Homogenization of daily temperature series in the European Climate Assessment & Dataset

Squintu, A. A., van der Schrier, G., Brugnara, Y., & Klein Tank, A.

Squintu, A. A., van der Schrier, G., Brugnara, Y., & Klein Tank, A. (2019). Homogenization of daily temperature series in the European Climate Assessment & Dataset. International Journal of Climatology, 39(3), 1243-1261. [5874]. https://doi.org/10.1002/joc.5874

Received: 29 May 2018 | Revised: 27 September 2018 | Accepted: 28 September 2018 | Published on: 15 November 2018

DOI: 10.1002/joc.5874

International Journal of Climatology RMetS

RESEARCH ARTICLE

Homogenization of daily temperature series in the European Climate Assessment & Dataset

Antonello A. Squintu[1,2] | Gerard van der Schrier[1] | Yuri Brugnara[3] | Albert Klein Tank[1,2]

[1]Koninklijk Nederlands Meteorologisch Instituut, De Bilt, RDW&D, Netherlands

[2]Department of Meteorology and Air Quality, University of Wageningen, Wageningen, The Netherlands

[3]Institute of Geography, University of Bern, Bern, Switzerland

Correspondence
Antonello A. Squintu, Koninklijk Nederlands Meteorologisch Instituut, RDW&D, Utrechtseweg 297, 3731 GA De Bilt, Netherlands.
Email: squintu@knmi.nl

Funding information
Horizon 2020 Framework Programme, Grant/Award Number: 640171; Seventh Framework Programme, Grant/Award Number: 607193

The daily maximum and minimum temperature series of the European Climate Assessment & Dataset are homogenized using the quantile matching approach. As the dataset is large and the detail of metadata is generally missing, an automated method locates breaks in the series based on a comparison with surrounding series and applies adjustments which are estimated using homogeneous segments of surrounding series as reference. A total of 6,500 series have been processed and after removing duplicates and short series, about 2,100 series have been adjusted. Finally, the effect of the homogenization of daily maximum and minimum temperature on trend estimation is shown to produce a much more spatially homogeneous and then plausible picture.

KEYWORDS
Europe, homogenization, quantile matching, temperature, trends

1 | INTRODUCTION

Modifications to meteorological stations, such as relocation, replacement of the instrument, recalibration, new buildings in the neighbourhood or growth of vegetation in the proximity, alter temperature measurements and introduce biases in the observational records that do not relate to weather and climate (Aguilar *et al.,* 2003; Hartmann *et al.,* 2013). The analysis of climatic variability and climatic change requires homogeneous temperature series (Peterson *et al.,* 1998): these series do not confuse the climatic signal with artificial biases which are present in non-homogenous series (Begert *et al.,* 2005; Thorne *et al.,* 2005; Brunetti *et al.,* 2006; Menne and Williams Jr, 2009). Prior to climate analyses, actions are required to aim at the removal of step-like or gradual changes related to these non-climatic effects in observational records (Caussinus and Mestre, 2004).

The registration of activities on meteorological stations in the metadata keeps track of these changes and sufficiently detailed metadata allow a precise temporal localization of the breaks. Unfortunately, the availability of metadata is often low, especially further back in time, and does not cover the whole set of inhomogenities that affect the measurements (Caussinus and Mestre, 2004). This implies that break-detection based on metadata only is not possible for many datasets, even though this approach is regarded as most accurate and reliable. This argument, and the sheer size of a dataset, motivates the use of an automated homogenization procedure (Caussinus and Mestre, 2004).

The aim of this study is to develop a pan-European homogeneous dataset of daily maximum and minimum temperature using such an automated homogenization procedure. It will use a recent agreement-based system to detect breaks (Kuglitsch *et al.,* 2012) and the quantile matching technique (Trewin, 2013) in combination with a pairwise-comparison (Menne and Williams Jr, 2005) approach to determine adjustments. The elements in this approach are introduced below.

Automated homogenization procedures consists of two steps: break detection and adjustment calculation (which follow, or are integrated with, a quality check procedure) (Alexandersson, 1986; Caussinus and Mestre, 2004). These have been focusing mainly on the detection of breaks in the monthly, seasonal, or annual values and use statistical tests accompanied with penalizing functions (Alexandersson, 1986; Caussinus and Mestre, 2004; Menne and Williams Jr, 2005; Wang *et al.,* 2007) or inspections on autocorrelation of residuals (Vincent, 1998). Recent comparisons (Domonkos, 2013; Lindau and Venema, 2013; Venema *et al.,* 2013) have pointed out advantages and drawbacks of the most common systems. Procedures that look for an agreement among methods (e.g., Kuglitsch *et al.,* 2012) go one step further and take benefits from the reduced uncertainty in break location by looking for consensus.

Homogenization of annual or monthly averages does not automatically imply a homogenization of higher-order moments (Trewin, 2013) since the processes that generates inhomogeneities on daily datasets are nonlinear, that is, introduced inhomogeneities to the temperature measurements depend, not in a linear way, on the temperature itself (Della-Marta and Wanner, 2006) and external factors as cloud cover, wind strength and direction can modify extreme daily values differently than the averaged conditions (Brandsma and Van der Meulen, 2008).

