



Reassessment of the homogenization of daily maximum temperatures in the Netherlands since 1901

Frans Dijkstra¹ · Rob de Vos² · Jan Ruis³ · Marcel Crok⁴

Received: 1 July 2020 / Accepted: 26 November 2021
© The Author(s) 2021

Abstract

In 2016, the Royal Dutch Meteorological Office (KNMI) homogenized the daily temperature records for the Netherlands from 1901 to 1950 to allow a realistic comparison of the temperatures from 1901 to the present. The homogenizations for the main station De Bilt were carried out using a Percentile Matching Method (PMM) with one reference station and a 56-month reference period. In this study, it is shown that the corrections in the number of tropical days (maximum temperature ≥ 30 °C) depend strongly on the choice of the reference station and the length of the reference period. A total of 116 different variants of the homogenization of De Bilt were carried out, using all combinations of five reference stations, five reference periods, two ways to calculate percentiles, and two ways to smooth the data. The parameters used for the KNMI's current homogenization of De Bilt result in a very sharp decrease of tropical days, which is not replicated by the majority of the 116 variants. Moreover, after homogenization, De Bilt appears to be an outlier compared to the other meteorological stations. Therefore, the current homogenized estimates of tropical days for De Bilt should be treated with considerable caution.

1 Introduction

In the Netherlands, there are five long-term active principal meteorological stations with temperature records from the beginning of the twentieth century up to present. These stations are Den Helder/De Kooy, De Bilt, Groningen/Eelde, Vlissingen/Souburg, and Maastricht/Beek (Fig. 1). All five stations have long time series with daily mean, minimum and maximum temperatures (T_m , T_n , T_x): De Bilt from 1901 onward and the other four stations from 1906.

Relocation took place at all five stations. Groningen station was moved to airport Eelde in 1951, about 10 km south of the city. Den Helder station was moved to airport De Kooy in 1972, about 5 km to the southeast; Maastricht station was moved to airport Beek in 1951, about 9 km northeast of Maastricht; Vlissingen station was temporarily located at Souburg (2 km further north) from 1947 to 1958. During these relocations, parallel measurements were carried out at the old and the new locations. For Groningen-Eelde and Maastricht-Beek, parallel data are available for a period of 6–7 years. A sixth station, Winterswijk, was also used in this study. It has a long-term station record starting in the beginning of the twentieth century, but was discontinued in 1970 (the station record spans the period 1901–1970). None of these stations remained unchanged regarding location and/or screen type during the whole period. See Fig. 1 and Table 1 for details.

For De Bilt, the situation was more complicated, because of a screen change and two relocations. In May 1950, the large Pagoda screen was replaced by a Stevenson screen. Parallel measurements were made from March 1947 to August 1950. In September 1950, the Stevenson screen was moved to a temporary location about 80 m to the west. In August 1951, the Stevenson screen was moved to a more open location some 300 m to the south. For both relocations in 1950 and 1951, no parallel observations were made.

✉ Frans Dijkstra
f.y.dijkstra@planet.nl

Rob de Vos
rob@klimaatgek.nl

Jan Ruis
j.ruis@casema.nl

Marcel Crok
marcel.crok@clintel.org

¹ Ommen, Netherlands

² Deurne, Netherlands

³ Rijswijk, Netherlands

⁴ Clintel Foundation, Zekeringstraat 41, 1014BV, Amsterdam, Netherlands



Fig. 1 The situation of the meteorological stations in the Netherlands having data starting in 1901 (De Bilt, Winterswijk) or 1906 (other stations)

Data of De Bilt are important, not only because of its long temperature series but also because it is one of the six stations on which the Central Netherlands Temperature record (CNT) is based. The CNT is a monthly daily mean temperature time series that is representative for a region in the center of the Netherlands. It has been constructed to study large-scale temperature changes and facilitate comparisons with climate models (Van der Schrier et al. 2011). Of importance is also that in the Netherlands, a national heat wave is defined only on the basis of the data of De Bilt.

2 Homogenization

Long instrumental temperature records are essential for the validation of climate models, for assessing long-term trends, and for detection of regional climate change (Venema et al. 2012). However, a problem with many long-term temperature series is the presence of non-climatological biases. External changes can cause jumps and trend breaks that are non-climatic. Peterson et al. (1998) and Trewin (2010) discussed these possible changes that can introduce non-random systematic biases into a data set.

Step biases are often introduced by station move or change in instrumentation or shelter. Trend biases are often associated with gradual changes in the environment of the station over multiple years, e.g., the growth of vegetation nearby (a micro-climate bias), or urbanization bias (a local climate bias).

It is likely that the stations of Groningen, Den Helder, Vlissingen, De Bilt, and Maastricht were affected by step biases as well as urbanization bias. Groningen (1951), Den Helder (1972), and Maastricht (1951) were all moved to airport locations in the neighborhood; Vlissingen (1958) regained its previous location at the harbor. In De Bilt, the screen moved twice, in 1950 and 1951. The old location was in a park with high trees, close to the buildings of the KNMI. The temporary new location in 1950 was in the same park but at greater distance from the buildings. The final location was situated in an open field on the KNMI grounds. Although De Bilt is not situated in a city, the distance to Utrecht is small and a continuing urban bias has been estimated. Brandsma et al. (2003) found for De Bilt station during the twentieth century an urbanization bias of $0.10 \text{ K} \pm 0.06 \text{ K}$. Koopmans et al. (2012) calculated for De Bilt an urban heat effect of 0.32 K in the previous century.

