

De homogenisatie van de maximumtemperaturen in De Bilt van 1906-1950

door Frans Dijkstra

Voorgeschiedenis

In 2016 heeft het KNMI de dagelijkse temperaturen (maximum, minimum, gemiddelde) van de vijf hoofdstations (De Bilt, Eelde/Groningen, Maastricht/Beek, Vlissingen, De Kooy/Den Helder) gehomogeniseerd, aanvankelijk zonder er veel ruchtbaarheid aan te geven. Mede dankzij EW-journalist [Simon Rozendaal](#) werd bekend, dat door deze operatie de temperatuur op vrijwel alle hete dagen in De Bilt vóór 1950 drastisch naar beneden is bijgesteld. Het jaar 1947, dat tot dan toe bekend stond als het jaar met de meeste tropische dagen en hittegolven duikelde vijf of zes plaatsen in de ranglijst van topzomers. De reden die het KNMI opgaf voor de homogenisatie was dat alle stations verplaatst waren en in De Bilt de meethut veranderd was. Tot 1950 werd in De Bilt gemeten in een grote pagodehut, waarbij de meetinstrumenten wel zijn afgeschermd van de zon, maar waarbij luchtstromen en reflecties van de zonnestrallen een andere werking kunnen hebben dan bij een moderne Stevensonhut. Bij de vier secundaire stations waren parallelle metingen beschikbaar van de oude en de nieuwe situatie, waardoor een eenvoudige vergelijking mogelijk was om de reeksen aan elkaar te knopen. Voor De Bilt waren zulke parallelle metingen wel beschikbaar, maar omdat de nieuwe thermometerhut in 1950 ook een paar honderd meter was verplaatst, zijn die metingen niet gebruikt voor de homogenisatie. In plaats daarvan is het 150 km verder gelegen station Eelde als vergelijkingsstation gebruikt.

Het effect van de homogenisatie

Direct na het lezen van het artikel van Simon Rozendaal heb ik het allemaal nagerekend. Dat kon ik doen omdat ik vóór de homogenisatie de ongehomogeniseerde data al eens had gedownload. Aanvankelijk hield het KNMI tot 2019 de oorspronkelijke data geheim, evenals de wel aanwezige parallelmetingen in 1947-1950.

Het viel me al snel op dat de correcties voor de maximumtemperaturen in De Bilt erg groot waren, in vergelijking met de andere stations. Alle temperaturen boven 27 graden in de zomermaanden waren met 1,6 tot 1,9 graad verlaagd. Het aantal tropische dagen in de periode 1906-1950 daalde daardoor in De Bilt veel sterker dan op de andere stations, van 160 naar 76. Was dat misschien met vooropgezette bedoeling gedaan? De Bilt is namelijk maatgevend voor het optreden van een hittegolf. Als er minder hittegolven vóór 1950 worden geteld, heeft het KNMI minder uit te leggen over zomerhitte vóór de moderne klimaatverandering. Het [technisch rapport](#) dat het KNMI uitgaf over de homogenisatie en het artikel van [Brandsma in Meteorologica \(2016\)](#) (2016) wekken wel de indruk dat het daarom ging. Onderzoeker Brandsma constateerde in de conclusies triomfantelijk, dat het aantal hittegolven vóór 1950 was teruggebracht van 23 naar 7. Opdracht uitgevoerd! Het hoeft niet te verbazen dat er over deze operatie nooit een gereviewde publicatie is verschenen in een internationaal tijdschrift.

Tegenspraak van vakgenoten bleef uit

In de blogsfeer werd de homogenisatie kritisch ontvangen, maar enige tegenspraak in de vakbladen en wetenschappelijke literatuur bleef nog uit. Mijn conclusie dat de maximumtemperaturen in De Bilt te sterk zijn gecorrigeerd in vergelijking met andere stations heb ik begin 2017 neergelegd in een [artikel](#) voor het vakblad Meteorologica. Over de methode van het KNMI sprak ik me niet uit, ik constateerde slechts dat de uitkomst niet klopte.

