**Climate
of the Past**

The construction of a Central Netherlands temperature

G. van der Schrier[1], A. van Ulden[*], and G. J. van Oldenborgh[1]

[1]KNMI, P.O. Box 201, 3730 AE De Bilt, The Netherlands
[*]retired

Abstract. The Central Netherlands Temperature (CNT) is a monthly daily mean temperature series constructed from homogenized time series from the centre of the Netherlands. The purpose of this series is to offer a homogeneous time series representative of a larger area in order to study large-scale temperature changes. It will also facilitate a comparison with climate models, which resolve similar scales.

From 1906 onwards, temperature measurements in the Netherlands have been sufficiently standardized to construct a high-quality series. Long time series have been constructed by merging nearby stations and using the overlap to calibrate the differences. These long time series and a few time series of only a few decades in length have been subjected to a homogeneity analysis in which significant breaks and artificial trends have been corrected. Many of the detected breaks correspond to changes in the observations that are documented in the station metadata.

This version of the CNT, to which we attach the version number 1.1, is constructed as the unweighted average of four stations (De Bilt, Winterswijk/Hupsel, Oudenbosch/Gilze-Rijen and Gemert/Volkel) with the stations Eindhoven and Deelen added from 1951 and 1958 onwards, respectively.

The global gridded datasets used for detecting and attributing climate change are based on raw observational data. Although some homogeneity adjustments are made, these are not based on knowledge of local circumstances but only on statistical evidence. Despite this handicap, and the fact that these datasets use grid boxes that are far larger then the area associated with that of the Central Netherlands Temperature, the temperature interpolated to the CNT region shows a warming trend that is broadly consistent with the CNT trend in all of these datasets. The actual trends differ from the CNT trend up to 30 %, which highlights the need to base future global gridded temperature datasets on homogenized time series.

Correspondence to: G. van der Schrier
(schrier@knmi.nl)

1 Introduction

In the Netherlands, the earliest temperature observations were made at the end of the 17th century. From 1706 onwards, systematic measurements were made and a continuous record, albeit constructed from several sources, exists (Labrijn, 1945). Because of the lack of standardization in observation procedures, instruments and observations screens, the construction of a homogeneous record on the basis of these early instrumental records is difficult and is not attempted here.

In 1906 a climatological network had become operational in the Netherlands that employed a highly standardized observation practice and a type of Stevenson screens at all stations but one. The exception was the station De Bilt, where this replacement happened on 17 May 1950 and where the Stevenson screen replaced a large thermometer screen (called "Pagoda"), which had a thermograph located at 2.20 m on the peak of the Pagoda's roof. A set of main stations made observations on an hourly basis, while secondary stations in the climatological network took measurements thrice daily. Around 1950, a new synoptic network was installed. This was operated by the Weather Forecasting department of KNMI in parallel to the climatological network, which was operated by the Climate Division at KNMI. This situation persisted until around 1990, when the two networks were integrated to form a single, fully automated, observation network.

The locations of the observing stations relevant for this study, both the stations that ceased operation and the operational ones, are shown in Fig. 1.

The aim of this study is twofold. The first is to construct a set of homogeneous monthly averaged records for daily mean temperature at various locations spread over the Netherlands. These records are either based on long continuous records from the KNMI network or, when these are not available, on combinations of two records from nearby stations to obtain time series as long as possible. Next, based on

Fig. 1. Map of the Netherlands with the station locations and station types.

a selection of these homogeneous records, a Central Netherlands Temperature (CNT) record is assembled that is, by construction, representative for a larger area.

In a precursor of this study, van Ulden et al. (2009) based an earlier version of the CNT on the same monthly averaged station records but used a different method to homogenize these records and made different choices in the application of their method than what is done here. Differences are found in the construction of the reference series, aggregation levels, window size etc. Despite these different approaches, we will show that the locations and sizes of most of the detected breaks in the station records are similar between the current study and that of van Ulden et al. (2009) and that consequently, the differences between the CNT as presented in this study and the one presented by van Ulden et al. (2009) are very small. The robustness of the CNT for

different approaches in arriving at an estimate of temperature representative for a larger region adds to the confidence in it.

2　Construction of long time series

At the secondary stations in the climatological network operating since the early 20th century, temperature readings have been made at 08:00, 14:00 and 19:00 LT (local time) as well as the minimum and maximum temperatures reached in the time period 19:00–19:00 LT. Based on these measurements, van der Hoeven (1992) made accurate estimates of daily (00:00-00:00 LT) mean temperature. This approach is a refinement of a method used in the 1980s at KNMI and aims to give a better representation of the daily cycle in temperatures. Note that only for the main stations in the climatological

Table 1. Climatological records analysed in this study. (H-records based on 24 hourly observations; G-records based on 5 observations per day until 1970 (at 08:00, 14:00 and 19:00 h plus minimum and maximum), 10 observations per day from 1971 onwards (at 3-hourly intervals plus minimum and maximum).

station	record id	period	missing data	filled with
De Bilt	D001 H	1901–1970		
Den Helder	D002 H	1906–1970	Sep 1944–May 1945	Hoorn
Vlissingen	D003 H	1906–1970	1918–1930, 1944–1945	Excluded from analysis
Eelde	D004 H	1946–1970		
Beek	D005 H	1946–1970		
Groningen	D006 H	1906–1951		
Maastricht	D007 H	1906–1952		
De Kooy	D009 H	1961–1970		
Winterswijk	D020 G	1906–1990	Nov 1944, Oct 1988	De Bilt
Hoorn	D029 G	1906–1990	Nov 1947–Apr 1948	Den Helder
Oudenbosch	D032 G	1906–1992		
Gemert	D033 G	1906–1990		
Sittard	D145 G	1906–1948	Apr–Aug 1940, Nov 1944–Feb 1945	Maastricht
Gilze-Rijen	D132 G	1953–1970		
Twenthe	D146 G	1947–1970		

Table 2. Synoptic records based on 24 hourly observations.

station	record id	period	missing data	filled with
Den Helder	235	1971–Jul 1972		
De Kooy	235	Aug 1972–2008		
Schiphol	240	1951–2008		
De Bilt	260	1951–2008		
Soesterberg	265	1953–2007		
Leeuwarden	270	1951–2008		
Deelen	275	1958–2008		
Eelde	280	1951–2008		
Hupsel	283	1990–2008		
Twenthe	290	1971–2008		
Rotterdam	344	1957–2008		
Gilze-Rijen	350	1971–2008		
Eindhoven	370	1951–2008	May, June 1952	Gemert and Oudenbosch
Volkel	375	1953–2008		
Beek	380	1971–2008		

network, minimum and maximum temperatures reached between consecutive readings, rather than those reached in a 24-h period, are recorded.

The approach of Van der Hoeven (1992) was to make five estimates of the daily mean temperature T_{24}:

$$T_{24}^{(i)} = T_i - C_i(t)\,(T_x - T_n), \quad i = 1, 2, 3, x, n \quad (1)$$

where subscripts 1,2,3 refer to temperature readings at 08:00, 14:00 and 19:00 LT and subscripts x and n to daily maximum and minimum temperatures. The coefficients $C_i(t)$ were seasonally dependant and obtained from a comparison with 24-h temperature observations at De Bilt in the period 1961–1970. The seasonal dependence of the coefficients is introduced to account for the annual variations in the times of sunset and sunrise and are computed for the 36 decades of the year. Finally, the five estimates of the daily mean temperature $T_{24}^{(i)}$ were averaged to give the best estimate:

$$T_{24} = (T_1 + T_2 + T_3 + T_x + T_n$$
$$\quad - (C_1 + C_2 + C_3 + C_x + C_n)\,(T_x - T_n))/5. \quad (2)$$

Applying this approach to the De Bilt data results in the value estimated from these five measurements agreeing with the true 24-h mean to within 0.006 °C. For the coastal station Den Helder, in the north-west corner of The Netherlands, the estimate of daily averaged temperatures are ≈0.1 to 0.2 °C too low, which is probably related to changes in the daily cycle of this location compared to that of the De Bilt which is more inland.

For the period up to 1970, daily averaged temperatures for the secondary stations are calculated following the method described above.

From 1970 until the introduction of the automated weather systems, ten measurements a day were made at the secondary stations: eight at three-hourly intervals plus the minimum and maximum temperatures. Twenty-four hour averages were made by an unweighted average of these values.

The secondary stations involved in this study are indicated with a "G" in Table 1.

Stevenson huts were used in all stations until about 1990, with the exception of De Bilt between 1901–1950. From around 1990 onwards, a new automated observing system was gradually introduced using small multi-plate thermometer screens. This transition has had a negligible effect on monthly mean temperatures (Brandsma and van der Meulen, 2008).