It is crucial that a data set reflects climate changes and not changes in observation conditions. Therefore, homogenization of temperature data sets is generally needed for construction of long-term datasets to be used for monitoring climate change. Applying homogenization to a data set typically starts with detection of inhomogeneities which is

Table 1 Meteorological stations in the Netherlands, having records from start of the twentieth century

Station	Start date	Major changes
De Bilt	1–1-1901	May 1950: screen change (Pagode—> Stevenson), August 1950: relocation 85 m West, August 1951: relocation 300 m South to an open location
Groningen/Eelde	1–3-1906	1951: city station moved to airport Eelde, 10 km South
Maastricht/Beek	1–1-1906	1951: city station moved to airport Beek, 8 km North, 90 m higher
Den Helder/De Kooy	1–1-1906	September 1944–July 1945: no data, 1972: city station moved to airport De Kooy, 6 km Southeast
Vlissingen/Souburg	1–1-1906	October 1944–May 1945: no data, 1947–1958: city station moved to airport Souburg, 4 km East
Winterswijk	1–1-1901	November 1944: no data, Possible inhomogeneities in 1950. Closed 1971

a well-described problem in climate science. Reviews on detection techniques include those of Reeves et al. (2007), Li and Lund (2012), Lindau and Venema (2015), and Acquavotta et al. (2016).

The next step in the process of homogenization is to remove inhomogeneities by adjusting the data. Homogenization has frequently been applied to data sets of monthly and annual mean temperatures and is described in various studies. In recent decades, a significant number of algorithms have been developed to remove inhomogeneities. Venema et al. (2012) did a blind intercomparison and validation study of monthly homogenization algorithms. The algorithms were validated against a realistic benchmark dataset containing real inhomogeneous data and added random independent inhomogeneities. The relative homogenization algorithms developed to work with an inhomogeneous reference performed best. In relative homogenization, neighboring stations are exposed to nearly the same climate signal. As a consequence, the differences between those stations can be used to detect inhomogeneities.

Studying developments in daily temperatures and extremes requires a homogeneous set of daily maximum temperatures (T_x) and minimum temperatures (T_n), not just mean temperatures (T_m). Homogenization of temperatures at a daily time scale is much more difficult than at monthly or annual scales because of the day-by-day changing meteorological situation. Della-Marta and Wanner (2006) used a nonlinear model to estimate the relationship between a candidate station and a highly correlated reference station. Trewin and Trevitt (1996) compared three homogenization methods for daily and monthly temperature series at two sites during a period of overlapping records and found that the “frequency distribution matching” method was superior in the case of extreme events. Mestre et al. (2011) used a method, SPLIDHOM,

that relies on an indirect non-linear regression method, with reliable estimation being ensured by cubic smoothing splines. Trewin (2013) used the percentile-matching (PMM) method, which applies adjustments to daily data depending on their position in the frequency distribution.

According to Brandsma (2016), all five Dutch principal stations show step biases due to relocation. As previously noted, the De Bilt series was also affected by a screen change in 1950. In order to be able to make realistic comparisons of the data of the first half of the twentieth century with recent data, the Dutch Meteorological Office (KNMI) homogenized the data of all five stations using the percentile matching method (PMM), as described by Brandsma (2016). For Groningen/Eelde, Maastricht/Beek, Den Helder/De Kooy, and Vlissingen/Souburg, parallel measurements were available for an overlapping period of 4–7 years to homogenize the data.

For De Bilt, parallel measurements were only available for the screen change in 1950, but not for the relocations in 1950 and 1951. Therefore, the KNMI used data of Eelde (148 km to the Northwest) before and after the change in De Bilt in 1950/1951, to homogenize the data of De Bilt for the period 1901–1951. The result of the homogenization was a strong decrease of tropical days ($T_x \geq 30$) in De Bilt in the period 1901–1950 from 164 to 76.

Figure 2 shows the ratio of days with $T_x \geq 25$ °C and $T_x \geq 30$ °C for the five principal stations before and after 1950. After homogenization, De Bilt appears to be an outlier compared with the other stations. In this study, we try to understand why the homogenization applied by Brandsma (2016) for De Bilt led to this apparently anomalous result. Specifically, we will examine the PMM homogenization algorithm as applied by Brandsma (2016) and investigate the effect of several choices Brandsma made in the algorithms.