'Gezien het belang van het onderwerp' liet de hoofdredacteur van dat blad het artikel beoordelen door 5 reviewers. Dat is een absoluut record in mijn eigen publicatiegeschiedenis sinds 1972. Het feit dat De Bilt na homogenisatie vóór 1950 nog maar 26,4% telde van de tropische dagen in de periode 1906-2016, terwijl dit voor de andere stations 35,4-39,5% was, gold voor mij als de 'smoking gun', het bewijs dat de homogenisatie niet klopte. Enige reviewers beaamden dit direct, maar het heeft veel moeite gekost om een paar tegenstribbelende reviewers te overtuigen. Na de derde versie van het manuscript hakte de hoofdredacteur de knoop door en het artikel verscheen in december 2017,

met een door mij gevraagd weerwoord van het KNMI waarvan het belangrijkste verweer was dat *'het een illusie is te denken, dat een homogenisatie perfect uitwerkt op de allerhoogste temperaturen'*. Waarom het KNMI eerder zo blij was dat het aantal hittegolven met deze homogenisatie sterk was teruggebracht liet men wijselijk in het midden.

Amateurs uit de provincie zoeken het uit

Samen met Rob de Vos, Marcel Crok en Jan Ruis heb ik geprobeerd de homogenisatie van het KNMI te reproduceren. In een rapport onder de titel *'Het raadsel van de verdwenen hittegolven'* constateerden wij met Excel-sheets dat de homogenisatie van het KNMI niet reproduceerbaar was, en dat een aangepaste homogenisatie meer tropische dagen en hittegolven liet zien. Navraag bij het KNMI leverde op, dat op een aantal punten was afgeweken van de in het technisch rapport beschreven methodiek. *'Het was beter geweest, dat deze punten in het technisch rapport waren vermeld'* gaf onderzoeker Brandsma toe.

Toen wij tenslotte de beschikking kregen over de R-code waarmee het KNMI had gerekend, bleek dat vrijwel elke keuze die het KNMI had gemaakt er toe leidde dat er minder tropische dagen in De Bilt werden gevonden¹. Door variaties aan te brengen op die keuzen vonden wij, dat je alleen nog minder tropische dagen in De Bilt kunt vinden door Vlissingen te nemen als vergelijkingsstation in plaats van Eelde. Neem je de mediaan van alle uitkomsten, dan zijn er in de periode 1906-1950 zeker 40 tropische dagen te weinig geteld.

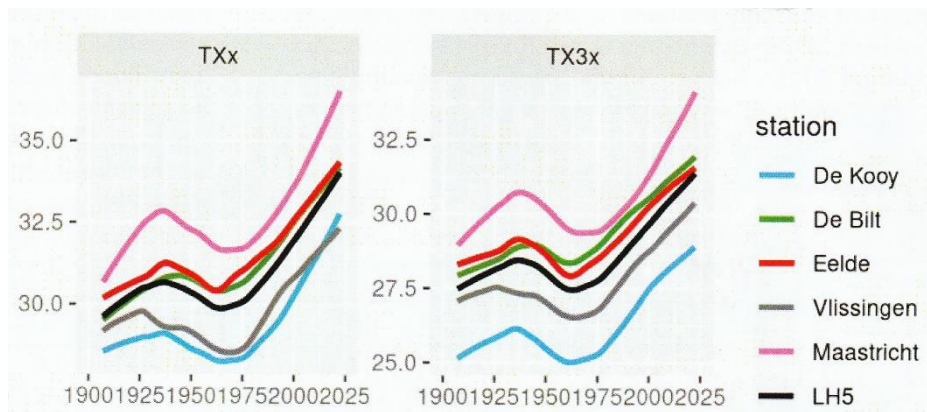
Zoals boven vermeld heeft het KNMI zijn homogenisatie nooit verantwoord in een wetenschappelijk artikel. Daarom hebben wij als amateurs uit de provincie deze taak op ons genomen door onze bovenstaande bevindingen neer te leggen in een wetenschappelijk artikel dat we in de zomer van 2020 indienden bij het tijdschrift Theoretical and Applied Climatology (afgekort als TAAC). De review heeft bijna anderhalf jaar geduurd, maar leidde niet tot veranderingen van onze conclusies. In december 2021 is het [artikel online](#) verschenen. Het KNMI, reageerde beleefd. *'Wij verwelkomen onderzoek naar homogenisatietechnieken en zullen in 2023 met een verbeterde homogenisatie komen waarin ook de onzekerheden in de correcties worden aangegeven.'*