Most records in Table 1 were complete. Den Helder and Sittard had 9 missing months, Winterswijk had 1 missing month, Hoorn had 6 missing months and Eindhoven missed May and June 1952. These missing data were filled with data from alternative stations with a monthly adjustment to account for any climatological differences (see Tables 1 and 2). The record from Vlissingen appeared to be too incomplete to be useful for this study. All records have been tested for outliers, but none were found with the exception of Eindhoven which is discussed in Sect. A12. Eight long records were constructed covering the period 1906–1908 by merging the records with records from nearby stations (see Table 3). The older parts of these merged records were adjusted to the recent parts using overlapping observation periods. The monthly adjustment factors were smoothed with a 5-point quasi-gaussian filter.

For the transition Winterswijk to Hupsel, the overlapping period was only 10 months, which is too short for a reliable

**Table 3.** Long composite records used for break and trend analysis.

station	record ids	period	transition	overlap
De Bilt	D001+260	1901–2008	Jan 1971	
Den Helder/De Kooy	D002+235	1906–2008	Aug 1972	1961–1970
Groningen/Eelde	D004/006+280	1906–2008	Jan 1946	1946–1951
Maastricht/Beek	D005/007+380	1906–2008	Jan 1946	1946–1952
Winterswijk/Hupsel	D020+283	1906–2008	Jan 1991	Mar–Dec 1990
Hoorn	D029+270	1906–1990	No suitable follow-up	
Oudenbosch/Gilze-Rijen	D032+350	1906–2008	Jan 1993	1988–1992
Gemert/Volkel	D033+375	1906–2008	Jan 1991	1990
Sittard/Beek	D145+380	1906–2008	Jan 1946	1946–1948
Twenthe	D146+290	1946–2008	Jan 1971	
Schiphol	240	1951–2008		
Soesterberg	265	1953–2008		
Deelen	275	1958–2008		
Rotterdam	344	1957–2008		
Eindhoven	370	1951–2008		

estimate of the adjustment factors. Therefore we used a 10 yr overlap of both time series with Deelen to determine the adjustments. The smoothed adjustment factors are shown in Fig. 2. We can see in this figure that the adjustment factors are all negative, meaning that the recent stations are cooler than the older stations. This may be related to an urban heat island effect for stations like Maastricht and Groningen, where observations were made in the centre of the city in the earlier period and at the airport later on. In Maastricht, the modern station at the airport is also located at a much higher and exposed position.

3 Method

The approach taken to identify possible changepoints and estimate the size of the break is based largely on the two-phase regression technique suggested by Vincent (1998). Potential discontinuities are detected on 40-yr sliding windows of the difference time series between the target series and a (homogeneous) reference series. The construction of the reference series is discussed in Sect. 4. Easterling and Peterson (1995) note that a windowing technique may obscure discontinuities which are close in time but have a sliding window with increments of one year will (at least partially) eliminate this problem. In order to prevent this problem, homogenized records are put through the detection algorithm to detect possible inhomogeneities which were left undetected in the first run. For the Eindhoven record, this procedure yielded a breakpoint which remained undetected in the first run of the program.

The two-phase regression technique applies four different models. The first model determines whether the time series is homogeneous over the tested interval of time. If the time
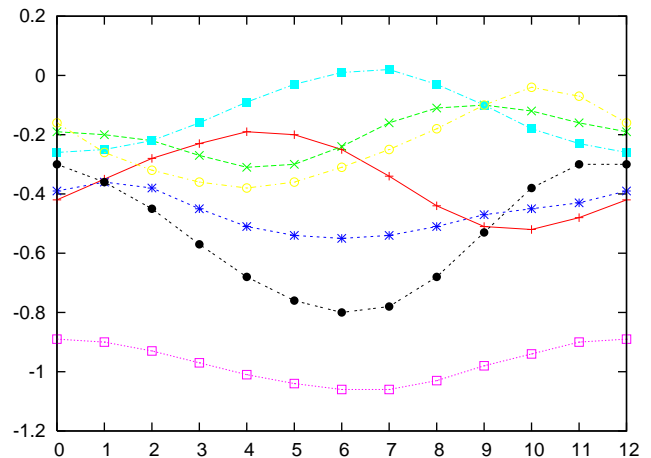
Fig. 2. Adjustment factors used to blend series from operational stations to those from stations which ceased operation. (red = Den Helder/De Kooy, green = Gemert/Volkel, blue = Groningen/Eelde, pink = Maastricht/Beek, magenta = Oudenbosch/Gilze Rijen, yellow = Winterswijk/Hupsel, black = Sittard/Beek) [°C].

series fails to pass this test, three different models are used to estimate the location and size of a potential step. Every year in the record is tested as a potential step with the exception of the first three and last three years. The regression is performed on the difference series resulting from the subtraction of the target series and a homogeneous reference series.

To test if a series is inhomogeneous, a straight line is fitted to the data. The goodness of fit is quantified using the Durbin-Watson statistic, which is a test for the correlation of regression residuals (Wilks, 1995, Sect. 6.2.6). It tests the null-hypothesis that the residuals are serially independent against the alternative that they are consistent with

a first-order autoregressive process. The threshold for the Durbin-Watson statistic relates to the 5 % level where the null-hypothesis of zero serial correlation can either be rejected or where this statistic is indeterminate. If adjacent residuals are of similar magnitude, as would occur in a bad fit, the Durbin-Watson statistic tends to be small. On the other hand, when residiuals are randomly distributed in time, this statistic tends to be large. Therefore one does not reject the null hypothesis that the residuals are independent if the Durbin-Watson statistic is sufficiently large. Upper and lower bounds for the significance of the Durbin-Watson statistic are calculated using the NAG routine g01epf.

If the difference series is judged inhomogeneous, the location and size of the break are estimated using a simple two-phase regression model (Vincent, 1998)

$$X_t = \begin{cases} \mu_1 + \varepsilon_t, & 1 \le t \le c \\ \mu_2 + \varepsilon_t, & c < t \le n, \end{cases} \qquad (3)$$

where μ_1, μ_2 are mean values before and after the break and ε_t is the zero-mean independent random error with a constant variance σ_ε^2. The time c is called a changepoint if $\mu_1 \ne \mu_2$. The F statistic for a changepoint at time c is:

$$F_c = \frac{(SSE_{red} - SSE_{full})/1}{SSE_{full}/(n-2)}, \qquad (4)$$

where SSE_{full} is the sum of squared errors of the "full" model Eq. (3), which includes the break and SSE_{red} is the sum of squared errors of the "reduced" model which assumes a constant mean.

Slightly more complex is the two-phase regression model with a common trend (Wang, 2003):

$$X_t = \begin{cases} \mu_1 + \beta t + \varepsilon_t, & 1 \le t \le c \\ \mu_2 + \beta t + \varepsilon_t, & c < t \le n, \end{cases} \qquad (5)$$

where β is the value of the trend. The F statistic for a changepoint at time c is:

$$F_c = \frac{(SSE_{red} - SSE_{full})/1}{SSE_{full}/(n-3)}. \qquad (6)$$

The fourth model allows for a combination of a discontinuous trend and a step (Lund and Reeves, 2002):

$$X_t = \begin{cases} \mu_1 + \beta_1 t + \varepsilon_t, & 1 \le t \le c \\ \mu_2 + \beta_2 t + \varepsilon_t, & c < t \le n, \end{cases} \qquad (7)$$

with β_1, β_2 values of the trend before and after the break. The F statistic for a changepoint at time c is:

$$F_c = \frac{(SSE_{red} - SSE_{full})/2}{SSE_{full}/(n-4)}. \qquad (8)$$

Under the null hypothesis of no changepoints and assuming Gaussian errors ε_t in models Eqs. (3), (5), (7), tables with the F_{max} percentiles are given by Jarusková (1996), Wang (2003) and Lund and Reeves (2002) respectively. The 95 %

siginificance level is used as a threshold to determine if a break is significant or not.

If a fit of a model fails to meet the significance level using the Durbin-Watson statistic, it is not considered further.

A review of modern methods, including the methods used here, is given by Reeves et al. (2007); they concluded that the common trend two-phase regression model seems optimal for most time series.

A hierarchy is used in determining which changepoint models Eqs. (3), (5), (7) are used to estimate step sizes. If a difference series is inhomogeneous, models Eqs. (3) and (5) are applied and information from model Eq. (7) is only used after a visual confirmation that a discontinuous trend is present. No distinction is made for information on the step size from models Eqs. (3), (5), the estimate of the continuous trend from model Eq. (5) is not used to correct for this trend. The motivation for not correcting for a continuous trend is related to the construction of the reference series. Both the construction of the reference series and this motivation are discussed in Sect. 4. However, discontinuous trends (which are output from model Eq. 7) are corrected for.