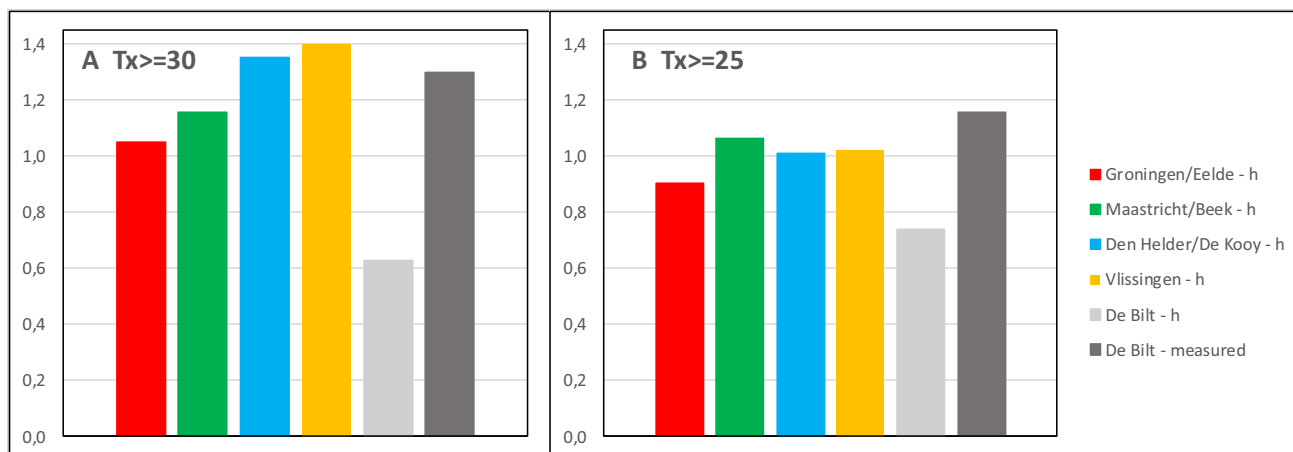


Fig. 2 Ratio of days with $T_x \geq 30$ °C (panel A) and $T_x \geq 25$ °C (panel B) in 1906–1949/1952–1995 at the 5 main stations in the Netherlands. Homogenized data of all 5 stations. For De Bilt also the measured data

The KNMI homogenized daily T_m , T_x , and T_n values at the five stations. We are concerned here with the effects of homogenization on T_x statistics. Since De Bilt is the only station where parallel measurements are not applied and the homogenization results differ notably from the other stations, we only deal with the effect of the homogenization on the T_x measurements for De Bilt. However, we would encourage further research into the effects on T_x of the other stations as well.

3 Methods

The PMM algorithm has different variants that have been described well by various authors, among others Squintu et al. (2020), Trewin (2013), Kuglitsch et al. (2009), Della-Marta and Wanner, (2006), Cao et al. (2016). We focus here on the variants that Brandsma (2016) used for homogenizing daily temperatures:

- (1) The overlap case, when parallel data of two situations at (approximately) the same location are available.
- (2) The non-overlap case, when there are no parallel data, so that data from reference stations must be used.

In the overlap case, for every month, percentiles of T_m , T_x , and T_n are calculated for the two situations. The difference between the monthly percentiles leads to a correction factor that can be applied to the data from the old situation, to make the old data comparable to the new situation. This procedure was followed by Brandsma (2016) for the homogenization of four stations: Maastricht/Beek, Den Helder/De Kooy, Vlissingen/Souburg, and Groningen/Eelde.

For De Bilt, Eelde was used as reference station with two overlapping periods: 1946–1950 and 1952–1956. This is the non-overlap case of the PMM algorithm, where the data for the reference stations must be homogeneous, while the candidate station has an inhomogeneity between the two periods.

The procedure that Brandsma (2016) applied is as follows:

- (1) Calculate for each month the percentiles (5, 10, ..., 90, 95) in the overlapping periods of the daily temperatures before and after the inhomogeneity, both for the candidate station and the reference station.
- (2) Calculate for each month before and after the inhomogeneity the difference (ΔT_1 , ΔT_2) between the percentiles of the candidate station and the reference station.
- (3) Calculate for each month and percentile the correction for the inhomogeneity as $\Delta T = \Delta T_1 - \Delta T_2$. Smooth the columns and rows of the resulting 19×12 matrix. For this smoothing, he used a Loess smoothing with a span of 0.6 and degree = 2, the latter meaning that the local regression is done with a quadratic function.

Trewin (2013) stated that because of the distance between a reference station and the candidate station, it is preferable not to use one reference station, but a set of reference stations located in different directions from the candidate station. For the final corrections, the results of the homogenizations with the separate reference stations can be combined by taking the median.

Della-Marta and Wanner (2006) conclude about using one or more reference stations: "Since the method only uses one reference station to estimate the adjustments, it is important that all inhomogeneities in that series be accurately determined for the creation of the HSPs [homogeneous subperiods]. If more than one suitable reference station is available, a comparison between the adjustments from each could be made. If the results were consistent then we may be more confident that [...] the use of this method [is] justified." Brandsma's method slightly differs from the method described by Della-Marta and Wanner.

Brandsma (2016) described the homogenization of the Dutch stations in a technical report and provided additional details in a personal communication to the authors of this paper. Important additional details are:

- (1) The reference periods were 1–1-1946 to 31–8-1950 and 1–1-1952 to 31–8-1956. Brandsma did so to include as much overlapping data as possible: Eelde started on 1–1-1946 and the Pagoda screen in De Bilt was abandoned on 31–8-1950. This choice results in reference periods of 56 months, with an overrepresentation of calendar months 1–8.
- (2) Because of the occasionally large differences between percentiles of adjacent months, Brandsma calculated the percentiles for each month from running groups of three months: for January DJF, for February JFM, etc.
- (3) Because these choices seem somewhat arbitrary and not underpinned, we decided to explore the effect of other choices:
- (4) Not only Eelde as reference station but also the other three principal stations as well as Winterswijk (five variants),
- (5) Reference periods of 48, 56, 60, 120, and 168 months (five variants),
- (6) Calculating the percentiles for each month from single months and from running groups of three months (two variants),
- (7) Loess smoothing not only with 2nd degree regression but also with linear regression (two variants).