Some methods homogenize daily records simply by interpolating monthly adjustment to a daily resolution via a polynomial (Vincent *et al.,* 2002) or trigonometric regressions (Brunetti *et al.,* 2006). While this approach assures that daily adjusted values reflect the same temporal behaviour as those observed in the monthly series (Vincent *et al.,* 2002), adjustments for the higher-order moments are not guaranteed (Mestre *et al.,* 2011). A more advanced set of methods considers the temperature distribution which, split into quantile bins, is compared with expected values obtained from surrounding stations. A nonlinear regression or cubic smoothing splines (Mestre *et al.,* 2011) are used for the calculation of the correcting factors. Finally, a method not based on model parameterization or regressions is the quantile matching, which compares quantiles of the distributions of measurements before and after the break and calculates adjustments by requiring similarity between these distributions.

Ideally, adjustments are made by comparing measurements from the original and the disturbed situation for overlapping periods (World Meteorological Organization, 2011). The difference between these records eliminates the background climatic signal and highlights the effects of, for example, the change in location. When such parallel measurements are not available, the most reliable source of information about the climate background is the net of neighbouring stations being exposed to the same climatic conditions as the target series (Aguilar *et al.,* 2003; Menne and Williams Jr, 2005; Della-Marta and Wanner, 2006; Venema *et al.,* 2013).

Approaches to construct the reference series are weighted (or simple) averages of surrounding recorded anomalies (Alexandersson and Moberg, 1997; Vincent *et al.,* 2002; Begert *et al.,* 2005; Štěpánek *et al.,* 2009), using a high-correlated homogeneous series (Della-Marta and Wanner, 2006) or performing the *pairwise comparison* (Menne and Williams Jr, 2005; Trewin, 2013). The calculation of an averaged series may incorporate the inhomogeneities of neighbouring series into the reference series (Menne and Williams Jr, 2005; Della-Marta and Wanner, 2006), might get misleading features from uncorrelated series and can be affected by the change in the number of contributing series, introducing strong changes of mean and variance in the reference (Brunetti *et al.,* 2006), thus compromising the ability to represent statistical features of the climate background (Caussinus and Mestre, 2004). On the other hand, while the use of single series allows the analysis to be independent from the changes in data availability, this approach is risky since it relies totally on a series whose quality might not be certain and whose climatic features might not be consistent with the target series. Isolation of the artificial signal with pairwise comparison, where each reference series provides an estimate of the adjustment of the target series, is shown to be more robust at detection undocumented changes (Menne and Williams Jr, 2009) and more reliable for the calculation of estimates of the adjusting factors (Trewin, 2013).

The study is organized as follows: Section 2 introduces the dataset and the methods, Section 3 shows the results, a few case studies and the effects of the homogenization on trends in temperature. The study is discussed and concluded in Section 4.

2 | DATA AND METHODS

2.1 | European Climate Assessment & Dataset

The European Climate Assessment & Dataset (ECA& D, Klein Tank *et al.,* 2002; Klok and Klein Tank, 2008) is a collection of daily station observations of currently 12 elements and contains at the time of writing (July 2018) data from nearly 11,100 European stations (more than 7,500 with temperature measurements) and is gradually expanding. ECA&D contains more than 200 temperature series starting before 1900, but a strong increase in the number of series is found in the 1950s. Data from the station network at ECA&D is updated on a monthly basis using data kindly provided directly by the National Meteorological and Hydrological Services (NMHSs), individual researchers affiliated with a university, global data centres like the National Centers for Environmental Information (Asheville, USA) or the synoptic messages from the NMHSs delivered through the Global Telecommunication System (World Meteorological Organization, 2007).

Data coverage varies depending on the countries and on the time coverage (Figure 1). Early series (start before 1890) are well distributed in western Europe with the exception of southern Italy and northern Scandinavia. Further data-rescue work is in progress for the improvement of data coverage in low density areas and periods.

The quality check procedure in ECA&D is documented elsewhere ECA&D (Project Team, 2012) and at the time of writing simply insists on consistency between maximum and minimum temperature, does not allow more than five repeating values and flags data when the difference from the climatological value exceeds five times the standard deviation. A more sophisticated quality check procedures for ECA&D is currently developed, as an evolution of (Štěpánek *et al.,* 2009), based on spatial consistency of measurements in cooperation with Global Change Research Institute (Brno, CZ).

2.2 | Break detection

The detection of breaks is done using a completely automated procedure which is blind to metadata. The break detection method is inspired by the approach of Kuglitsch *et al*. (2012), which seeks agreement (i.e., common detected timing of breaks) of two out of three common break detection methods (Prodige: Caussinus and Mestre, 2004, RHtest: Wang *et al.,* 2007 and GAHMDI: Toreti *et al.,* 2012). The main difference with the Kuglitsch *et al*. (2012) approach is that both the selection of reference series and the combination of the three detection methods have been automated. Both are performed separately on annual and winter/summer half means of standardized differences between candidate and a maximum of eight reference series, selected based on completeness, correlation of annual averages (minimum of 0.6) and distance (maximum of 1,000 km). At least three reference series must confirm a breakpoint in any of the temporal aggregations in a pairwise approach.

The breakpoints are detected at annual resolution. Breakpoints detected in adjacent years by different methods, reference series, or temporal aggregations, are considered the same breakpoint.

Simultaneous changes made to the measurement networks at a national scale are difficult to detect simply because surrounding reference series will suffer from the same resulting break. For these breaks documented metadata is required.

2.3 | Calculation of adjustments: quantile matching

The adjustment of daily temperature is inspired by the work of Trewin (2013) and is based on a quantile matching algorithm which compares the probability density distribution of temperature before and after the considered break, not taking into account metadata. By making use of a set of homogeneous references series, the climatic signal is accounted for and the assumption is made that the difference series between the candidate and the reference (in their homogeneous sub-periods) is random noise.