Het KNMI weer aan zet

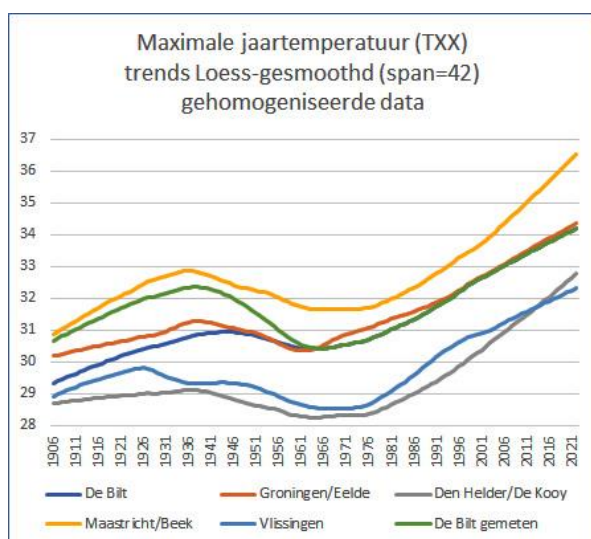
In juni 2023 verscheen in Meteorologica een [artikel van KNMI-onderzoeker Theo Brandsma](#) over trends in temperatuurextremen in Nederland, waaruit valt af te leiden op welke manier het KNMI probeert de homogenisatie van 2016 zoveel mogelijk overeind te houden. Tropische dagen en hittegolven zijn voor het KNMI niet meer zo relevant, omdat die afhangen van arbitraire definities. Liever kijkt het KNMI nu naar de ontwikkeling van de hoogste jaartemperaturen en het hoogste voortschrijdende gemiddelde over 3 dagen. Die reeksen worden afgevlakt met een door het KNMI ontwikkelde speciale smoother (waarvan de meerwaarde ten opzichte van algemeen gebruikte standaardmethoden mij niet duidelijk is) en daaruit wordt een lineaire trend over de hele periode van 1906-2022 berekend uit het verschil van de gesmoothde waarden van 1906 en van 2022. Dat levert de volgende figuur op, waaruit volgens Brandsma blijkt dat de trends in alle gehomogeniseerde reeksen vergelijkbaar zijn.

¹ Deze (in onze ogen) arbitraire keuzen betreffen:

- De keuze voor één vergelijkingsstation, namelijk Eelde
- De lengte van de vergelijkingsperiode, namelijk 56 maanden, wat niet een geheel aantal jaren is, waardoor van het laatste jaar de maanden januari-augustus oververtegenwoordigd zijn
- Het gebruik van een tweedegraads smoothing functie
- Het bepalen van de temperatuurverdeling per maand over die maand en de voorafgaande en de volgende maand. Voor de maand mei wordt daardoor het hoogste percentiel gedomineerd door dagen in juni, voor juni door dagen in juli.



Figuur 1 Uit Fig. 2 in Brandsma (2023). De ontwikkeling van de jaarlijkse maximumtemperatuur (TXx) en de hoogste gemiddelde maximumtemperatuur gedurende drie opeenvolgende dagen (TX3x) voor de 5 KNMI-stations. LH5 is het gemiddelde van de vijf stations.



Figuur 2 Reproductie van Fig 2 in Brandsma (2023)