We used an R-code script similar to the one Brandsma (2016) used and executed 116 runs with all combinations of the above-mentioned variants.

Because the homogenized De Bilt record seems to be an outlier regarding the number of tropical days (see Fig. 2),

we suggest that Brandsma's procedure may have led to an overcorrection. We explored this possibility by comparing the average differences between homogenized Tx of De Bilt and the other stations. Tx-corrections per percentile were calculated as $\Delta T = \Delta T_1 - \Delta T_2$, where ΔT_1 = the average Tx De Bilt minus average Tx reference station over 1935–1949, and ΔT_2 = likewise but over 1952–1966. We define a relative “overcorrection” as being when ΔT is positive, and a relative “undercorrection” when ΔT is negative. The relative corrections were calculated for percentiles 1, 2, 3, 4, 5, 10, ... 90, 95, 96, 97, 98, 99 and the results were smoothed with the Loess-function, span = 0.4, degree = 1.

4 Data

We used as much as possible uninterrupted data sets of the Dutch stations. All of them were obtained from the KNMI-website or were provided by Brandsma on our request. The data sets are presented by KNMI as homogenized. Only the data of Winterswijk seem to have an inhomogeneity around 1950. Nevertheless, we used these data because no homogenization is available, and because these data were also used by KNMI for the construction of the Central Netherlands Temperature record (CNT) by Van Schrier et al. (2011). The following data sets were used:

- De Bilt 1–1-1901/31–8-1950: measured data using Pagoda screen on the old location.
- De Bilt 1–1-1901/31–12-1951: homogenized data with Eelde as reference station.
- De Bilt 1–1-1952/31–12-1995: measured data using Stevenson screen on the final location.
- Groningen/Eelde 1–3-1906/31–12-1965: data from city station Groningen homogenized with Eelde as reference station.
- Eelde 1–1-1946/31–12-1965: measured data from airport station.
- Maastricht/Beek 1–1-1906/31–12/1965: data from city station Maastricht homogenized with Beek as reference station.
- Beek 1–1-1946/31–12-1965: measured data from airport station.
- Den Helder 1–1-1935/31–12-1965: measured data from city station.
- Vlissingen 1–1-1935/31–12-1965: measured data from city station. This data series was interrupted 1–1-1947/31–12-1958, for which period homogenized data of airport station Souburg were used.
- Winterswijk 1–1-1935/31–12-1965: measured data.

5 Results and discussion

Figure 3 shows the results of 116 variants of the homogenization for De Bilt. Long-term records are sensitive to changes in degree of urbanization and siting exposure. This may probably have influenced most of the data sets used here, so it could be impossible to make reliable adjustments to daily records from 1901 to 1951 on the basis of overlapping data from shorter subseries. The results presented here only show the effect of the choice of the reference stations, length of the reference period, and the way of calculating percentiles and smoothing.

It is obvious from this figure that the choice of the reference station has considerable effect on the results. Using Groningen/Eelde or Vlissingen as reference station gives the smallest number of 1901–1951 tropical days for De Bilt. Using Maastricht/Beek as reference station gives much higher numbers. Using Den Helder or Winterswijk gives approximately the same results, intermediate between Groningen/Eelde and Maastricht/Beek.

The KNMI decided to use only Eelde as reference station for the homogenization of De Bilt because these two stations are located at comparable distances to large water surfaces. This is not quite true, because De Bilt is close to the former inland sea (Zuiderzee), which had an open connection to the North Sea until 1932, when the sea was closed by a 30 km long dike. Between 1940 and 1970, the former sea was partly impoldered. The effects of these geographical changes on the local climate are absent at station Eelde. Remarkably, using Vlissingen as reference station (a coastal station in the south west) gives very similar results as using Eelde as reference. On the contrary, using Den Helder (a coastal station to the north west) and Winterswijk (an inland station to the east) also gives similar results, but higher than when Eelde or Vlissingen are used. The conclusion could be, that an ideal reference station to homogenize the data of De Bilt does not exist. In the absence of a better reference station, the median of the results with all surrounding stations could be used, as was done by Trewin (2013) with the Australian data. An alternative could be to leave the historical data unchanged, and deduce climatic trends from the separate data of all stations.

The length of the reference period used seems to noticeably influence the spread of the results. With periods of 48–60 months, the largest spread is found. Any number of tropical days between 65 and 140 can be found. However, when applying longer reference periods, the results converge strongly. With a reference period of 168 months, the median of the number of tropical days found for De Bilt is 113 with margins 104–119. This result deviates considerably from the present KNMI data (76 tropical days before 1950).