It is possible to formalize the choice between the various model using a statistical test. In an attempt to do this, we noticed that model Eq. (7) was chosen in more cases than what could be confirmed by the available metadata. This observation made us change the procedure and adjusted both the step and the trend when the metadata indicated that these adjustments were required.

4 Reference time series

In the absence of homogeneous time series, constructing a (near-)homogeneous reference series requires a special approach. Instead of using a tailored approach where an average of a small number of selected time series is used as reference series from the vicinity of the target record, we use the most dominant mode of variability from a Principal Component Analysis (PCA) based on all available long time series. The first mode of variability accounts for the maximum amount of joint variability of the variance-covariance matrix (Wilks, 1995), which is based, in its turn, on a selection of long station data homogeneously spread over the country. The principal mode of variability is a weighted average of the input series and contains a large fraction of the common variability of the series. This time series will have the warming trend common to all time series and, due to the averaging of all available long records, inhomogeneities in the individual records are damped. However, a reference series constructed from time series scattered over the country will not reflect any regional signal. Other considerations are that the tailored approach is more labour-intensive and can hardly be automated, but the expectation is that a tailored approach will provide a reference series capturing more of the month-to-month variability, thus reducing the noise in the

difference series. This should make it easier to detect smaller breaks. The PCA-based method is a rather straightforward procedure, easily automated, but potentially less suited to homogenize a series which is not near the centre of the country. The decorrelation length of the interannual variability of monthly mean temperature throughout the Netherlands varies from about 1000 km in summer to 2000 km in winter, which is much larger than the size of the Netherlands, so regional effects are not expected to be very large.

This approach contrasts with that of van Ulden et al. (2009), who use the average of nearby stations as the reference series. Peterson et al. (1998) note that the construction of a reference series by simply averaging series from surrounding stations has been done earlier by Potter (1981), although he used an average of 18-stations for this. More specifically, Peterson and Easterling (1994) average the three best correlating series from the 5 nearest stations to build the reference series.

Input to the PCA are the time series of the stations De Bilt, Groningen/Eelde, Winterswijk/Hupsel, Maastricht/Beek, Volkel/Gemert, Oudenbosch/Gilze-Rijen and Den Helder/De Kooy. This excludes the station Vlissingen in the southwestern corner of the Netherlands, which is too gappy and has seen too many relocations to be allowed into the reference time series.

Over the more recent period, from the 1950s onward, more stations have become available to construct a reference series. In order not to introduce inhomogeneities in the reference series, we did not include these stations in the reference series. Moreover, the seven stations used for the reference series over the first part of the 20th century are scattered around the country and should be sufficiently able to pick up on large-scale variations of temperature.

When using a weighted sum of series as a reference series, a correction has to be made when one of the series which is input to the PCA analysis is adjusted. The reference series is written as

$$\sum_{i=1}^{N} c_i A_i(t), \tag{9}$$

where c_i is the weight associated with the i-th series and $A_i(t)$ is the i-th series at time t. The sum of the weights c_i equals 1. When the adjustments to the j-th series itself needs to be calculated, a reference series excluding the j-th series would be required. The difference from which an adjustment is computed is:

$$\Delta T = A_j(t) - \frac{1}{1 - c_j} \sum_{i=1, \, i \neq j}^{N} c_i A_i(t). \tag{10}$$

By writing Eq. (9) as

$$\sum_{i=1}^{N} c_i A_i(t) = \left(\sum_{i=1, \, i \neq j}^{N} c_i A_i(t) \right) + c_j A_j(t), \tag{11}$$

we see that

$$\Delta T = \frac{1}{1 - c_j} \left(A_j(t) - \sum_{i=1}^{N} c_i A_i(t) \right). \tag{12}$$

One can therefore just use the total reference series and multiply corrections by $1/(1-c_j)$ afterwards. Alternatively, one could recalculate the reference series for each target series by using this PCA-based technique but excluding the target series from this calculation. An adjustment to the corrections is then not neccessary.

To test the merits of the PCA-based approach, the reference series used by van Ulden et al. (2009) and the PCA-based reference series are compared for a selection of stations. The RMS between the target series and either of the two reference series is calculated (correcting for any offset). It turns out that the RMSs are very similar (not shown), indicating that for this study involving stations that are much closer to each other than the decorrelation length, a PCA-based reference series does not give higher noise-levels in the difference series, compared to a tailored approach.

The different models discussed in Sect. 3 are combined in the approach of this study which has merit in cases where the "true" regression model is unknown (Reeves et al., 2007). Since the reference series used in this study only holds information associated with country-wide spatial scales and is not specifically pin-pointed at a certain region, we expect that difference series may have a continuous trend throughout the record. This makes the use of model Eq. (5), which includes a step and a continuous trend, particularly suited to the approach.

The principal mode explains 96.7 % of the variability and includes the warming trend. The dominant mode of variability is a weighted average of the input series, with the weights shown in Table 4. The weights for the various stations are very similar, the relative difference between the extremes ((maximum−minimum)/maximum) is only 0.16. The largest weights are found in Maastricht/Beek and Winterswijk/Hupsel, the lowest is found in Den Helder/De Kooy, located at the North Sea coast.

5   Quality check

In order to assess the quality of the various records used in this study, a running standard deviation of the difference of the annual average of each series with the reference series is shown over 41 yr sliding windows (Fig. 3). The standard deviations vary considerably with time. Fig. 3a shows that Sittard has a maximum in the first few decades of the record, which is in part related to a warm bias (not shown). Gemert has a very pronounced peak around 1950, which is related to a very significant break in that period (discussed in Sect. A6). The running standard deviation for the series composed of Oudenbosch and Gilze Rijen has a maximum

Table 4. Loading of the seven long records on the first Principal Component.

station	loading
De Bilt	0.143
Den Helder/De Kooy	0.128
Groningen/Eelde	0.144
Maastricht/Beek	0.150
Winterswijk/Hupsel	0.149
Oudenbosch/Gilze-Rijen	0.141
Gemert/Volkel	0.145

in the late 1960s which is related to a break in that period (discussed in Sect. A4).

De Bilt, which is situated in the centre of the Netherlands, has low standard deviations for the whole observation period. This relates to the fact that the reference series reflects climatic conditions of the central part of the Netherlands best.

Figure 3b shows that Soesterberg, despite its central location, and Twenthe both show high noise levels.

Around 1990, the noise levels of all stations (except for Beek) are significantly reduced (not shown). This is probably related to the transition to an automated network and improved observation practices.

6 The Central Netherlands temperature

6.1 Definition

The Central Netherlands Temperature (CNT) record is based on homogenized monthly means of daily averaged temperatures from a selection of series from the central part of The Netherlands. These series are from De Bilt, Winterswijk/Hupsel, Oudenbosch/Gilze-Rijen and Gemert/Volkel. The record from Eindhoven is included from 1951 onwards and Deelen is included from 1958 onwards. The CNT is a simple unweighted average of these records. Monthly adjustments were applied to the CNT prior to the inclusion of the Eindhoven record in 1951 and to the CNT record from 1951 to the inclusion of Deelen in 1958 to account for the transition from 4 to 5 to 6 stations. These adjustments are calculated over the 1961–2008 period, smoothed by a 5-point Gaussian filter, similar to the adjustments in Sect. 2 and are small at $O(0.01\,°\mathrm{C})$.