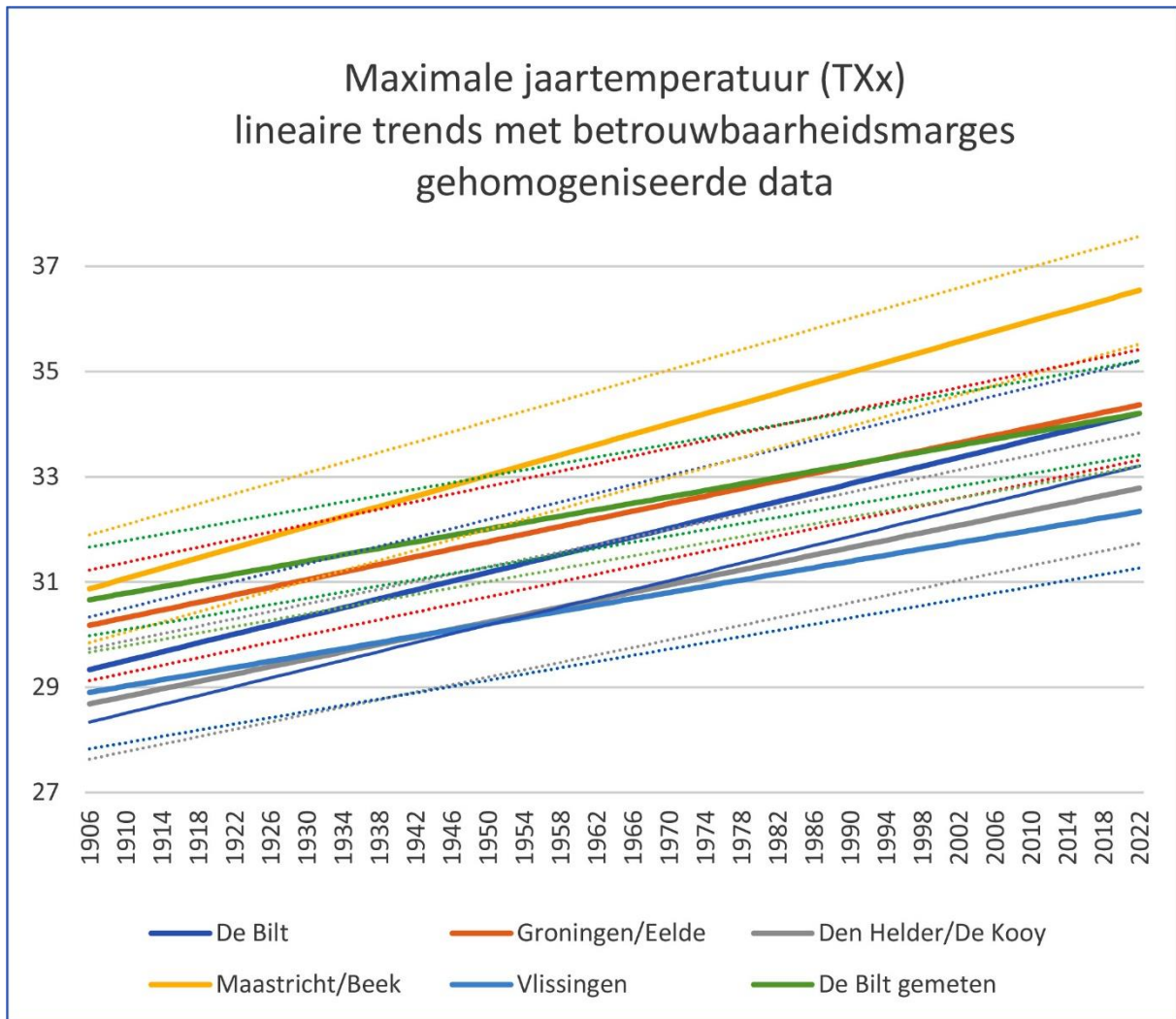
Brandsma suggereert, dat de trend in de gehomogeniseerde data van De Bilt niet afwijkt van die van de andere stations. Als we nauwkeurig kijken, zien we wel degelijk verschillen, die te maken hebben met de correcties die voor de Bilt in de periode 1906-1950 zijn aangebracht: bij TXx vallen de data voor De Bilt en voor Eelde na 1950 vrijwel samen. Vóór 1950 ligt De Bilt onder Eelde. Voor TX3x geldt dat De Bilt na 1950 boven Eelde ligt, en voor 1950 onder Eelde. We zien de verschillen nauwkeuriger in de door mij gemaakte figuur 2. Hierbij zijn dezelfde gehomogeniseerde stationsdata gebruikt als Brandsma doet. Ik heb de normale standaard smoothing (Loess) gebruikt met een span van 42, die volgens Brandsma overeenkomt met zijn methode. Bovendien heb ik de gemeten data van De Bilt toegevoegd. Duidelijk is te zien, dat de gehomogeniseerde data van De Bilt afwijken van die van Eelde. Ook is duidelijk dat de gemeten data van De Bilt inderdaad te hoog lijken te zijn. De groene lijn kruipt vóór 1940 zelfs dicht tegen die van Maastricht aan.

De hamvraag is nu, of deze verschillen statistisch significant zijn. Daarvoor moeten we kijken naar de niet afgevlakte data van de afzonderlijke jaren. Zie Tabel 1.