The adjustment process targets each series individually. The break detection produces a sequence (t_1, t_2, t_3, ..., t_n) of the timing of the detected breaks in the candidate series (from the most recent to the earliest). Following these breaks, the candidate is divided into $n + 1$ sub-series:

$$\mathbf{S}_0(t|t_1<t),\quad \mathbf{S}_1(t|t_2<t<t_1),\text{etc.,} \tag{1}$$

which are homogeneous by definition (Caussinus and Mestre, 2004). These segments will be considered independently during the following steps of the process. Segments shorter than 5 years are not adjusted because of insufficient length required for a robust calculation of quantiles.

2.3.1 | Reference selection and use

The references are selected from a box of 6° centred on the candidate station and with an elevation difference smaller than 500 m. For high-elevation stations (≥1,000 m), this threshold is changed to find neighbouring stations within half the elevation of the candidate station (which increases the number of reference series for mountain stations). Among the series that fulfil these requirements, in case of densely covered areas, the set union of the 40 longest ones and the 20 earliest starting are chosen.

With the same splitting procedure used for the candidate series, the results of break detection are used to divide the reference series into homogeneous sub-series. Only the sub-series with at least 5 years of overlap with both segments of the candidate (i.e., at either side of the break) are selected. This constraint helps to avoid series that have breaks in the same period to be references to each other (e.g., in case of simultaneous breaks in a national network). Since the presence of a trend on temperatures is likely to happen, the maximum length of the sub-series used for calculating the adjustments is set to 20 years, so that the temporal evolutions affecting the moments do not alter the shape of the distribution, for example, making it broader.

For each reference sub-series, the daily raw correlation with the segment of the candidate after the break is calculated.

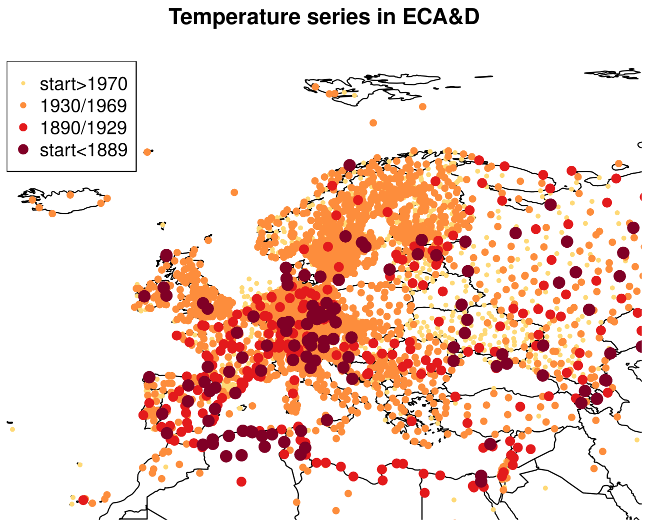

FIGURE 1 Minimum and maximum temperature series in ECA&D. Colour code indicates the start of the series [Colour figure can be viewed at wileyonlinelibrary.com]

In order to limit the computational time but simultaneously preserve statistical significance, out of the set of reference series, the 18 series with the highest correlation are chosen, provided they have correlations higher than 0.75. Note that this threshold is higher than earlier suggested (Domonkos, 2013). Figure 2 illustrates the selection of homogeneous references for a detected break in the candidate series.

In areas with a sparse network, which are common for the early periods, the number of available references may be low. In these cases, non-split (and thus inhomogeneous) series (up to 5 in total, meeting the correlation, geographical and temporal overlapping requirements) are added to the reference set, avoiding stations having a sub-segment already selected. In any case, a minimum of 3 reference series is required for the procedure to be performed; otherwise, the candidate is temporarily discarded. The larger size and density of the station network of ECA&D in comparison to the Australian dataset of Trewin (2013) makes the availability of reference series higher for the European situation.

Note that the selection of reference series for the adjustment calculation is different than the one used for the break detection.

2.3.2 | Calculation of quantile-based adjustments

Adjustment calculation has been developed taking inspiration from Trewin (2013). It is performed backwards considering the breaks from t_1 to t_n successively. The segment of the candidate series after the considered break (t_i) is termed the *basis* series (**B**), while the *segment* ($\mathbf{S}_i$, where i = 1, 2, ..., n) immediately before it is to be adjusted. For each reference $\mathbf{R}_j$, we define $\mathbf{R}_j^{\text{aft}}$ and $\mathbf{R}_j^{\text{bef}}$ as the portion of the reference after and before the break t_i.

The quantile-based adjustments for daily data are calculated on a monthly basis and applied on a daily resolution, depending on which quantile of the monthly distribution the daily data belong to. The distribution of temperatures is then considered for each month separately, introducing the seasonal cycle in the adjustments. Absolute temperatures from the target month m are considered, and values from the preceding month and the following month are gathered to reduce the noise and to make sure that a sufficient amount of data are available to determine the quantiles. This approach makes that weather types from the spring and autumn transition seasons, like March and May, contribute to determining the adjustment for homogenization of April temperatures, which is likely to be influenced both by typical March and May conditions. These temperature measurements are sorted in ascending order and, for example, the value associated to 10th quantile is calculated as the median value of all data points between the 7.5th and 12.5th quantiles. This process generates quantile sequences for data before and after the break in the target ($\mathbf{s}_{q,m}$, $\mathbf{b}_{q,m}$) and in the reference series ($\mathbf{r}^{\text{bef}}_{j,q,m}$, $\mathbf{r}^{\text{aft}}_{j,q,m}$) (thus obtaining four quantile sequences).