The long records from the coastal station Den Helder/De Kooy, the series from Groningen/Eelde and Leeuwarden in the north and Maastricht/Beek and Sittard/Beek in the south of the Netherlands have not been included since they are too far at the outer extremes of The Netherlands and are therefore less representative of the central Netherlands. The principal motivation not to include the records from the airports of Rotterdam and Amsterdam (Schiphol) is that these stations are
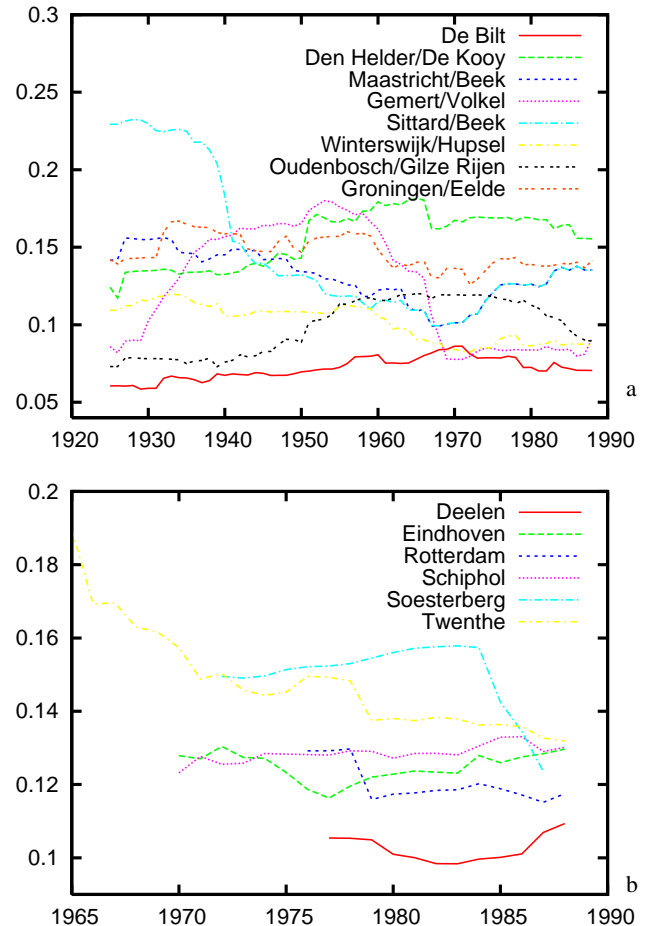

Fig. 3. Running standard deviations over 41 yr windows of annual averages of the difference between target series and reference series. Upper panel shows the long records, lower panel shows the shorter records.

relatively close to the sea and that they may be influenced by the large cities in their vicinity and, in the case of Schiphol, the rapid development of the airport itself.

Other long records in The Netherlands have either been discontinued (Hoorn, Soesterberg) or are too gappy (Vlissingen) to be included in the CNT. The running standard deviations discussed in Sect. 5 and Fig. 3 indicate that the records from Twenthe and Sittard are too noisy to be included.

Although the CNT is constructed from stations roughly centred in the central Southeast of the Netherlands, a correlation analysis between winter (DJF) and summer (JJA) averages of the CNT and similar averages of the E-OBS gridded dataset based on daily averaged temperature from surrounding stations from the European Climate Assessment & Dataset (Haylock et al., 2008) (Fig. 4a,b) shows that the CNT is representative of a much larger area. The correlations remain high for stations in the Netherlands, including those in the North and along the coast; high correlations are found in Belgium and Germany as well. The monthly mean
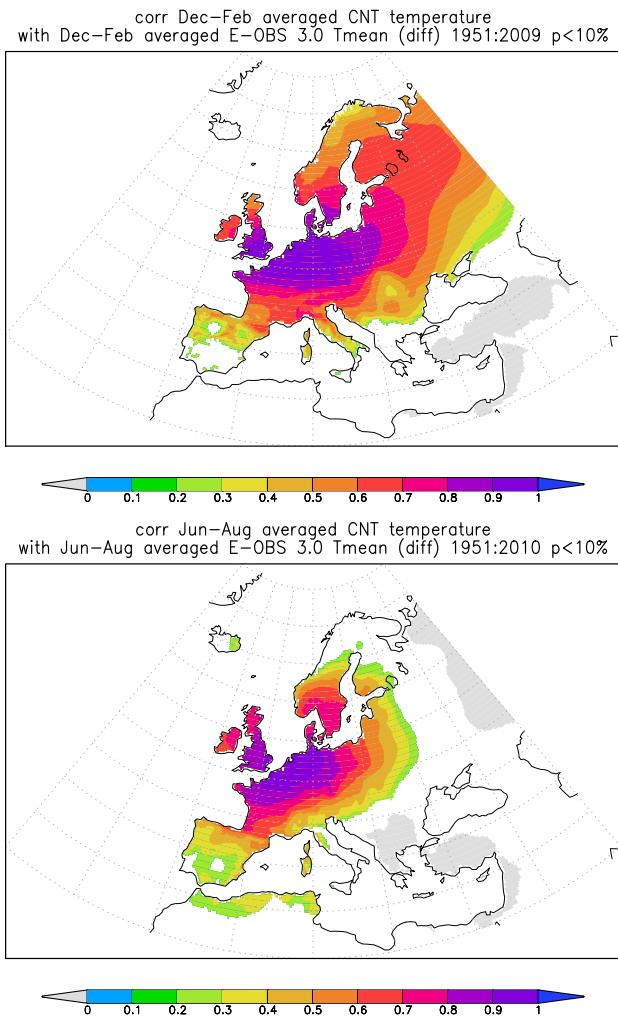
Fig. 4. Correlation of the interannual fluctuations of the CNT series with the E-OBS v3 (Haylock et al., 2008) temperature analysis for three winter months (December, January and February) and three summer months (June, July and August) over 1950–2009. The trend was removed by taking year-on-year differences.

values show smaller correlations, especially in summer (not shown).

6.2 Comparison with an earlier version

Preceding this study, van Ulden et al. (2009) constructed an earlier version of the Central Netherlands Temperature record. This earlier version, to which we attach the version number 1.0, is based on the same selection of series as the current version 1.1. However, the homogenisation procedure between the van Ulden et al. (2009) study and the current study is different. These differences relate to the construction of the reference series (see Sect. 4) and also to the break and spurious trends detection algorithms. For the construction of CNT1.0, homogeneity tests based on Easterling and Peterson

(1995) for the detection of breaks were used and a method based on that of Alexandersson and Moberg (1997) was used to detect spurious trends. In contrast to Easterling and Peterson (1995), van Ulden et al. (2009) used moving windows, both for breaks and trends. In both tests, the critical significance levels derived from Alexandersson and Moberg (1997) were used.

Because of the differences between the current approach and that of van Ulden et al. (2009), differences in the homogenized versions of the underlying station data can be expected. Figure 5 shows the adjustments made to the records in comparison with the adjustments made in version 1.0. Below we argue that the differences between CNT1.1 and CNT1.0 are very modest, despite the difference in approach, which adds to the robustness of the CNT.

Figure 6 shows the RMS between CNT1.1 and CNT1.0 as a function of the month. The RMS has been determined over the time period 1906–2008. This figure shows that the RMS is near 0.04 °C, except for late spring and for July to September. In the first period, the RMS rises to nearly 0.04 °C, in the second to approximately 0.07 °C. The rise in April–May can be attributed to differences in the homogenisation of the Eindhoven record (Fig. 7). The current study has an additional correction for a break in 1958. The rise in the period July to September is mainly attributed to the Winterswijk/Hupsel, Deelen and Gemert/Volkel series (Fig. 7). In the homogenisation of the first record (Sect. A5), a small break in 1960 is detected, but we have not corrected for this break due to the absence of metadata related to it. However, van Ulden et al. (2009) do correct for this break and the largest amplitude of the correction can be found in the months June–September. In the Deelen record, breaks are detected which are corrected for in the current study but not in the series used to construct CNT1.0. Finally, the adjustments to the Gemert/Volkel record are slightly different in the current study compared to that of van Ulden et al. (2009).

Regarding the trends, Table 5 shows that the differences between CNT1.0 and CNT1.1 are very small.

7 Comparison of the CNT v1.1 with the HadCRUT3, GISTEMP and NCDC datasets

In Fig. 8, the annual mean CNT1.1 time series is compared with other series that are frequently used to estimate the temperature changes in the Netherlands. These are the CNT1.0, the observed De Bilt temperature, and the interpolated temperature of datasets used to construct estimates of the global mean temperature (CRUTEM3, NOAA/NCDC and NASA/GISS) at the mean of the co-ordinates of the six CNT stations (using land temperatures only).