Tabel 1 TXx-De Bilt minus TXx-Eelde. Gehomogeniseerde reeksen.			
Periode	Gemiddeld verschil	Standaarddeviatie	Aantal
1906-1950	-0,54	1,33	45
1951-2022	-0,06	1,42	72

De gehomogeniseerde TXx-data voor 1906-1950 geven voor De Bilt in vergelijking met 1951-2022 gemiddeld een 0,48 graden lagere TXx dan voor Eelde, bij een standaarddeviaties van ongeveer 1,4. Dat lijkt een klein verschil en een grote spreiding, maar omdat het om 127 punten gaat is dit verschil wel significant. Student's t-toets geeft $t=1,80$. Bij 125 vrijheidsgraden (n_1+n_2-2) is de significantie 96%. In termen van het IPCC: het is uiterst waarschijnlijk dat de gehomogeniseerde TXx voor De Bilt van 1906-1950 afwijkt van het vergelijkingsstation Eelde/Groningen in die periode.

Maar Brandsma baseert zijn conclusies niet op Figuur 2. Hij vereenvoudigt de figuur door alleen de lineaire trends van 1906-2022 te berekenen. Zie figuur 3.



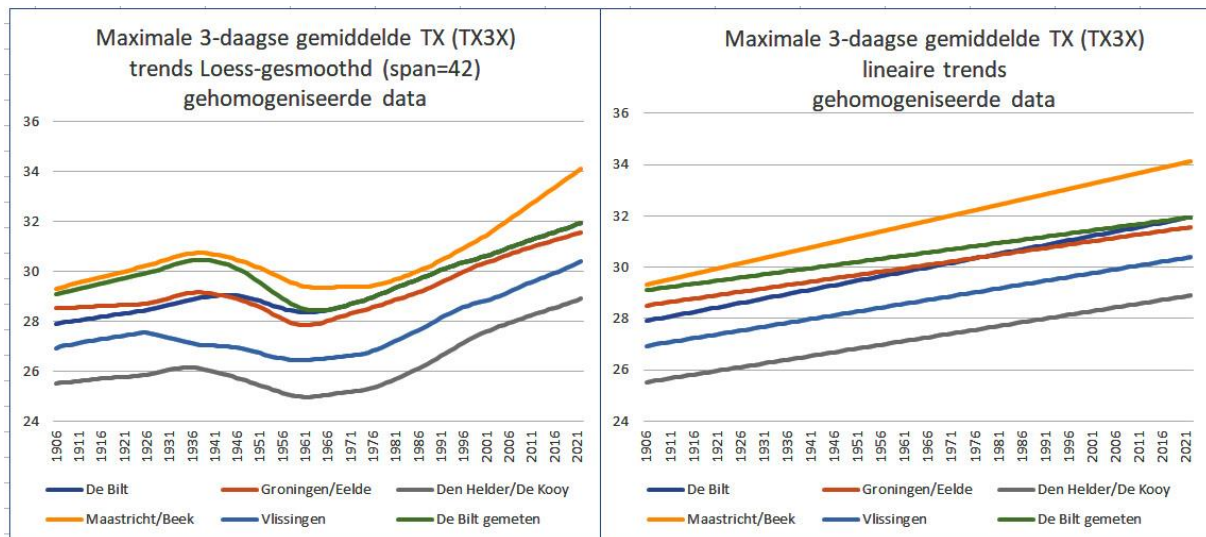
Figuur 3 Dezelfde begin- en einddata als in figuur 2, maar voorgesteld als lineaire trends. De gestippelde lijnen geven de 95%-betrouwbaarheidsmarges van de trendlijnen aan volgens Brandsma

Deze figuur komt niet uit het artikel van Brandsma, maar geeft precies aan wat hij doet. Alle nuances in figuur 2 zijn hier gereduceerd tot rechte lijnen. Uit Brandsma's tabel 1 blijkt, dat hij voor de trend van deze lijnen onzekerheidsmarges van 4,0 tot 4,3 graden aanhoudt, wat betekent dat de marges van de trendlijnen ongeveer een graad boven en onder de trendlijnen liggen. Die marges zijn in figuur 3 met stippellijnen aangegeven. Bij zulke marges zijn er weinig significante verschillen tussen de stations. Zo komt Brandsma tot zijn hoofdconclusie: *“de gegevens van de andere vier hoofdstations – die onafhankelijk van De Bilt gehomogeniseerd zijn – laten voor de hier gepresenteerde indices hetzelfde beeld zien als voor de gegevens van De Bilt (...) De aannames die bij*

de homogenisatie van De Bilt zijn toegepast hebben dus niet tot afwijkende resultaten geleid wat betreft de verschillende temperatuurtrends.”

