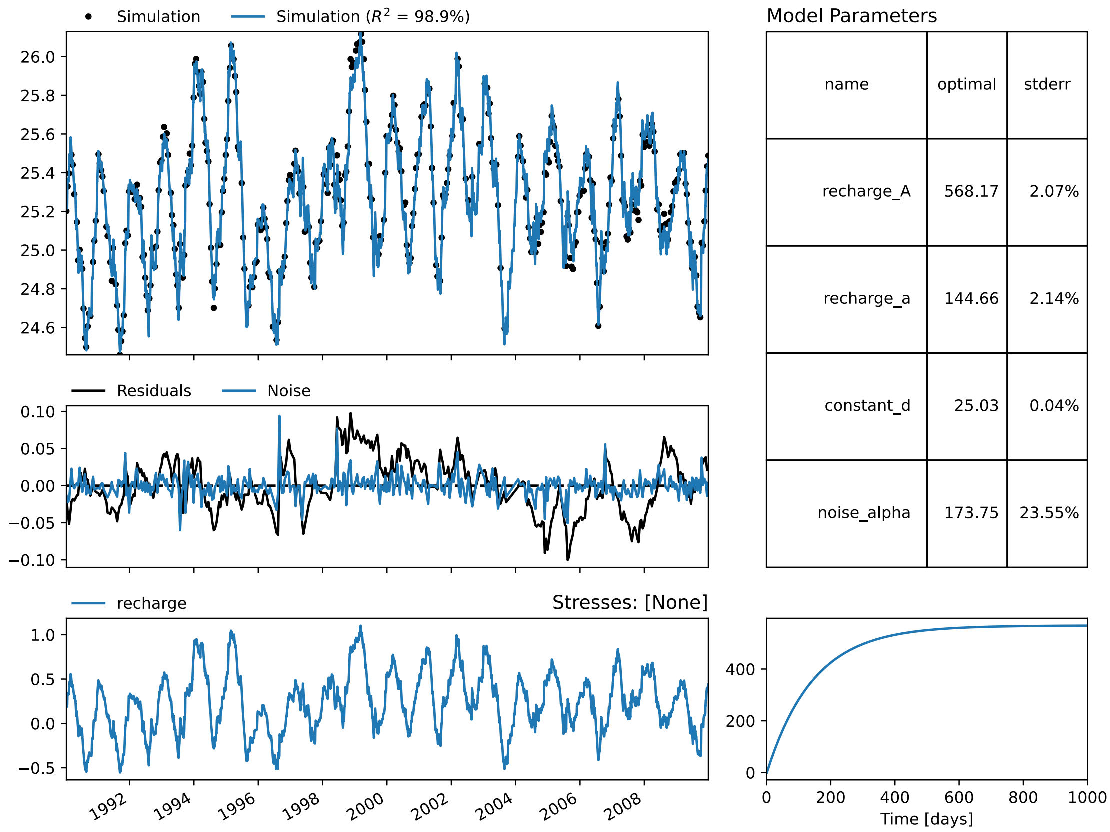
```
number of times true d within estimated confidence interval: 98
number of times true alpha within estimated confidence interval: 87
```

In bovenstaand exeriment is uitgelegd wat het voordeel van het gebruik van een ruismodel is. Als er geen ruismodel gebruikt wordt, dan worden de parameters goed geschat, maar de betrouwbaarheidsintervallen van de parameters zijn te klein. Als het ruismodel wel meegenomen wordt, dan wordt het betrouwbaarheidsinterval veel beter geschat. In bovenstaand voorbeeld was de fout in de stijghoogte gegenereerd met een model dat precies hetzelfde is als het ruismodel dat in het tijdreeksmodel gebruikt werd. Zoals uitgeleged is de correlatie in de residuën van in de werkelijkheid het gevolg van fysische processen, bijvoorbeeld een verschil in de gemeten regenval bij het weerstation en de gemeten regenval bij de peilbuis. In de volgende test wordt het experiment herhaald, maar deze keer wordt er geen gecorreleerde fout opgeteld bij de stijghoogte, maar wordt er een grondwateraanvullingsreeks (regen min verdamping) gebruikt die een fout bevat. Er zijn vele manieren waarop er een realistische fout toegevoegd kan worden aan regen of verdampings reeksen. Hier is er voor gekozen om de gemeten grondwateraanvulling te vermenigvuldigen met een factor $(1+\epsilon_g)$, waarbij ε_g normaal verdeeld is met een gemiddelde nul en standaardafwijking σ_g. Zodoende is de grondwateraanvulling elke dat dus iets meer of iets minder dan gemeten. De waarde van σ_g= is 0.2.