By eye these six time series look very similar. Two series with obvious errors are given in Fig. 8d,h: the unadjusted De Bilt temperature from GHCN v2 and the value corresponding to the CNT in the GISTEMP 250 km dataset, which is
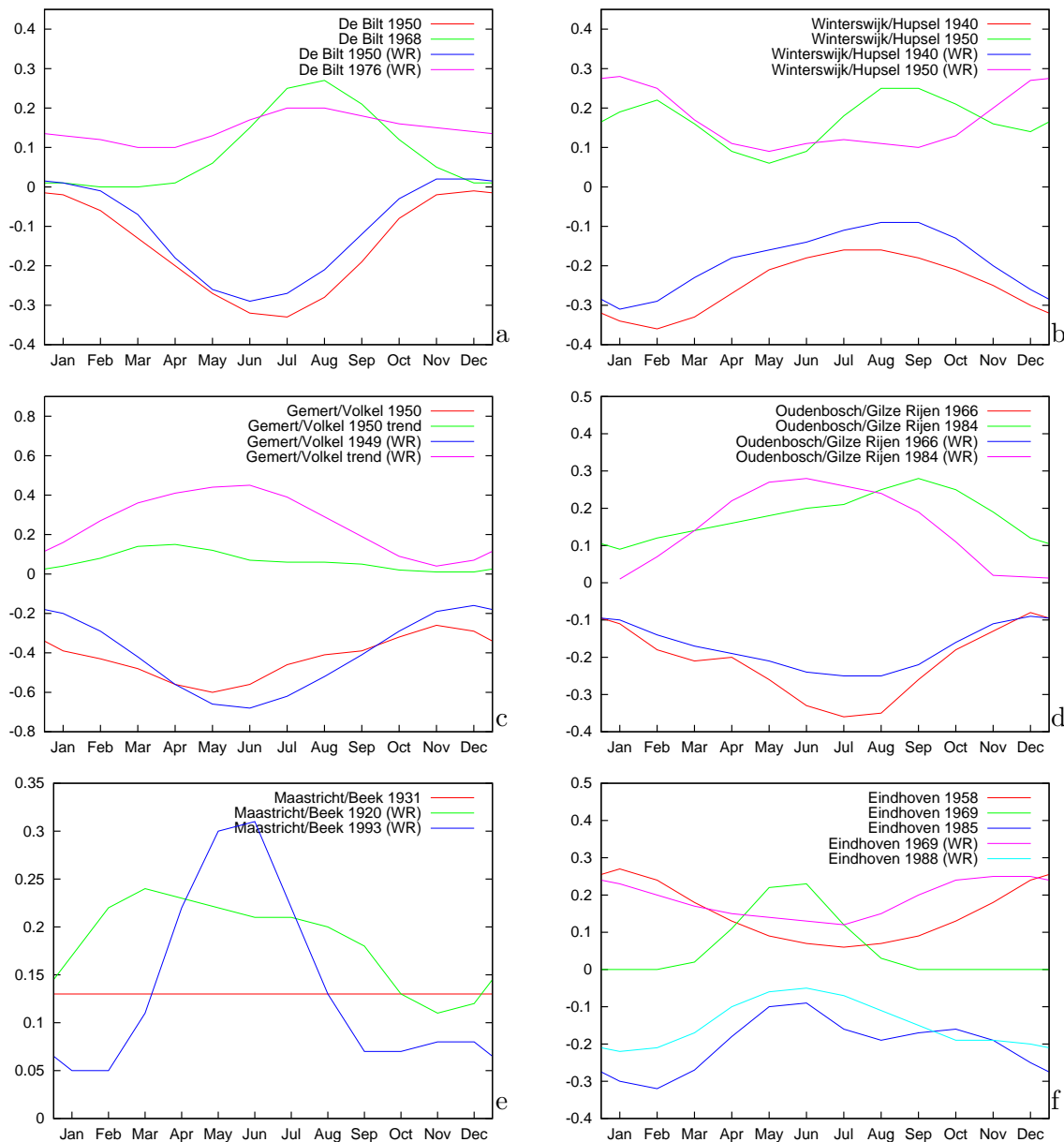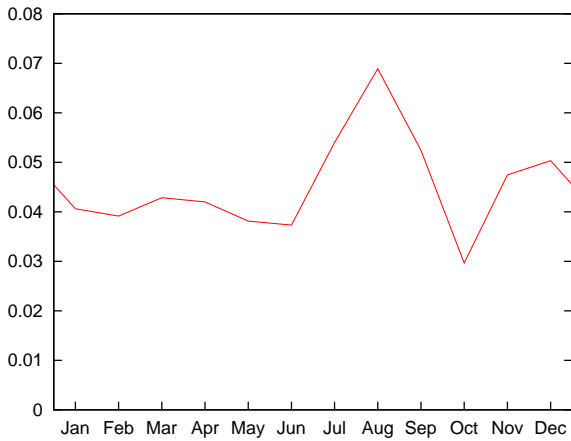
Fig. 5. Break and trend corrections for various stations for this study and those of van Ulden et al. (2009), the latter marked with (WR). Shown are corrections for Winterswijk/Hupsel **(a)**, Oudenbosch/Gilze Rijen **(b)**, Gemert/Volkel **(c)** and Eindhoven **(d)**, Maastricht/Beek **(e)** and Schiphol **(f)**.

not used to construct an estimate of the global mean temperature. The GISTEMP datasets use the GHCN De Bilt time series with an inhomogeneity in 1950 of about 1.5 K. This inhomogeneity is caused by a combination of the inhomogeneities around 1950 discussed in Sect. A1 and a change in observational practice. Prior to 1950, three daily observations and the minimum and maximum temperatures were recorded next to the use of a thermograph at the main stations. From the thermograph and the five daily observations, hourly estimates of the temperature at De Bilt (and the other

main stations) were made. These hourly estimates are used in the current study. However, only the three daily measurement (excluding the minimum and maximum temperatures) were communicated to international databases. After 1950, 24 hourly observations were communicated. Daily temperatures in the GHCN De Bilt record prior to 1950 simply averages the three measurement without application of the estimate of the daily mean temperature of Eq. (2). This inhomogeneity is reflected in the GISTEMP 250 km dataset, which uses only stations within a 250 km radius of each grid box.

Table 5. Trends (°C/10 yr) in CNT1.0 and CNT1.1 for winter (DJF), spring (MAM), summer (JJA) and autumn (SON), calculated over the periods 1906–1908, 1950–1908 and 1975–1908.

	1906–1908		1950–1908		1975–1908	
	CNT1.0	CNT1.1	CNT1.0	CNT1.1	CNT1.0	CNT1.1
DJF	0.111	0.114	0.334	0.332	0.544	0.540
MAM	0.126	0.130	0.308	0.314	0.655	0.662
JJA	0.154	0.148	0.241	0.240	0.381	0.391
SON	0.127	0.123	0.149	0.149	0.284	0.277

Fig. 6. Root Mean Square error of the difference between CNT1.1 and CNT1.0.

Table 6. Trends in °C yr^{-1} of the time series shown in Fig. 8 over the periods 1906–2010 and 1975–2010.

	1906–2010	1975–2010
CNT v1.1	0.013 ± 0.002	0.039 ± 0.011
CNT v1.0	0.013 ± 0.002	0.038 ± 0.011
De Bilt observed	0.013 ± 0.002	0.049 ± 0.011
De Bilt GHCN v2	-0.001 ± 0.003	0.049 ± 0.011
CRUTEM3	0.009 ± 0.002	0.036 ± 0.010
NOAA/NCDC	0.010 ± 0.002	0.052 ± 0.011
NASA/GISS	0.011 ± 0.002	0.038 ± 0.010
NASA/GISS 250 km	0.013 ± 0.002	0.042 ± 0.011

Fig. 7. Root Mean Square error of the difference between the homogenized stations records which construct CNT1.1 and CNT1.0.
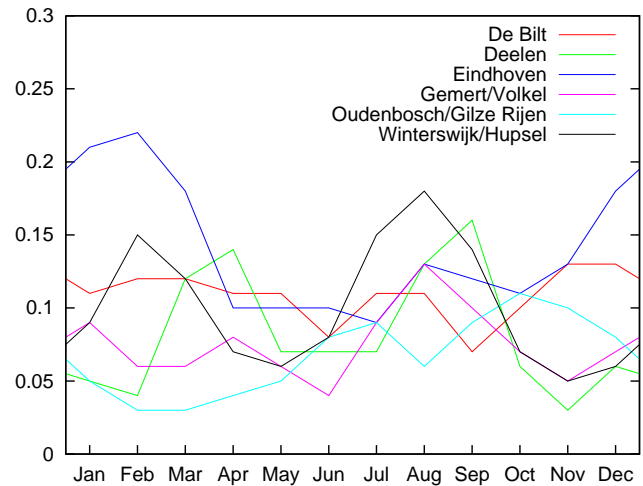
However, in the 1200 km dataset used for global temperature estimates, the inhomogeneity is not visible due to the large number of stations used.

The CRUTEM3 data shows a discontinuity when compared with the CNT1.1 in 1951 (van Ulden, 2008). This break is probably partially related to the use of the Groningen and Eelde records as one continuous record in the CRU data without corrections made for the relocation as done in Sect. 2 and Fig. 2. Similarly, the Maastricht-Beek transition is uncorrected for. Both these transitions occurred in January 1946 and overlapped in the periods 1946–1951 and 1946–1952 respectively. Furthermore, the relocations and change of thermometer screen at the De Bilt site around this time (discussed in Sect. A1) may have added to this break.