Saillant is, dat ook de trend in de niet gehomogeniseerde data niet significant afwijkt van de gehomogeniseerde stations. Maar dat vermeldt Brandsma niet. **Ik concludeer daarom dat we hier een schoolvoorbeeld zien van ‘liegen met statistiek’, waard om op te nemen in de handboeken.**

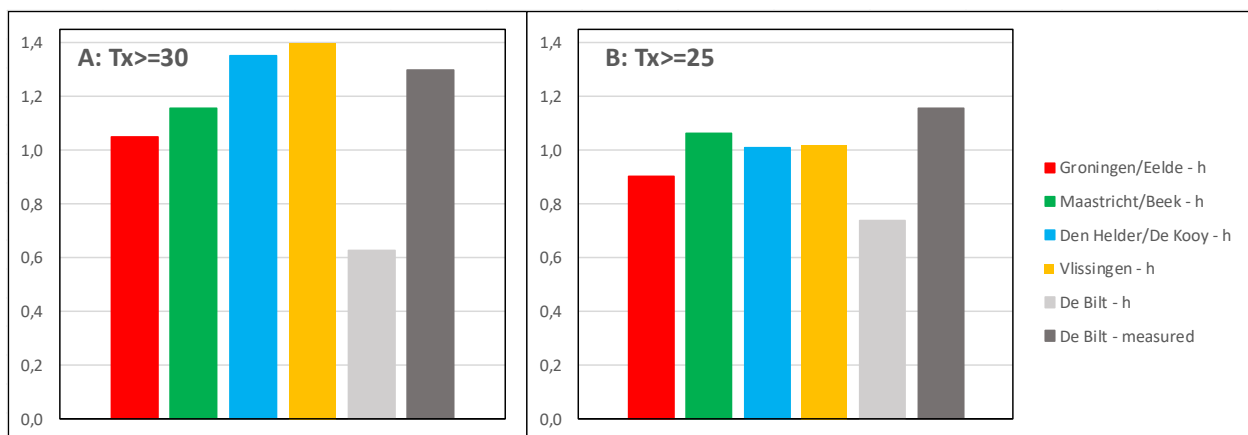
Over de andere hitte-indicator die Brandsma gebruikt kan ik kort zijn. TX3x geeft de hoogste waarde per jaar van het voortschrijdend gemiddelde van drie opeenvolgende maximale dagtemperaturen. Dat is een betere indicator voor zomerhitte dan het aantal hittegolven. Zie figuur 4.



Figuur 4 Trends in TX3x volgens de methode van Brandsma, gesmoothde data (links) en de lineaire trends (rechts)

Ook bij deze indicator blijken de gehomogeniseerde data van De Bilt vóór 1950 onder die van Groningen/Eelde te liggen. Na 1950 liggen ze boven die van Eelde. Het verschil tussen de perioden 1906-1950 en 1951-2022 is 0,73 graden en de significantie van het verschil is meer dan 99,5%.

De smoking gun



Figuur 5 Uit Dijkstra et al (2021): Ratio of days with $T_x \geq 30^\circ\text{C}$ (panel A) and $T_x \geq 25^\circ\text{C}$ (panel B) in 1906-1949/1952-1995 at the 5 main stations in the Netherlands. Homogenized data of all 5 stations. For De Bilt also the measured data.

Ik ga nog even terug naar de ‘smoking gun’, die volgens mij bewijst dat de homogenisatie van De Bilt niet klopt. In mijn artikel van december 2017 liet ik door vergelijking van de periodes voor en na 1950 voor de 5 KNMI-stations zien dat in De Bilt het aantal tropische dagen vóór 1950 onevenredig is teruggebracht. In ons TAAC-artikel van december 2021 hebben we dit opnieuw in beeld gebracht, nu met wat andere periodes, maar met dezelfde conclusie. Zie figuur 5. Met het aantal tropische dagen wijkt De Bilt significant af van de andere gehomogeniseerde stations, terwijl de gemeten data van De Bilt niet afwijken van de andere stations. Bij het aantal zomerse dagen ($T_x \geq 25$) is er ook een

verschil, maar dat is niet significant. Deze onevenredige correctie van het aantal tropische dagen in De Bilt vóór 1950 is door vriend en vijand erkend, bijvoorbeeld op het mainstream blog *klimaatveranda.nl*. Maar in een e-mail-correspondentie met Brandsma noemt hij dit verschil niet significant. In hoeverre heeft hij wat dit betreft gelijk?