The adjustment for each of the quantiles is then calculated in a three-step approach. First, the difference between quantile sequences of the candidate series before and after the breaks is calculated, this difference is affected by both the artificial and the climatic signal. Second, to identify the climatic signal, the difference between quantile sequences of the reference series before and after the break is calculated. As third step, in order to isolate the artificial signal, the above differences are subtracted to each other. These steps are summarized by the following equation:

$$\mathbf{a}_{i,j,q,m} = \left(\mathbf{b}_{q,m} - \mathbf{s}_{i,q,m}\right) - \left(\mathbf{r}^{\text{aft}}_{j,q,m} - \mathbf{r}^{\text{bef}}_{j,q,m}\right). \qquad (2)$$

In order to reduce noise, the adjustments are smoothed using a simple mean of adjustments from the neighbouring months and neighbouring quantiles.

$$\bar{\mathbf{a}}_{j,q,m} = \frac{\mathbf{a}_{j,q,m} + \mathbf{a}_{j,q+5,m} + \mathbf{a}_{j,q-5,m} + \mathbf{a}_{j,q,m+1} + \mathbf{a}_{j,q,m-1}}{5}. \qquad (3)$$

Further check has been performed to avoid situations in which negative slopes of the smoothed sequences cause, after their application, changes in the rank of the data (i.e., data changing quantile after homogenization). This check has interested a very small portion of the series, more details on this can be found in Appendix A.

As mentioned above, the application of adjustments is performed considering each daily value of the series individually, depending on the location in the monthly temperature distribution.

A set of estimations of the correction is produced, each one corresponding to the different overlapping periods each reference series $\mathbf{R}_j$ has with the segments of the candidate series. The value to be corrected may belong to a different quantile in each of these overlapping periods. After

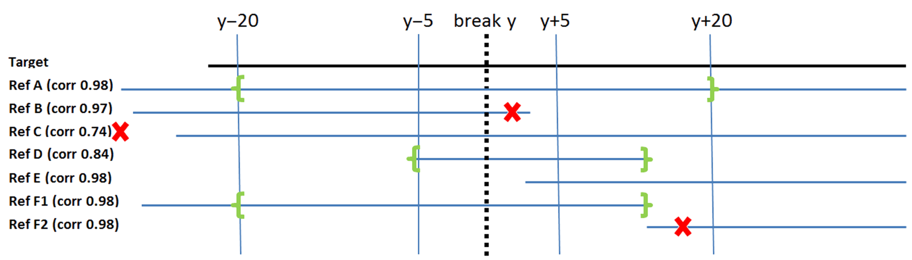

FIGURE 2 Fictional example of reference selection where the sections between curly brackets are used and sections marked with a cross are not. Reference series with correlation below 0.75 are discarded, as well as series with overlap shorter than 5 years or data recorded more than 20 years before or after the break [Colour figure can be viewed at wileyonlinelibrary.com]

determining these quantiles ($\widetilde{q}_j$) the estimation ($\widetilde{v}_j$) of the adjusted value related to $\mathbf{R}_j$ is:

$$\widetilde{v}_j = v + \mathbf{a}_{j,\widetilde{q},m}, \tag{4}$$

where v is the original value. The final adjusted value is then calculated by taking the median of the estimations:

$$\overline{v} = \text{median}\left(\widetilde{v}_j\right), \tag{5}$$

where $j = 1, \ldots, r$.

The method described above has been applied to the whole ECA&D dataset. The high number of breaks detected caused a high number of short homogeneous segments, which often were not long enough to be homogenized or had too short overlapping period with the homogeneous segments of the surrounding reference series, making it impossible to perform the quantile matching. These portions have been integrated in the temporary homogenized version of the dataset, which has undergone a second round of homogenization. In this second run the break detection and quantile matching have been launched again, so that additional adjustments were calculated.

TABLE 1 Number of series involved in the stages of the homogenization process

	TN	TX
Original series (complete set of considered series)	6,438	6,404
Homogeneous original series (original series labelled as homogeneous by the break detection)	560	670
Adjusted series, Iteration 1 (series that have been corrected during Iteration 1)	2,111	2007
Homogenized dataset, Iteration 1 (union of series that were already homogeneous and adjusted series)	2,671	2,677
Homogeneous series after Iteration 1 (union of original homogeneous series and series successfully homogenized after Iteration 1)	1,165	1,131
Adjusted series, Iteration 2 (series that have been corrected during Iteration 2)	1,526	1,571
Homogenized dataset (final) (union of series that were already homogeneous and adjusted series)	2,691	2,702
Final non-homogeneous series (according to a third run of break detection)	1,357	1,260
Final homogeneous series (according to a third run of break detection)	1,334	1,441

Homogenized series, for each iteration, consist of the sum of the adjusted series and the already homogeneous series. Low percentage of homogenized series is due to the exclusion of short and duplicate series.