In Table 6 linear trends over 1906–2010 and 1975–2010 are shown for all the series in Fig. 8. The De Bilt observed temperature is seen to have a larger trend over the last 36 years than the CNT, the difference is due to the inhomogeneity in 1968/1969 (discussed in Sect. A1). The unadjusted GHCN v2 time series shows no trend over 1906–2010 due to the 1.5 K inhomogeneity in 1950. Although the inhomogeneity is retained in the NASA/GISS 250 km dataset, the trend is not affected due to the trend adjustment used (Hansen et al., 2010). Finally, it is unknown why the NCDC dataset shows a higher trend over 1975–2010. This may be due to

the western grid box (50–55° N, 0–5° E), which includes stations in Belgium and East Anglia (UK).

8 Conclusions

Climate models compute meteorological variables at a typical scale of 100 km. Local effects caused by vegetation, small lakes or small variations in altitude are not resolved by the models. In order to make a sensible comparison between model output and observations, the latter need to be defined at a spatial scale similar to the model results. The Central Netherlands Temperature record (CNT) has been designed to meet this demand. Additionally, the CNT is expected to be of interest for climate research, being based on high-quality homogenized records and representative for a larger area.

The CNT is based on a selection of homogeneous monthly averaged records for daily mean temperature from the KNMI network. Long records have been constructed by blending data from nearby stations, using the overlap period to
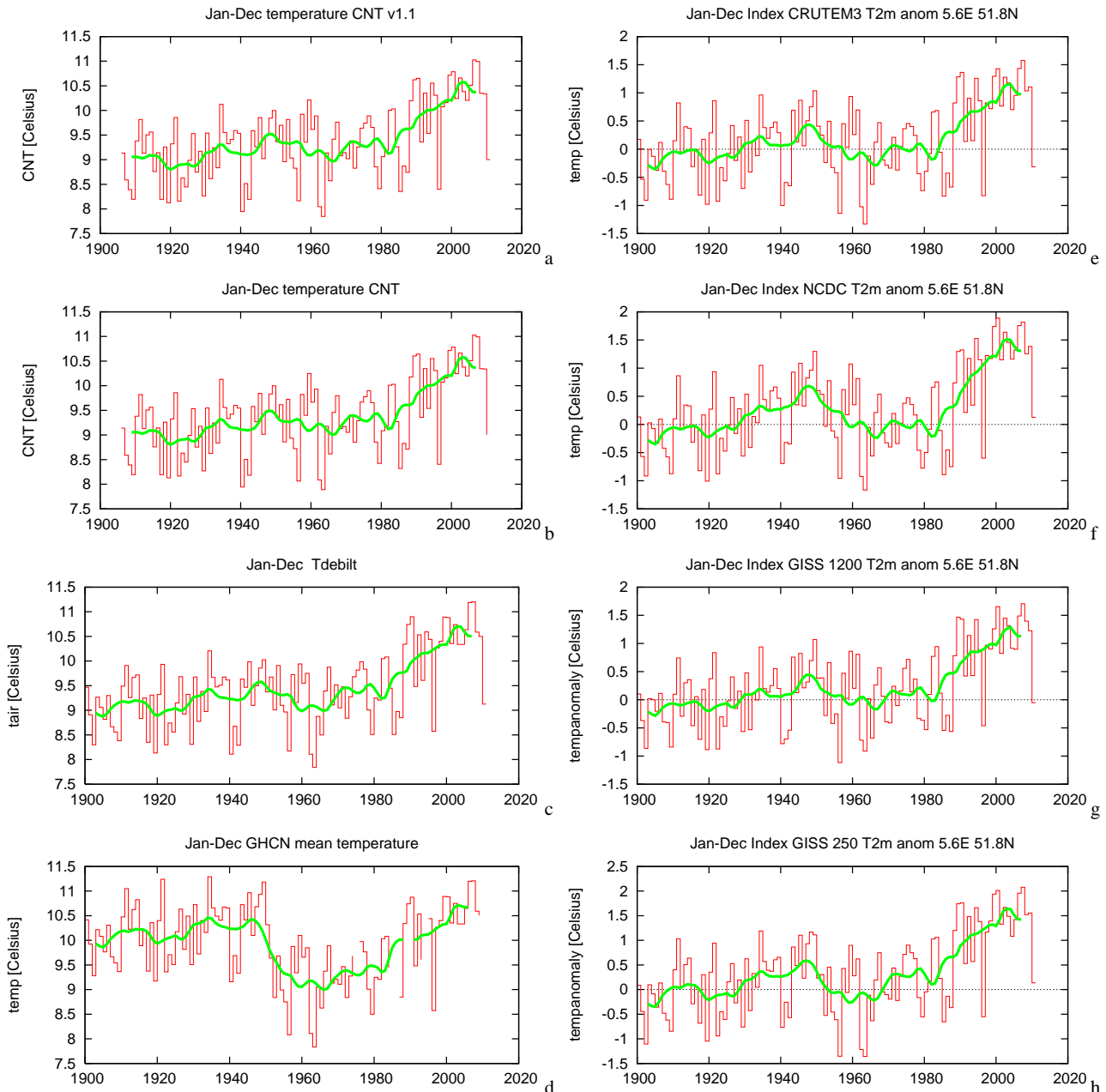
Fig. 8. Various measures of the temperature in the Central Netherlands. **(a)** CNT1.1, **(b)** CNT1.0, **(c)** observed De Bilt temperature. **(e–g)** The point 51.8° N, 5.6° E bilinearly interpolated in land temperature datasets used to construct global mean temperature estimates. **(e)** CRUTEM3, **(f)** NOAA/NCDC, **(g)** NASA/GISS. For reference, the GHCN v2 unadjusted series for De Bilt is shown in **(d)** and the 250 km GISTEMP dataset in **(h)**. The green lines denote 10-yr running means.

calibrate the differences. This resulted in nine records which start in the early 1900s. Using seven of these records in a Principal Component Analysis, a weighted average of these seven series is obtained which contains a large fraction of the common variability of the series. This time series contains the warming trend common to all time series and, due to the averaging of all available long records, inhomogeneities in the individual records are damped. The weighted average is used as reference series to homogenize the available records.

Based on an assessment of the noise levels of each difference record, the location of the record and whether or not the station is still operational, a selection of four records is made which span the period from 1906 onwards. Two additional records are included from 1951 and 1958 onwards.

The CNT is constructed as an unweighted average of these records.

The CNT series presented in this study, to which we attach the version number 1.1, shows only small differences to an earlier version of the CNT put forward by van Ulden et al. (2009) (to which version number 1.0 is attached). The differences between CNT1.1 and CNT1.0 are modest despite the differences in approaches to reach homogeneous input series. This observation adds to the robustness of the CNT.

In a comparison between the CNT1.1 and various other measures of the daily averaged temperature for central Netherlands, we conclude that the trends in most datasets that are used to construct global mean temperatures are compatible with the trend of the Central Netherlands Temperature within about 30 %. This is despite the fact that these datasets use grid boxes that are far larger then the area associated with that of the Central Netherlands Temperature and that these datasets are based on unhomogenized time series. The global datasets have their various methods to homogenize input data, but they lack knowledge on detailed metadata records. This shows that the warming trend in these datasets is robust, but also points to the need to base these datasets on homogenized time series of temperature.

The CNT series and the ten homogenized station series are updated monthly and can be downloaded from the KNMI web site, http://www.knmi.nl.

Appendix A

Detected breaks and trends

In this section, the available metadata and the results from the statistical tests described in Sect. 3 are combined. In the metadata, there are many details like routine maintenance and replacements of thermometers. These routine changes are made about once a year, which makes it unlikely that a step change or a discontinuous trend is related to a replaced thermometer. However, it shows that the detail of reporting for these stations is high, when relatively minor maintenance is added to the metadata records. Consistent and detailed reporting for a station is difficult to combine with vast changes in the surroundings (leading to the break or discontinuous trend) which remained unnoticed. Based on this argument, we were more inclined to dismiss statistical indications of a break in a record for stations that have detailed metadata records when no metadata is supportive of a break. For stations where metadata is scarce, the absence of supporting metadata for the existence of a break was less prohibitive for making adjustments to the record.

A1   De Bilt

The tests indicate breaks in annual averaged values near 1918–1920, near 1950, near 1968–1969 and near 1976.

There is no information from the metadata which could explain the breaks near 1918–1920 and 1976. The break at 1950 is clearly related with a string of modifications which happened in the early 1950s. On 17 May 1950, the large thermometer screen (the so-called "Pagoda") was replaced by a Stevenson screen. On 16 September 1950, a relocation westwards had taken place, subsequently followed by a 300 m relocation southwards on 27 August 1951.