Voor het onderkennen van uitschieters (outlier detection) bestaan statistische toetsen. In het boekje *'Practical statistics for chemical research'* van J.D. Hinchin (1969), dat vanaf het begin van mijn loopbaan mijn trouwe gids is geweest in statistische kwesties, staat onder de titel *'The best two out of three'* (pagina 26) een handige manier om uitschieters op puur statistische gronden als fout aan te merken. De formule is:

$$R = (x_n - x_{n-1}) / (x_n - x_1)$$

waarin x_n de verdachte waarde is, x_{n-1} de waarde die daar het dichtst bij ligt, en x_1 de waarde die er het verst van af ligt. Simpel voorbeeld: als je een reeks hebt van 1,6,7,8,9 waarbij je vermoedt dat de 1 een uitschieter is, dan bereken je $R=5/8=0,625$. Er is een tabel (zie bijlage) die aangeeft dat de significantie in dit geval ligt bij ongeveer 94%. Toegepast op de data van figuur 5A (tropische dagen) bereken je $R=0,583$, wat een significantie betekent van 92%. In IPCC-termen is dat zeer waarschijnlijk, maar de kans dat De Bilt geen uitschieter is, is dus niet nul maar 8%. Er is dus een kleine kans van 8%, dat de homogenisatie toch klopt.

Waarom doet het KNMI dit?

In het slot van zijn artikel refereert Brandsma aan de publicatie van Dijkstra et al. (2021) die laat zien *“dat de resultaten van de homogenisatie gevoelig zijn voor een aantal aannames in Brandsma (2016). De gegevens van de andere vier hoofdstations – die onafhankelijk van De Bilt gehomogeniseerd zijn – laten echter voor de hier gepresenteerde indices hetzelfde zien als voor de gegevens van De Bilt. Dat geldt ook voor andere temperatuurtrends in Nederland. De aannames die bij de homogenisatie van De Bilt zijn toegepast hebben dus niet tot afwijkende resultaten geleid wat betreft de verschillende temperatuurtrends. Vanwege de consistentie in temperatuurtrends van de onafhankelijk van elkaar gehomogeniseerde reeksen, verwachten we dat de door het KNMI aangekondigde versie 2.0 homogenisatie van deze reeksen geen grote veranderingen zal laten zien in temperatuurtrends.”*

Hierin kunnen wij dus lezen, dat het KNMI de hakken in het zand zet. Met hoog overvliegende trends over lange perioden en grote onzekerheidsmarges in die trends probeert Brandsma aan te tonen, dat het allemaal niets uitmaakt. Op zichzelf kan dat wel kloppen. De onzekerheidsmarges in de van dag tot dag gehomogeniseerde temperaturen zijn waarschijnlijk groter dan de correcties die zijn aangebracht. Dan blijft de grote vraag staan: waarom zijn die correcties dan toch aangebracht? En waarom zal het KNMI – gezien dit statement van Brandsma – daar geen rectificaties op aanbrengen?

Om te laten zien dat het klimaat in Nederland opwarmt heeft het KNMI het niet nodig om de historische gegevens te vervalsen. Dat we de laatste 20 jaar vaker hete dagen en hittegolven hebben meegemaakt dan in de vorige eeuw wordt door niemand betwist. Dat we tegenwoordig minder kunnen schaatsen op open water dan 25 jaar geleden weet ook iedere schaatser (overigens waren de jaren 2009-2013 goede schaatsjaren, met in 2012 bijna een Elfstedentocht en in 2021 nog een spectaculair staartje). Maar waarom moet ontkend worden dat er ook van 1930-1950 veel zomerwarmte was in Nederland? Weliswaar niet zoveel als de laatste 20 jaar, maar wel veel meer dan tussen 1950 en 1975? Waarom moet de ontwikkeling van het klimaat voorgesteld worden met rechte lijnen? Is dat om de aandacht af te leiden van de natuurlijke interne variabiliteit van het klimaat?