3 | RESULTS

3.1 | Statistics of the adjustments

Figure 3 shows the number and timing of the detected breaks for the first and second iterations. The relatively high number of breaks detected in the original series during the first iteration resulted in a high number of short homogeneous segments. These segments, serving as references, are often not long enough to be homogenized or had a too short

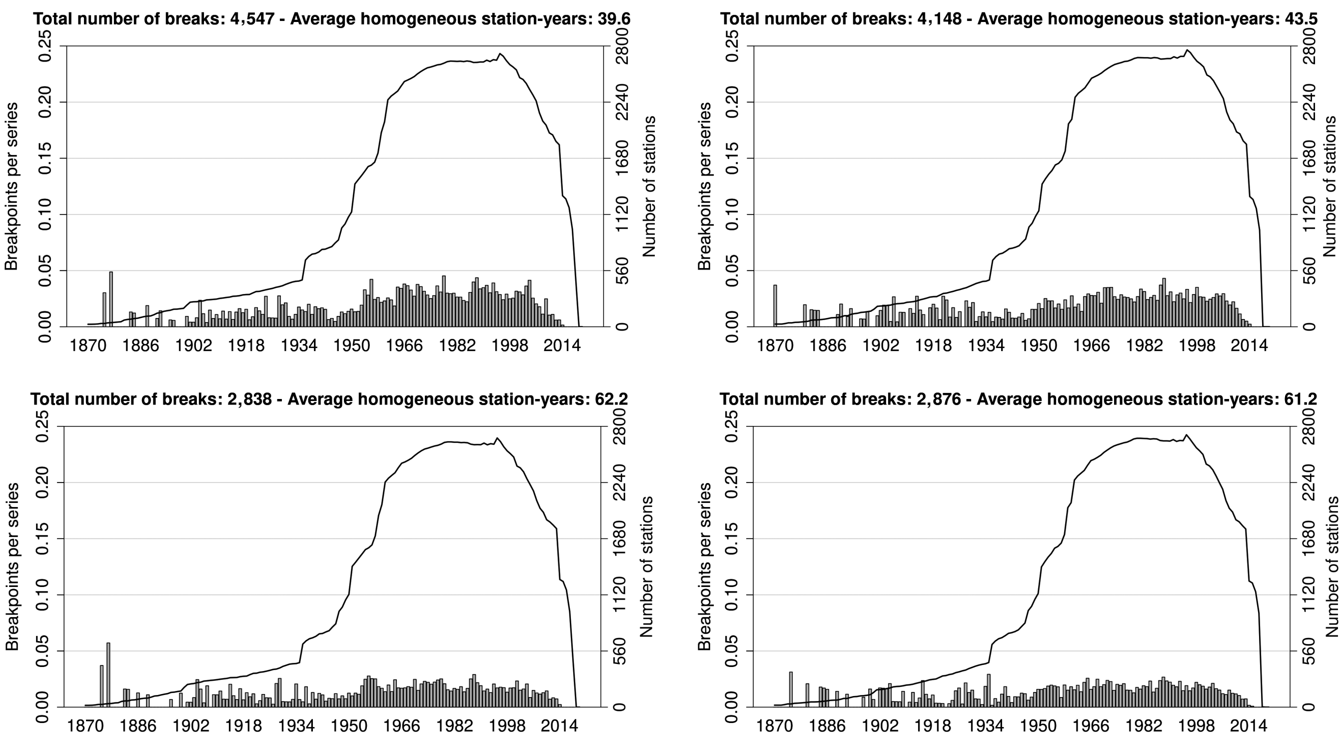

FIGURE 3 Statistics regarding application of break detection on ECA&D temperature dataset. Histogram describes number of breakpoints per series, line describes number of stations. Left (right) panels are about minimum (maximum) temperatures. Top and bottom panels are, respectively, for first and second break detection runs