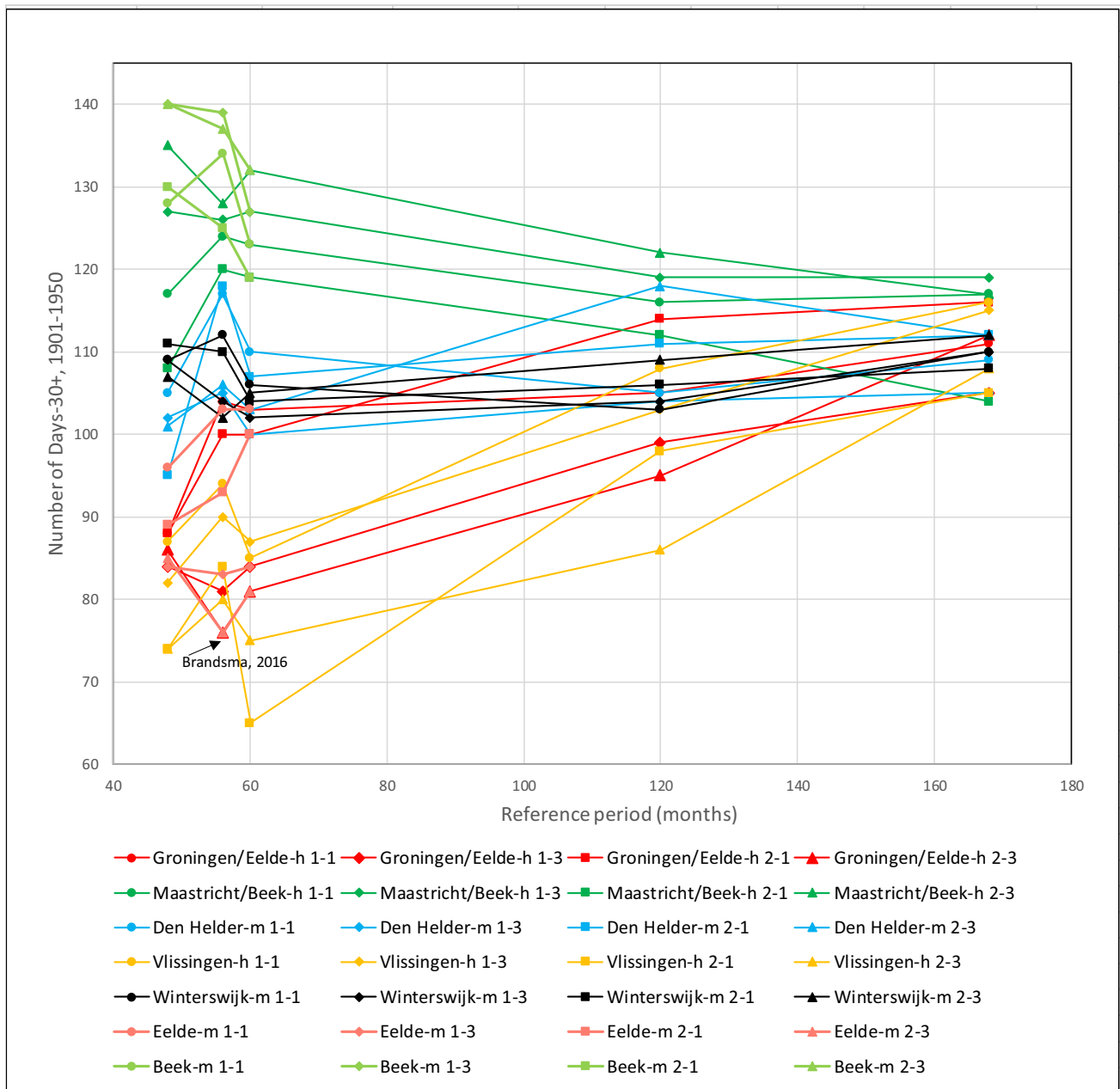


Fig. 3 Number of tropical days, from 1901 to 1950 found for De Bilt after homogenization with 5 reference stations, reference periods 48–168 months, 2 types of percentiles, 2 types of smoothing (116 runs of the algorithm). Explanation of the legends: Suffix -h: homog-

enized data. Suffix -m: measured data. First number: degree of the smoothing (1=linear, 2=quadratic). Second number: way of calculating the monthly percentiles (1=from single months, 3=from running groups of three months)

Calculating the monthly percentiles from running groups of three months has in some cases considerable effect, but in other cases less effect, as Table 2 shows. Coincidentally, this effect is most pronounced when using Groningen/Eelde as reference station with a 56-month reference period as was done by Brandsma (2016). In that case, the number of tropical days found for De Bilt decreases from 93 to 76. On the other hand, when using Maastricht/Beek as reference station, this way of calculating the percentiles leads to an increase

of the number of tropical days for De Bilt from 125 to 137. With a reference period of 168 months, the effect of the method of calculation percentiles has much smaller effects with all reference stations.

The degree of the Loess smoothing has a moderate downward effect on the number of tropical days found, except for Den Helder as reference station.