The break at 1968–1969 is most probably related to the uprooting of the nearest part of the neighbouring orchard (which was turned into a parking lot) at the end of March 1967 and/or the construction of a new 25 m high KNMI office which started in November 1967 and had reached its highest point on 2 January 1969.

The only metadata which could possible relate to the detected break at 1919–1920 are routine replacements of thermometers. The detected break near 1918–1920 is left uncorrected due to the absence of metadata indicating possible causes for this jump.

After correcting for the 1950 and 1968 jumps, a new run through the program did detect the possible inhomogeneity at 1976 but it failed to exceed the 95 % significance level.

Brandsma (KNMI, personal communication, 2010) homogenized the De Bilt record using more physical methods rather than the statistical approach used here. He corrected for the replacement of the screen to the Stevenson screen and its transition (September 1950), a relocation of the Stevenson screen (August 1951), the lowering of Stevenson screen from 2.2 m to 1.5 m (June 1961) and the transition of the artificial ventilated Stevenson screen to the KNMI round-plated screen (June 1993). Finally, Brandsma corrected for a warming trend of 0.11 °C per century caused by urban warming. The correction factors Brandsma used for the changes around 1950 and the correction factors used in this study are shown in Fig. A1 and are very similar. However, the use of smoothed adjustment factors in this study rather than the original, more noisy monthly break-values, makes the adjustments for the De Bilt record associated with this break more conservative than those by Brandsma. The additional correction for the break around 1968 and the lack of a correction of a possible warming trend makes the two records different. The RMS difference between annual values of the De Bilt record obtained using the adjustments detected statistically in this study and the one homogenized using Brandsma's method is 0.1 °C.

A2   Den Helder/De Kooy

No inhomogeneities have been detected.

A3   Groningen/Eelde

No inhomogeneities have been detected in the annually averaged values. However, evaluating the months separately gives possible inhomogeneities in the years 1952–1953, 1973
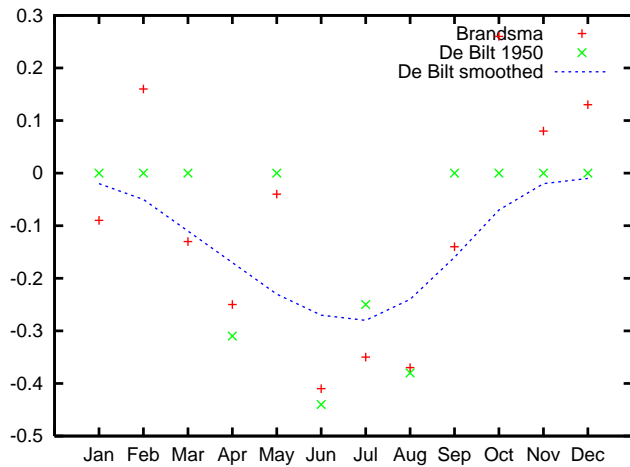
Fig. A1. Breaks at De Bilt for the changes around 1950. Shown are the break amplitudes as determined by Brandsma based on a physical methodology (red) and based on the current statistical approach (green) and the smoothed curve based on the latter approach [°C].

and 1996. The 1952–1953 break has an amplitude of a roughly equal size for July and December, but of opposite sign. The metadata indicate that on 28 February 1952 the thermograph was corrected with 1 °C and that it was replaced on 2 June 1953. The evidence from the metadata is judged too scanty to justify a correction for this break.

On 1 May 1973 the measurement field was relocated to the west side of the runway. This move coincides with changing measurement practice from manual to the use of electronic equipment. The Groningen/Eelde series is adjusted for this break.

No metadata indicating a possible cause for the 1996 break could be found. Apparently, some suspicion at the KNMI staff of the time must have existed, since on 4 October 1996, a comparison is made between the temperature sensors and a calibrated sensor. No deviations were found though. The series is not corrected for this detected break.

A4 Oudenbosch/Gilze Rijen

Breaks in the Oudenbosch/Gilze Rijen series are detected around 1946–1948, 1966–1967, 1971 and near 1984.

The metadata indicates that corrections on the minimum and maximum recording thermometers changed frequently and significantly around the 1946–1947 period. Breaks in the minimum and maximum temperatures could result in breaks in the daily averaged temperature since Oudenbosch is one of the stations where Eq. (2) is used. However, the changes in corrections to the min. and max. temperatures affected temperatures below freezing mostly and amounted to approx. 0.1 °C to 0.2 °C. Application of models Eqs. (3) and (5) indicate statistically significant breaks in summer and autumn. Based on incongruous evidence from the metadata and the tests, we leave this break uncorrected.

The break around 1966–1967 is most likely associated with two changes: the construction of a paved road on the southeast side of the terrain and the related uprooting of high trees which made the surroundings more open and the move of the instrument field 1400 m southwards, further away from trees in a more exposed setting in July 1966. The series is adjusted for this break.

The 1971–1972 break has only a detectable break for March, August and September, not in the other months and not in the annual mean values. The only metadata from around this year was the reinstallment of the thermometer screen following 1971 "new style" specifications at 5 October 1970. These reinstallments were made at all stations around this period, making it unlikely that only this station suffers from adverse effects. Furthermore, KNMI noted in their logs at 14 October 1970 that a garage had been erected in the vicinity of the location. In the log the comment is added "However, this garage has no effects on the measurements". It is not clear on what analysis this conclusion is based. No corrections are made for this break.

The metadata for the years around 1984 only indicate routine maintenance and replacements of the instruments, the most profound being a replacement of the thermograph on 19 March 1983 due to a bended pen arm and adjustments to the thermograph on 6 October 1983 and 29 November 1984 by 1.0 °C and 0.5 °C respectively. Thermographs were replaced approximately once a year, which makes it improbable that these defects relate to the observed breaks in the record. However, this latter break was large and robust enough to justify adjustment in the absence of metadata.

A5 Winterswijk/Hupsel

Large breaks in the Winterswijk/Hupsel record are detected in 1940 and 1950. Much smaller breaks are detected near 1960 and 1970–1972. The break in 1940 is possibly related to a relocation in 12 March 1940 to a more ideally located site. The new site is open, facing the observer's house to the north and a meadow to the south, but the thermometer screen is placed between two shrubs. On 27 February 1950, the thermometer screen was relocated 5 m eastwards away from the shrubs. The growth of the shrubs might have introduced an artificial trend in the data. The annually averaged difference record does show a trend over this 10-year period, but it is small (approx. 0.18 °C in 10 years) and has been left uncorrected.

The breaks near 1960 and 1970–1972 have been left uncorrected due to absence of metadata which could be related to the break.

Interestingly, a break is detected near 1984–1985. Based on the F-statistic, this break is not significant and is not corrected for. However, it seems to be related to a modest relocation of the station some 50 m in SW direction on 27 March 1985.

A6 Gemert/Volkel

A large break and a discontinuous trend are detected around 1950. The break and discontinuous trend are obvious from a visual inspection of the difference record (Fig. A2). This is clearly associated with the reinstallment of the station on 27 September 1949. Preceding this reinstallment, reports had been made (in June and July of 1949) indicating that the site did not meet regulations regarding the surroundings. The clearing was too small for proper ventilation. Additionally, the height of the thermometer screen was not according to regulations.

Having identified the combination of a step and a discontinuous trend for 1950, estimates of the adjustments from the other model, Eq. (7), are to be used. It turns out that the estimates of the size of the trend for the 1950 break, calculated for each sliding window, show much variability. This is probably related to the relatively high year-to-year variability of the difference series in relation to the modest trend. Because of this, the few estimates ($<10\%$) of a negative trend were not used in the final estimate of the trend, nor were estimates used which showed trends of $>0.5\,°C/10$ year (7 instances). The record is corrected for the break and the discontinuous trend.

There is some discussion on the validity of the measurements from Volkel airbase. When the transition between the Volkel and Gemert records is set at 1980, then the homogenisation procedure detects a break and discontinues trend at this year, with a negative trend in the difference series after 1980. No break nor discontinuous trend is detected when the transition between the series is set at 1990. The poor quality of the measurements from the 1980s of the Volkel air base (not shown) may be related to the spurious trend. The introduction of automated measurements in the early 1990s will have improved the quality of the data. With the nearest tree line at 245 m of the thermometer screen in N–NW direction, with trees of heights between 18–20 m, the situation at the observation site is in line with the WMO regulations for a meteorological observation site.