Conclusies

1. Brandsma claimt ten onrechte dat de arbitraire keuzen die het KNMI heeft gemaakt bij de homogenisatie van De Bilt niet tot afwijkende resultaten hebben geleid wat betreft de verschillende temperatuurtrends.
2. Een nauwkeurige analyse laat zien dat de gehomogeniseerde TXx en TX3x van De Bilt voor de periode 1906-1950 significant lagere waarden (0,48 resp. 0,73 graden) geven ten opzichte van het vergelijkingsstation Eelde. Deze afwijking is voor TXx 96% betrouwbaar en voor TX3x meer dan 99,5%. Aangezien TX3x een indicator is voor het mogelijk optreden van hittegolven volgt hieruit, dat de homogenisatie vrijwel zeker teveel hittegolven heeft geschrapt. Het *'Raadsel van de verdwenen hittegolven'* is hiermee opgelost.
3. Brandsma komt tot zijn misleidende claim door uitsluitend lineaire trends te presenteren over een lange periode waardoor de verschillen tussen de stations klein lijken en de onzekerheidsmarges groot. Deze methode is niet geschikt om te beoordelen of de homogenisatie van De Bilt tot afwijkende resultaten heeft geleid, en verdient vermelding in de handboeken over *'liegen met statistiek'*.
4. De methode van Brandsma suggereert een lineair model en leidt de aandacht af van de grote interne variabiliteit van het klimaat.
5. Volgens de methode van Brandsma wijkt ook de temperatuurtrend van de niet gehomogeniseerde data van De Bilt niet af van de andere stations. Dit roept de vraag op of de homogenisatie nodig was.

Bijlage. Outlier detection.

Ontleend aan J.D. Hinchon (1969) 'Practical statistics for chemical research'

- x_n = the suspected value
- x_{n-1} = the nearest to it
- x_{n-2} = the nearest, once removed
- x_1 = the furthest away from the suspected value
- x_2 = the next furthest away
- etc.

Critical values and criteria for testing for extreme values

α	0.30	0.20	0.10	0.05	0.02	0.01	0.005	Criterion
100(1 - α)	70	80	90	95	98	99	99.5	
$n = 3$	0.684	0.781	0.886	0.941	0.976	0.988	0.994	$r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}$
4	0.471	0.560	0.679	0.765	0.846	0.880	0.926	
5	0.373	0.451	0.557	0.642	0.729	0.780	0.821	
6	0.318	0.386	0.482	0.560	0.644	0.698	0.740	
7	0.281	0.344	0.434	0.507	0.586	0.637	0.680	
8	0.318	0.385	0.479	0.554	0.631	0.683	0.725	$r_{11} = \frac{x_n - x_{n-1}}{x_n - x_2}$
9	0.288	0.352	0.441	0.512	0.587	0.635	0.677	
10	0.265	0.325	0.409	0.477	0.551	0.597	0.639	
11	0.391	0.442	0.517	0.576	0.638	0.679	0.713	$r_{21} = \frac{x_n - x_{n-2}}{x_n - x_2}$
12	0.370	0.419	0.490	0.546	0.605	0.642	0.675	
13	0.351	0.399	0.467	0.521	0.578	0.615	0.649	
14	0.370	0.421	0.492	0.546	0.602	0.641	0.674	$r_{22} = \frac{x_n - x_{n-2}}{x_n - x_3}$
15	0.353	0.402	0.472	0.525	0.579	0.616	0.647	
16	0.338	0.386	0.454	0.507	0.559	0.595	0.624	
17	0.325	0.373	0.438	0.490	0.542	0.577	0.605	
18	0.314	0.361	0.424	0.475	0.527	0.561	0.589	
19	0.304	0.350	0.412	0.462	0.514	0.547	0.575	
20	0.295	0.340	0.401	0.450	0.502	0.535	0.562	
21	0.287	0.331	0.391	0.440	0.491	0.524	0.551	
22	0.280	0.323	0.382	0.430	0.481	0.514	0.541	
23	0.274	0.316	0.374	0.421	0.472	0.505	0.532	
24	0.268	0.310	0.367	0.413	0.464	0.497	0.524	
25	0.262	0.304	0.360	0.406	0.457	0.489	0.516	