From Fig. 3 and Table 2, it is clear that the choices Brandsma (2016) made to homogenize the daily Tx of De

Table 2 Number of tropical days ($T_x > 30$) found for De Bilt 1901–1950 after homogenization with 5 reference stations, reference periods 56 and 168 months, 2 types of percentiles, 2 types of smoothing

Reference station	Percentile method	Smoothing method	Reference period 56 months	Reference period 168 months
Groningen/Eelde	Groups of 3 months	Quadratic	76 (Brandsma, 2016)	112
	Groups of 3 months	Linear	81	105
	Single months	Quadratic	100	116
	Single months	Linear	104	111
Maastricht/Beek	Groups of 3 months	Quadratic	128	117
	Groups of 3 months	Linear	126	119
	Single months	Quadratic	120	104
	Single months	Linear	124	117
Den Helder/De Kooy	Groups of 3 months	Quadratic	106	112
	Groups of 3 months	Linear	105	105
	Single months	Quadratic	118	112
	Single months	Linear	117	109
Vlissingen/Souburg	Groups of 3 months	Quadratic	80	108
	Groups of 3 months	Linear	90	115
	Single months	Quadratic	84	105
	Single months	Linear	94	116
Winterswijk	Groups of 3 months	Quadratic	102	112
	Groups of 3 months	Linear	104	110
	Single months	Quadratic	110	108
	Single months	Linear	112	110

Bilt from 1901 to 1951 lead to a very low number of tropical days. With Groningen/Eelde as reference station, 56-month reference period, percentiles from groups of three months and quadratic smoothing 76 tropical days are found. However, with different choices, different results can be found. Using other reference stations leads to 80–137 tropical days; using different reference periods leads to 81–112 tropical days. Taking percentiles from single months leads to 100–104 tropical days, and applying a linear Loess smoothing instead of quadratic smoothing gives 81–104 tropical days.

These results show that homogenization of temperature extremes by using the PMM is only possible when the results are reported with a wide uncertainty range. As was shown in Fig. 2, the homogenized data of De Bilt appear to be an outlier regarding the number of tropical days, compared to the other stations.

In Fig. 4, we explore the difference between station De Bilt and the other stations in two 15-year periods before and after 1950 for the lower percentiles. This shows that up to the 80th percentile the corrections for De Bilt do not differ much from the other stations. The median of the difference between De Bilt and the other stations is about -0.2 °C. However, for the highest percentiles, the difference increases strongly, up to $+0.7$ ° for the highest percentiles. Note that tropical days are all above the 98th percentile, and warm days

($T \geq 25$ °C) are above the 95% percentile. This suggests that the large reduction of tropical days in Brandsma's

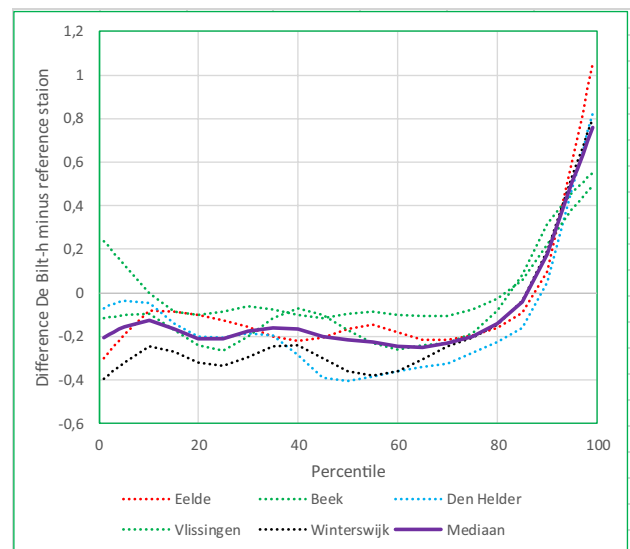


Fig. 4 Possible overcorrection of homogenized T_x De Bilt in highest percentiles (corresponding to T_x above 25°). The dotted curves show the difference of the homogenized data sets of De Bilt relative to each of the five other reference stations in the periods 1935–1949 and 1952–1966. The solid curve shows the median of the five series

homogenization is caused by a statistical effect that occurs especially at the high end of the temperature distribution.

A possible explanation could be that Brandsma (2016) calculated monthly percentiles from running groups of three months. When, for example, the percentiles of the month of May are calculated for April–June, the highest percentile is dominated by June days. This will cause a bias in the temperature distribution for May, so that the highest Tx for May days will be corrected as if they were June days, but this effect will be similar for the reference station and for the two reference periods. A bias could occur if this effect changes in a different way for the two reference periods at different reference stations. Considering these complicated consequences, it seems advisable not to use percentiles from groups of three months. Brandsma (2016) used this procedure because he found large differences between percentiles of adjacent single months and adjacent percentiles. This problem can be overcome by using longer reference periods. In Fig. 3, it is shown that with long reference periods, the results converge to a rather narrow interval of 104–119 tropical days.

The effect of the smoothing procedure is quite small. To smooth data, the Loess function uses extrapolation of local trends. This extrapolation can be performed with degree 1 (linear) or 2 (quadratic). Quadratic regression may in some cases give a somewhat better data fit, but it requires much more computing power. As Table 2 shows, the effect of quadratic regression in the PMM homogenization results in most cases in a lower number of tropical days. To our knowledge, there is no climatological or statistical argument to use a degree 2 smoothing.