No corrections are made to the Volkel record which is blended to the Gemert record from 1990 onwards.

A7 Maastricht/Beek

The Maastricht/Beek record shows a break in 1931 in the annually averaged values, but fails to show significant values in an analysis for each month separately. The Maastricht observations were made on top of a tower (at approx. 20 m above street level) on a building in the centre of the city. From 1 July 1951 onwards, parallel measurements were made at the outskirts of the city in a garden area which show considerable differences to the Maastricht observations. The suboptimal setting of the Maastricht station and the relocation to a new site some 65 m higher in altitude may be the principal
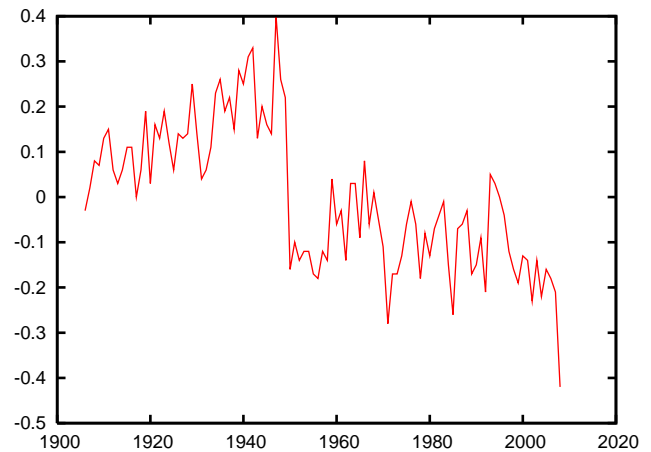
Fig. A2. Trends and breaks in the annual mean temperature difference between the Gemert/Volkel series and the reference series [°C].

reasons for having the largest adjustment factors associated with the move to a more ideally located setting (Fig. 2).

A8 Twenthe

Twenthe shows a break in 1969 in the annually averaged values, but significant values fail to show in an analysis for each month separately. Twenthe is a military air base and no metadata exists which might substantiate this break. No adjustments have been made to this record.

A9 Hoorn

Hoorn shows a break which is barely significant in 1948. Initially, the observation site was located on a terrain for agricultural use. On 1 November 1946, the observation terrain was relocated to the gardens of the local slaughterhouse, facing nearby buildings in southeast to southwest directions. On 21 November 1947, measurements ceased and on 28 April 1948, the station was relocated back to its original terrain. No adjustments are made for this break.

Another break is detected around 1970–1973, but only for the month of March. The metadata indicate that a school was built at a distance of approximately 20 m from the observation site in the early 1970s and a relocation in SW direction of 15 m was effectuated on 2 July 1971, but it is unclear if this rather modest relocation could explain the break to warmer conditions. No correction for this break have been made to the Hoorn record.

A10 Schiphol

The Schiphol (Amsterdam International Airport) record shows a significant break in 1981. The Schiphol metadata is not very clear about the cause for this break. Presumably it is related to a relocation of the measurement field from

the vicinity of the main buildings to the outer edge of the Schiphol area, near a runway. The KNMI archives holds a map for the situation around 1960 and one for 1986 from which a change in position of the measurement field is evident but a more precise timing of the relocation cannot be given. Given the rapid growth of the airport, this period is likely to have seen more than one relocation of the instrumentation. The record is adjusted for this break.

A11 Deelen

In the period 1954–1957, measurements were taken during weekdays only at Deelen airbase, reliable monthly means could only be constructed for January 1958 onwards. Breaks have been detected near 1962 and around 1984–1985. However, the metadata from this military airport provided no leads as to what might have caused these breaks. The breaks are large enough to convincingly exceed the critical significance levels and have been adjusted.

In the annual averaged values for Deelen and to a lesser extent in the February monthly means, a discontinuous trend was detected with a break in 1977. Before 1977, a distinct upwards trend was detected indicating that Deelen warms faster than the reference record, after that year the warming trends are similar. Again, there is no indication as to what might be the reason for the discontinuous trend and it is left unadjusted.

A12 Eindhoven

A curiously low value for annual averaged temperature for Eindhoven airport was observed for 1952, which is 0.95 °C lower than the reference. A comparison with adjusted records for Oudenbosch/Gilze Rijen and Gemert/Volkel indicated that the monthly averages for the month May to July were up to 3 °C lower than surrounding stations. The monthly averages of 1952 for these months were replaced by an average of the corresponding months of the adjusted records of Oudenbosch/Gilze Rijen and Gemert/Volkel. The metadata indicated that observations from 1 May 1952 onwards were made by airforce personnel rather than civil servants from the aviation authority, and reports of KNMI inspectors of the mid-1950s complained of many false readings.

Breaks were detected at 1969–1970 and 1986–1988. The exact timing of the latter break is vague, non-significant breaks were also reported for 1985, but strangely enough, none for 1987. The metadata provided no information on the possible origin of the first break. The relocation to a new terrain on 3 July 1984 may be related to the latter break.

There is some indication of a break in the month of May only, around 1980–1981; only 3 sliding windows indicated this break. No metadata had been found which might account for this break and it is left uncorrected.

After corrections for the 1969–1970 and 1986–1988 breaks, the Eindhoven series was put through the break detection script again and this yielded a break in 1980–1981 and a newly detected break in 1958, which was apparently not large enough (in terms of the F-statistic) to be reported in the uncorrected series. This break must be related to a station relocation at 17 October 1958. Before this date, the site did not meet KNMI specifications. This break is corrected for.

A13 Rotterdam

No breaks have been detected at Rotterdam.

References

Alexandersson, H. and Moberg, A.: Homogenization of Swedish temperature data, part I: homogeneity test for linear trends, Int. J. Climatol., 17, 25–34, doi:10.1002/joc.1097, 1997.

Brandsma, T. and van der Meulen, J. P.: Thermometer screen intercomparison in De Bilt (the Netherlands) – Part II: Description and modeling of mean temperature differences and extremes, Int. J. Climatol., 28, 389–400, doi:10.1002/joc.1524, 2008.

Easterling, D. R. and Peterson, T. C.: A new method for detecting undocumented discontinuities in climatological time series, Int. J. Climatol., 15, 369–377, doi:10.1002/joc.3370150403, 1995.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change, Rev. Geophys., 48, RG4004, doi:10.1029/2010RG000345, 2010.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded daat set of surface temperature and precipitation for 1950–2006, J. Geophys. Res.-Atmos., 113, D20119, doi:10.1029/2008JD010201, 2008.

Jarusková, D.: Change-Point Detection in Meteorological Measurement, Mon. Weather Rev., 124, 1535–1543, 1996.

Labrijn, A.: The climate of the Netherlands during the last two and a half centuries, Scientific Rep. KNMI No. 102, Royal Netherlands Meteorological Institute, mededelingen en Verhandelingen, 1945.

Lund, R. and Reeves, J.: Detection of Undocumented Change-points: A Revision of the Two-Phase Regression Model, J. Climate, 15, 2547–2554, 2002.

Peterson, T. C. and Easterling, D. R.: Creation of homogeneous composite climatological reference series, Int. J. Climatol., 14, 671–680, 1994.

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Toumenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Forland, E., Hanssen-Bauer, I., Alexandersson, H., Jones, P. D., and Parker, D.: Homogeneity adjustments of *in situ* atmospheric climate data: A review, Int. J. Climatol., 18, 1493–1517, 1998.

Potter, K. W., Illustration of a New Test for Detecting a Shift in Mean in Precipitation Series, *Mon. Wea. Rev.*, *109*, 2040–2045, 1981.

Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q.: A Review and Comparison of Changepoint Detection Techniques for Climate Data, J. Appl. Meteorol. Clim., 46, 900–915, 2007.

van der Hoeven, P. C. T.: Etmaaltemperatuur en Dagextremen uit Termijnwaarnemingen, memorandum KD 92-08, KNMI, 1992.

van Ulden, A.: The Construction of a Central Netherlands temperature, EMS8/ECAC7 Abstracts, Vol. 5, EMS2008-A-00421, 8th Annual Meeting of the EMS/7th ECAC, 2008.

van Ulden, A., van Oldenborgh, G. J., and van der Schrier, G., The Construction of a Central Netherlands temperature, Scientific Rep. WR 2009-03, KNMI, 2009.

Vincent, L. A.: A Technique for the Identification of Inhomogeneities in Canadian Temperature Series, J. Climate, 11, 1094–1104, 1998.

Wang, X. L.: Comments on "Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model", J. Climate, 16, 3383–3385, 2003.

Wilks, D. S.: Statistical methods in the Atmospheric Sciences, Academic Press, 1995.