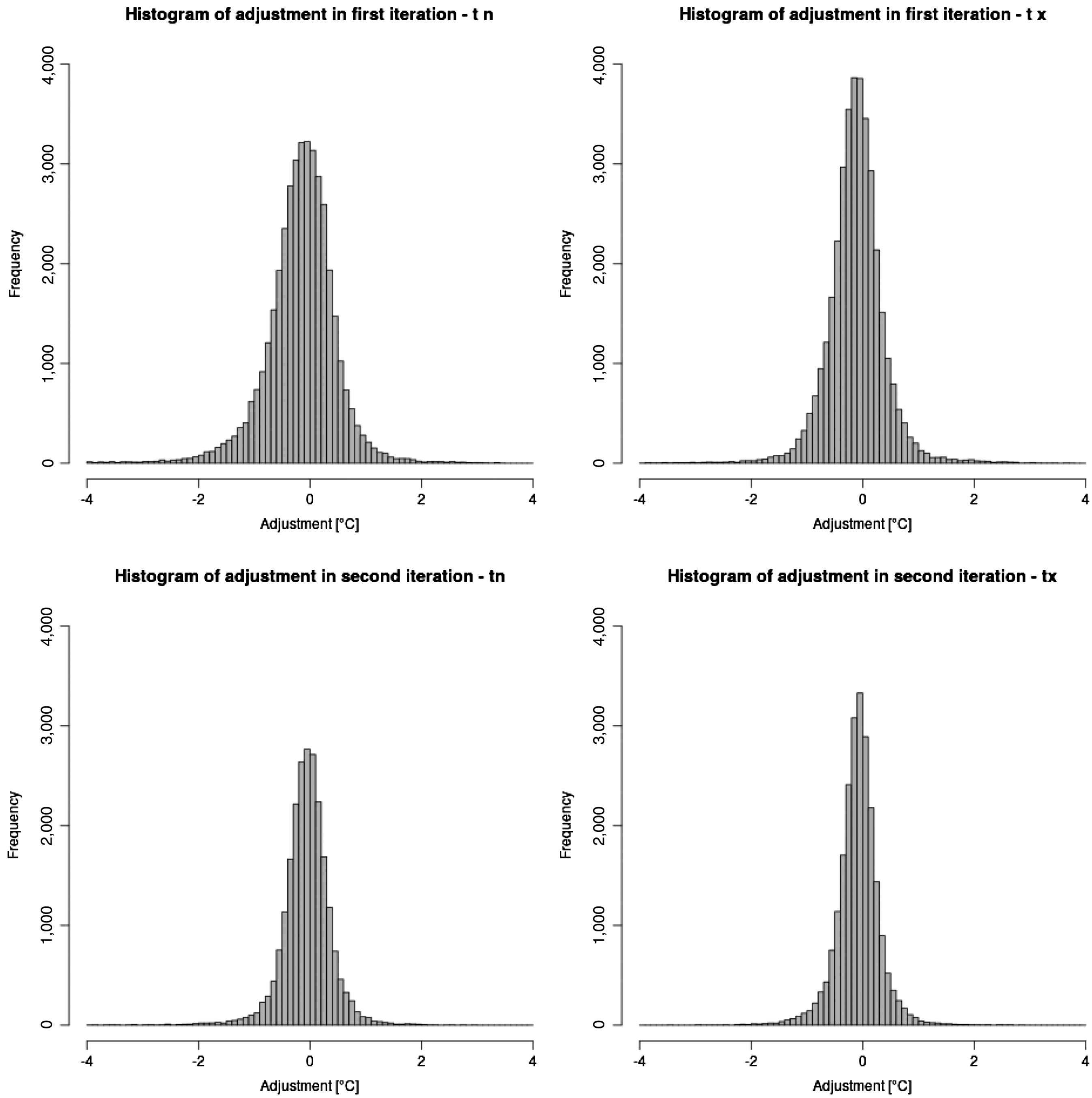

FIGURE 4 Histograms of adjustments for values in the median quantile (q50), for TN (left column) and TX (right columns) and for first iteration (top row) and second iteration (bottom row). Difference in width between first and second iteration proves the different role of the two phases

overlapping period with the target series. In such a situation no adjustments are possible. Nevertheless, the first homogenization iteration improved the number and the length of the homogeneous segments which made it possible to adjust additional breaks in the second iteration. The number of series for daily maximum and minimum series in the ECA&D dataset and the number of adjusted series after the first and second iterations are shown in Table 1.

Since the second iteration takes the results of the previous iteration as input, the possibility exists that this complex system diverges too strongly from the initial situation (e.g., positive feedbacks in the iterative processes, tendency

TABLE 2 Averages of adjustments for 5th, 50th and 95th percentile for TN and TX in first and second iterations

	TN (°C)			TX (°C)		
	05th percentile	50th percentile	95th percentile	05th percentile	50th percentile	95th percentile
First iteration	−0.12	−0.11	−0.10	−0.06	−0.05	−0.06
Second iteration	−0.03	−0.01	−0.01	−0.04	−0.03	−0.03