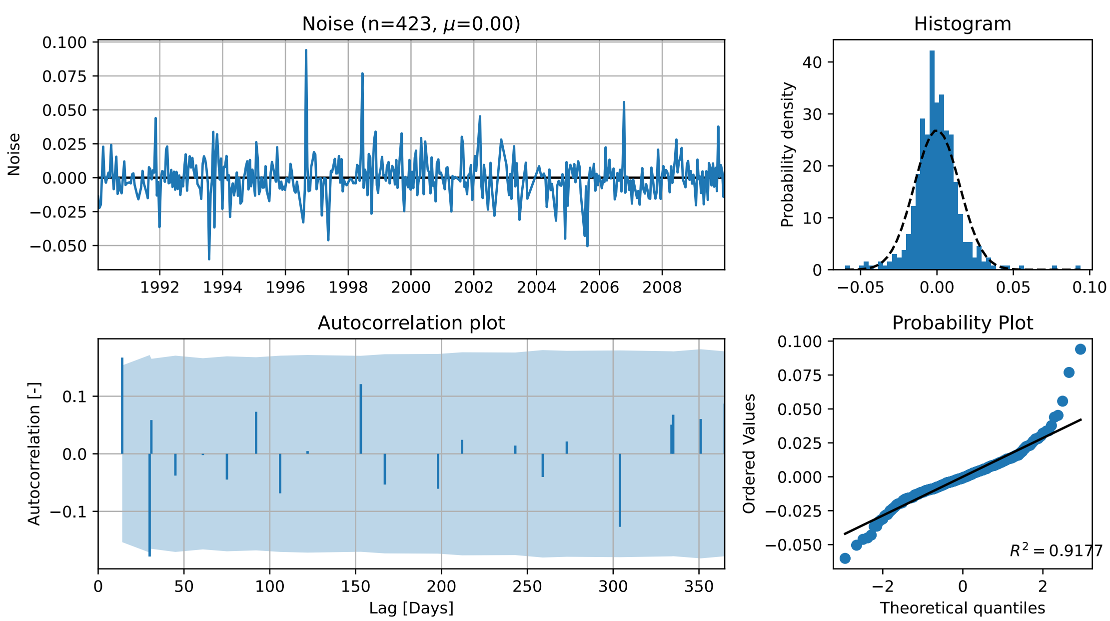
Hieronder wordt een tijdreeksmodel gemaakt, waarbij wederom een AR1 ruismodel meegefit wordt. Zoals te zien is, is de fit zeer goed en is er vrijwel geen autocorrelatie in de ruis, maar de ruis is niet erg normaal verdeeld. Dit baart zorgen, want in het bepalen van het betrouwbaarheidsintervallen van de parameters wordt er vanuit gegaan de de ruis normaal verdeeld is.

```
/Users/mark/git/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")
```

Results for time series model with error in recharge



Diagnostics for time series model with error in recharge



We gaan wederom het hierboven beschreven experiment uitvoeren. Wederom blijken de tijdreeksmodellen zonder noise model te parameters goed te schatten, maar de betrouwbaarheidsintervallen van de parameters zijn veel te klein: bij slechts 30-40% van de modellen liggen de echte parameters binnen de betrouwbaarheidsintervallen van de geschatte parameters. Ook de tijdreeksmodellen met noise model geven goede resultaten voor de geschatte parameters. Bij 80-85% van de modellen liggen de echte parameters binnen de betrouwbaarheidsintervallen van de geschatte parameters. Dat is minder dan de verwachtte 95% en komt hoogst waarschinlijk omdat de ruis niet normaal verdeeld is. Om dit beter te krijgen zou een tijdreeksmodel gebruikt moeten worden gebaseerd op, bijvoorbeeld, een maximum likelihood schatting van de parameters, waarbij opgegeven kan worden wat de verwachtte verdeling is van de ruis.

```
0, 20, 40, 60, 80,
```

```
0, 20, 40, 60, 80,
```

```
Results of models with error in recharge, without a noise model
Mean value of parameter A: 591.7426782992483
Mean value of parameter a: 149.23428129938452
Mean value of parameter d: 25.004585739875516
Mean value of evp: 99.17619220048779
```

```
number of times true A within estimated confidence interval: 36
number of times true a within estimated confidence interval: 39
number of times true d within estimated confidence interval: 32
```

```
Results of models with error in recharge, with a noise model
Mean value of parameter A: 593.895547885876
Mean value of parameter a: 150.5514298698745
Mean value of parameter d: 25.003663104071393
Mean value of parameter alpha: 143.01175591702426
Mean value of evp: 99.16060120631344
```

```
number of times true A within estimated confidence interval: 84
number of times true a within estimated confidence interval: 83
number of times true d within estimated confidence interval: 83
```

0.3.1 Voorbeeld van een echte reeks waarbij AR1 niet zo goed werkt, maar ARMA(1,1) wel

0.3.2 Voorbeeld van een reeks waarbij AR1 niet goed werkt bij hoge frequentie, maar wel redelijk bij lagere frequentie