6 Conclusions

The homogenization of the daily maximum temperatures at the five principal stations in the Netherlands has caused a significant decrease in the number of tropical days from 160 to 76 at the main station De Bilt in the period 1901–1950. As a result, De Bilt has become an outlier compared to the other stations. We reassessed the homogenization procedure using the same percentile matching algorithms as used by KNMI.

Since there were no parallel measurements for location changes around 1950 at De Bilt, KNMI used data of only one other station, Eelde located 148 km to the northeast, with two reference periods of 56 months before and after 1950. For our reassessment, we executed 116 runs of the homogenization algorithm, using five other reference stations, five reference periods up to 168 months, and different procedures for calculating percentiles and smoothing the data. We conclude that the results of the homogenization strongly depend on the choice of these parameters.

Only with a minority of the runs, we found as little tropical days before 1951 as KNMI. Alone with Eelde as a reference station, we found several combinations of reference period and calculation of percentiles that result in much higher numbers of tropical days for De Bilt (up to 116) before 1951. With other reference stations, we found up to 140 tropical days for De Bilt before 1951.

Major causes of the strong reduction of tropical days in the homogenization by KNMI are the calculation of percentiles from running groups of three months and the reference period of 56 months that is too short to correct daily Tx-values over a period of 50 years. Comparison of the Tx-corrections for De Bilt with the corrections for the other stations suggests that the average Tx-corrections for De Bilt do not deviate strongly from those of the other stations, but for the highest percentiles of Tx, the corrections are much higher than could be expected on the basis of the other stations.

Our observations suggest that the homogenization by KNMI of the temperature extremes in De Bilt from 1901–1950 needs to be reconsidered.

7 Detailed information

Annex 1 specifies the results of the 116 runs of the PMM algorithm that were used for Fig. 3 and Table 2. Annex 2 specifies the number of hot days from three time series: the originally measured data of De Bilt, the homogenized data by Brandsma, and the median of our runs with 168-month reference period.

Appendix 1: detailed information PMM-runs

Number of days $T_x \geq 30$ found for De Bilt 1901–1950 in all 116 PMM-runs after homogenization with five reference stations, reference periods 48–168 months, two types of percentiles (from one or three months), two types of smoothing (linear or quadratic).

Reference period (months)			48	56	60	120	168
Reference station	Number of months to take monthly percentiles	Degree of smoothing					
h=homogenized	m=measured						
Gron./Eelde-h	3	2	86	76	81	95	112
Gron./Eelde-h	3	1	84	81	84	99	105
Gron./Eelde-h	1	2	88	100	100	114	116

Reference period (months)		Degree of smoothing	48	56	60	120	168
Reference station h=homogenized m=measured	Number of months to take monthly percentiles						
Gron./Eelde-h	1	1	88	104	103	105	111
Mstr./Beek-h	3	2	135	128	132	122	117
Mstr./Beek-h	3	1	127	126	117	119	119
Mstr./Beek-h	1	2	108	120	119	112	104
Mstr./Beek-h	1	1	117	124	123	116	117
Den Helder-m	3	2	101	106	103	118	112
Den Helder-m	3	1	102	105	100	104	105
Den Helder-m	1	2	95	118	107	111	112
Den Helder-m	1	1	105	117	110	105	109
Vliss./Souburg-h	3	2	74	80	75	86	108
Vliss./Souburg-h	3	1	82	90	87	103	115
Vliss./Souburg-h	1	2	74	84	65	98	105
Vliss./Souburg-h	1	1	87	94	85	108	116
Winterswijk-m	3	2	107	102	105	109	112
Winterswijk-m	3	1	109	104	102	104	110
Winterswijk-m	1	2	111	110	104	106	108
Winterswijk-m	1	1	109	112	106	103	110
Eelde-m	3	2	85	76			
Eelde-m	3	1	84	83			
Eelde-m	1	2	89	93			
Eelde-m	1	1	96	103			
Beek-m	3	2	140	137			
Beek-m	3	1	140	139			
Beek-m	1	2	130	125			
Beek-m	1	1	128	134			

Appendix 2: detailed information on hot days

Time series of days $T_x \geq 25$, $T_x \geq 30$ in De Bilt from measured data, homogenization by Brandsma (2016) and median of all runs with 168-month reference period.

	Days $T_x \geq 25$			Days $T_x \geq 30$		
	Measured	Brandsma	Median-168	Measured	Brandsma	Median-168
1901	26	17	22	2	1	2
1902	14	11	12	3	1	2
1903	17	12	15	4	0	1
1904	19	16	17	2	2	2
1905	18	11	13	2	0	0
1906	18	10	14	1	0	0
1907	9	3	5	0	0	0
1908	15	10	10	0	0	0