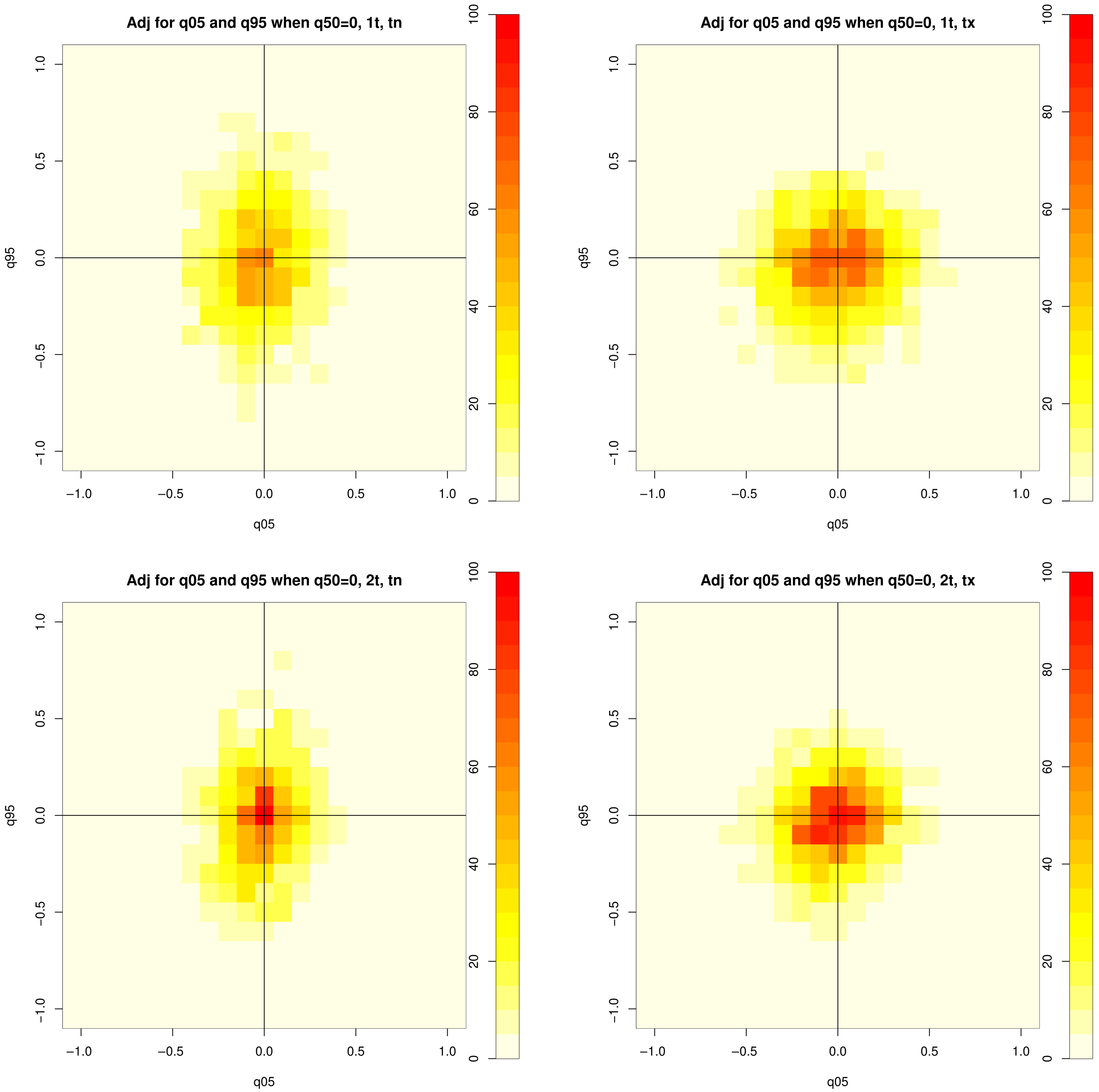

FIGURE 5 Density scatterplot of adjustments for quantile 95 versus quantile 05 when adjustments for the median are null. Minimum temperature (left column), maximum temperature (right column), first iteration (top row), second iteration (bottom row) [Colour figure can be viewed at wileyonlinelibrary.com]

for the removal of local signals or introduction of forced trends in the temperature series). This issue might be alleviated by applying the homogenization on the original series, using the homogenized series as reference in each iteration and setting a convergence threshold to stop the system. In the approach documented here, we simply limit the number of iterations to two.

The adjustments that have been applied to the breaks vary strongly with the month and the quantiles. Figure 4 shows that the adjustment of the median is symmetric around −0.1°C and smaller (and less frequent) adjustments are found for the second iteration. Furthermore, the distribution of adjustments of the median is wider for minimum temperatures and more peaked for maximum temperature. Averages of adjustments for the 5th, 50th and 90th quantiles (see Table 2) are more negative for TN. For both variables adjustments for lower quantiles are more negative, indicating a general tendency to broaden the probability density distribution. This is consistent with earlier findings (Lawrimore *et al.,* 2011; Trewin, 2013; Thorne *et al.,* 2016).

The peak in the distribution close to zero relates to: (a) the independence between break detection and adjustment calculation, that is, a break may be found but the comparison to the surrounding reference series does not give a reliable correction and (b) to the possibility that the median needs no adjustment but percentiles in the tails do.

This latter situation is illustrated in the scatterplots of adjustments for the 5th versus the 95th quantiles, for the series where adjustments of the median are null (Figure 5). These figures show a centred and symmetric distribution and indicate that adjustments are not skewed towards more positive or negative slopes. These figures also show that no thresholds on the absolute value of the adjustments are used, contrasting with the approach of Trewin (2013) who applied adjustments only if the resultant shift in annual mean exceeded the ±0.3°C threshold.

3.2 | Case studies

In order to demonstrate the method in more detail, two case studies are presented.

3.2.1 | Bamberg

An illustrative example is the adjustment of data from the station of Bamberg (Germany).[-999] Metadata reports a set of breaks (Table 3) which are only partially retrieved by the automatic break detection (first iteration: 1891, 1952; second iteration: 1920).

The two documented breaks are not reproduced exactly, but the 1948/1949 break is located within a few years. The further breaks that are detected are probably related to unrecorded changes in the features of the station.

The high density of stations in Germany and Austria provided by their respective meteorological services allow to have more than 18 reference series available for the break in 1952 on which we focus in this case study. The 18 highest correlated ones have been selected (Figure 6).

Shape and location of the probability distributions of the non-homogenized temperatures before (light red) and after (light blue) the break shows a clear distinction (Figure 7). Shifts in quantile sequences varies from very low values for the tails of the distribution to 1.1°C for the median, showing the different effect of the break on mean and extreme values. Probability plots after the two iterations of homogenization (red before and blue after the break) get closer to each other in different way depending on the quantiles, indicating that the difference between the two original distributions was not entirely due to the artificial intervention. The two sub-series (before and after the break) do not completely overlap due to the climatic variability that has been captured by the surrounding reference series and taken into account in determining the adjustments.

Estimates of adjustments related to each reference (for this case study) are shown in Figure 8. These are the results of the process described by Equation (2), followed by

[1]The German Weather Service now makes the data for station Bamberg available as separate series ranging from 1879 to 1958 and from 1949 to 2018.

TABLE 3 Available metadata regarding the station in Bamberg, Germany

January 1880 to December 1948	Bamberg (Sternwarte, city centre)	z = 283 m
January 1949 to March 1995	Bamberg (south, country side)	z = 243 m
March 1995 to present	Bamberg (south, country side)	z = 239 m

smoothing and check of negative slopes. Figure 8 shows that the lower quantiles has stronger (more negative) adjustments than the upper quantiles which increases the width of the distribution. As an example of the adjustment process, the estimates related to the value measured on 22 May 1951 (6.2°C) are highlighted in Figure 8. This measurement belongs to different quantiles (35th and 40th), depending on the reference series that is considered, since for each of these there is a different overlapping period with the target series. The final correction is taken as the median of the estimates, in this particular case the adjustment will be −1.3°C, with a final homogenized value of 4.9°C.

Effects on the series are evident when indices like the annual mean are plotted (Figure 9, top panel). In this particular case, the first iteration is able to correct the series almost entirely, since the break that has been detected during the second iteration (1920) had very low adjustments during the second one. Comparison of corrections for the mean and the two tails of the distribution show the expected differences: larger (smaller) corrections for the 5th (95th) quantile.