	Days $T_x \geq 25$			Days $T_x \geq 30$		
	Measured	Brandsma	Median-168	Measured	Brandsma	Median-168
1909	9	6	7	0	0	0
1910	10	8	9	2	0	1
1911	43	31	35	12	8	10
1912	16	11	13	3	0	2
1913	14	10	11	0	0	0
1914	31	17	26	0	0	0
1915	10	7	8	1	1	1
1916	11	7	7	1	0	0
1917	26	14	16	4	2	2
1918	12	9	11	1	1	1
1919	21	15	17	4	1	4
1920	15	8	8	0	0	0
1921	37	30	32	5	2	4
1922	11	10	10	3	3	3
1923	18	13	15	8	5	7
1924	11	6	8	1	0	0
1925	26	21	22	3	2	2
1926	17	9	14	2	0	1
1927	16	8	11	1	0	0
1928	17	8	11	3	1	1
1929	32	21	26	4	2	2
1930	23	16	20	5	2	3
1931	14	10	12	0	0	0
1932	35	23	26	4	2	3
1933	29	13	21	2	1	1
1934	35	15	20	2	0	1
1935	30	21	26	4	0	3
1936	20	12	13	4	2	3
1937	21	13	18	3	2	2
1938	18	11	13	3	1	2
1939	30	17	22	2	1	1
1940	20	9	9	0	0	0
1941	28	20	22	11	9	10
1942	21	14	17	4	1	3
1943	23	15	20	4	0	0
1944	28	14	18	5	4	5
1945	21	14	16	4	2	2
1946	12	9	10	3	1	1
1947	64	47	55	18	8	12
1948	23	15	17	7	6	7
1949	32	20	24	3	1	2
1950	20	12	13	4	1	2
1901–1950	1086	699	840	164	76	113

Acknowledgements The authors thank Mr. Theo Brandsma for providing unpublished technical details of his procedure and for providing the data of Eelde, Beek and Winterswijk. The authors thank Mr. Ben Lankamp for providing the R-code used in this study. The authors thank Mr. Ronan Connolly for reading and commenting an earlier version of the manuscript.

Author contribution All authors contributed to the study conception and design. Data collection and analysis were performed by Frans Dijkstra and Jan Ruis; description of the stations in the Netherlands and general literature review were performed by Rob de Vos. The first draft of the manuscript was written by Jan Ruis, modified by Frans Dijkstra, and all authors commented on further versions of the manuscript. Communication with external experts was done by Marcel Crok. All authors read and approved the final version of the manuscript.

Funding The research was supported by a grant from Clintel Foundation, Amsterdam.

Data availability The data used in this research can be downloaded from <https://klimaatgek.nl/wordpress/data/>.

Code availability The codes used can be downloaded from <https://klimaatgek.nl/wordpress/data/>.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Acquaotta FS, Fratianni S, Venema V (2016) Assessment of parallel precipitation measurements networks in Piedmont, Italy. *Int J Climatol* 36(12):3963–3974. <https://doi.org/10.1002/joc.4606>
- Brandsma T, Können GP, Wessels HRA (2003) Empirical estimation of the effect of urban heat advection on the temperature series of De Bilt (The Netherlands). *Int J Climatol* 23:829–845
- Brandsma, T. (2016) Homogenization of daily temperature data of the five principal stations in the Netherlands (version 1.0). KNMI Technical Report TR-356
- Cao L, Zhu Y, Tang G, Yuang F, Yan Z (2016) Climatic warming in China according to a homogenized data set from 2419 stations. *Int J Climatol* 36(13):4384–4392
- Della-Marte PM, Wanner H (2006) A method of homogenizing the extremes and mean of daily temperature measurements. *J Climate* 19(17):4179–4197
- Koopmans S, Theeuwes NE, Steeneveld GJ, Holtslag AAM (2012) De invloed van urbanisatie op de meetreeks van De Bilt. *Meteorologica* 21(2):9–14
- Kuglitsch FG, Toreti A, Xoplaki E, Della-Marta PM, Luterbacher J, Wanner H (2009) Homogenization of daily maximum temperature series in the Mediterranean. *J Geoph Res* 114:D15108. <https://doi.org/10.1029/2008JD011606>
- Li S, Lund R (2012) Multiple changepoint detection via genetic algorithms. *J Climate* 25(2):674–686
- Lindau R, Venema V (2015) The uncertainty of break positions detected by homogenization algorithms in climate records. *Int J Climatol* 36:576–589
- Mestre O, Gruber C, Prieur C, Caussinus H, Jourdain S (2011) SPLID-HOM: a method for homogenization of daily temperature observations. *J Appl Meteor Climatol* 50(11):2343–2358
- Peterson TC et al (1998) Homogeneity adjustments of in situ atmospheric climate data: a review. *Int J Climatol* 18:1493–1517
- Reeves J, Chen J, Wang XL, Lund R, Lu QQ (2007) A review and comparison of changepoint detection techniques for climate data. *J Appl Meteor Climatol* 46:900–915
- van der Schrier G, van Ulden A, van Oldenborgh GJ (2011) De Centraal Nederland Temperatuurreeks *Meteorologica* 20(4):10–15
- Squintu AA, van der Schrier G, Stepanek P, Zahradnicek P, Klein Tank A (2020) Comparison of homogenization methods for daily temperature series against an observation-based benchmark dataset. *Theor Appl Climatol* 140:285–301
- Trewin B (2010) Exposure, instrumentation, and observing practice effects on land temperature measurements. *Wires Clim Change* 1:490–506
- Trewin B (2013) A daily homogenized temperature data set for Australia. *Int J Climatol* 33:1510–1529
- Trewin BC, Trevitt ACF (1996) The development of composite temperature records. *Int J Climatol* 16:1227–1242
- Venema VKC et al (2012) Benchmarking homogenization algorithms for monthly data. *Clim past* 8:89–115

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.