3.2.2 | Salzburg

A particularly representative case is the station of Salzburg (Austria), where the metadata reports a set of breaks (Table 4). Break detection detects almost all these breaks, except the most recent one which is probably small in amplitude (only

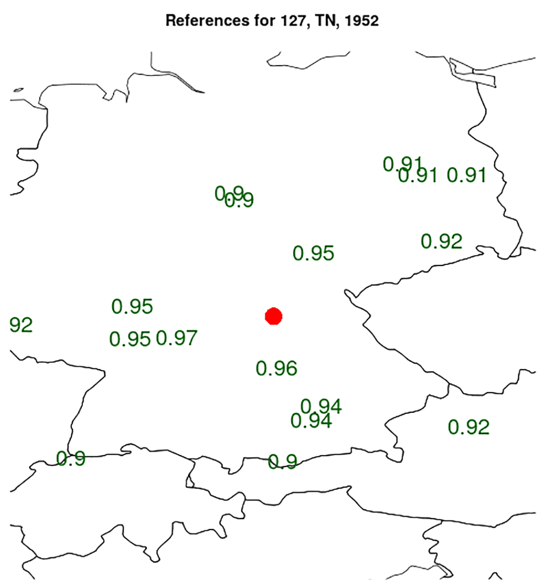

FIGURE 6 Reference series that have been used for the homogenization of the break in 1952 in the station of Bamberg (big dot). Numbers represent correlation calculated between references and basis series (considering the 20 years following the break) [Colour figure can be viewed at wileyonlinelibrary.com]

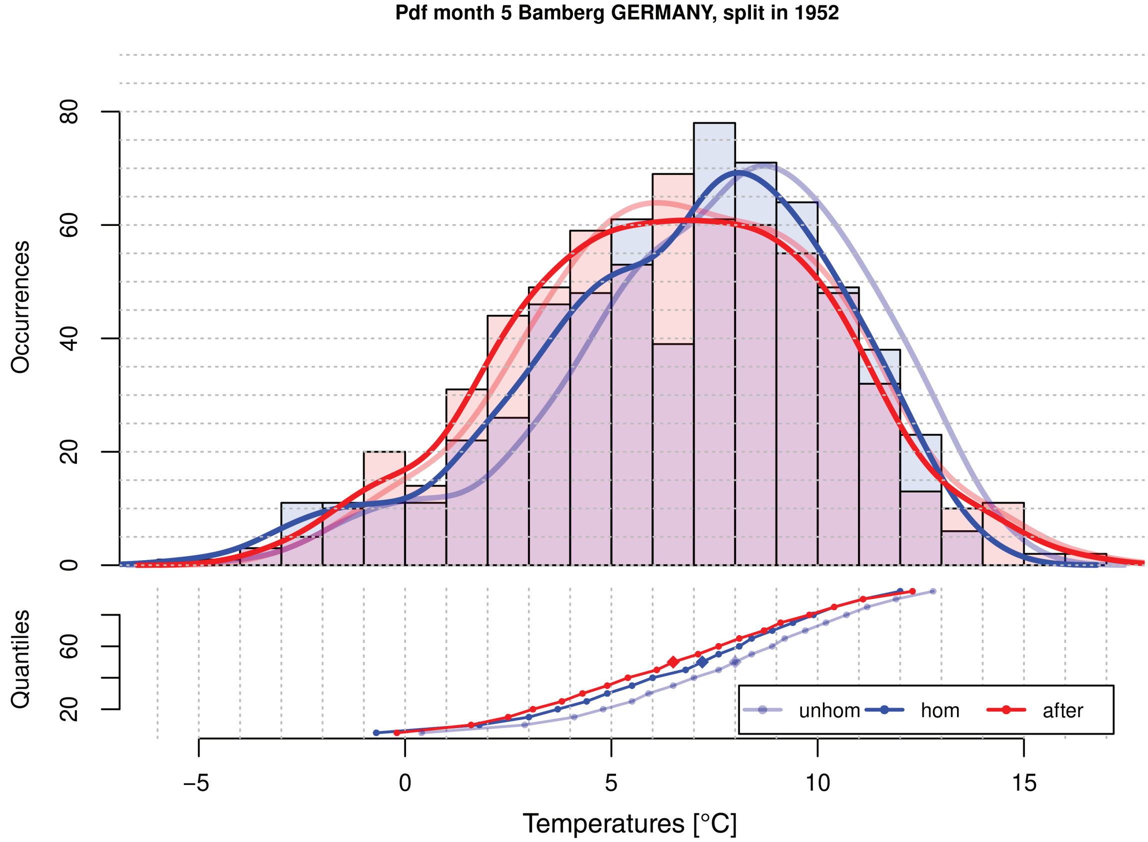

FIGURE 7 Top: Histograms and pdf of adjusted minimum temperature in May month for the 20 years before 1952 (blue) and 20 years after that (red). Light blue and light red curves represent original probability density functions. Bottom: Quantile functions related to the above distributions, same colour code

3 m of change in height), and detecting some further breaks which probably derive from unreported changes in the stations features. For this case study, focus will be on the break located near 1938 which is associated with the relocation of the station from the city to the nearby airstrip.

For the 1938 break, 12 reference series meet the requirements, see Figure 10. Shape and location of the probability distribution of temperatures before and after break in 1938 (Figure 11) show a clear shift of the distribution of records before the break. Shift in quantiles varies from 0.6°C for the 5th

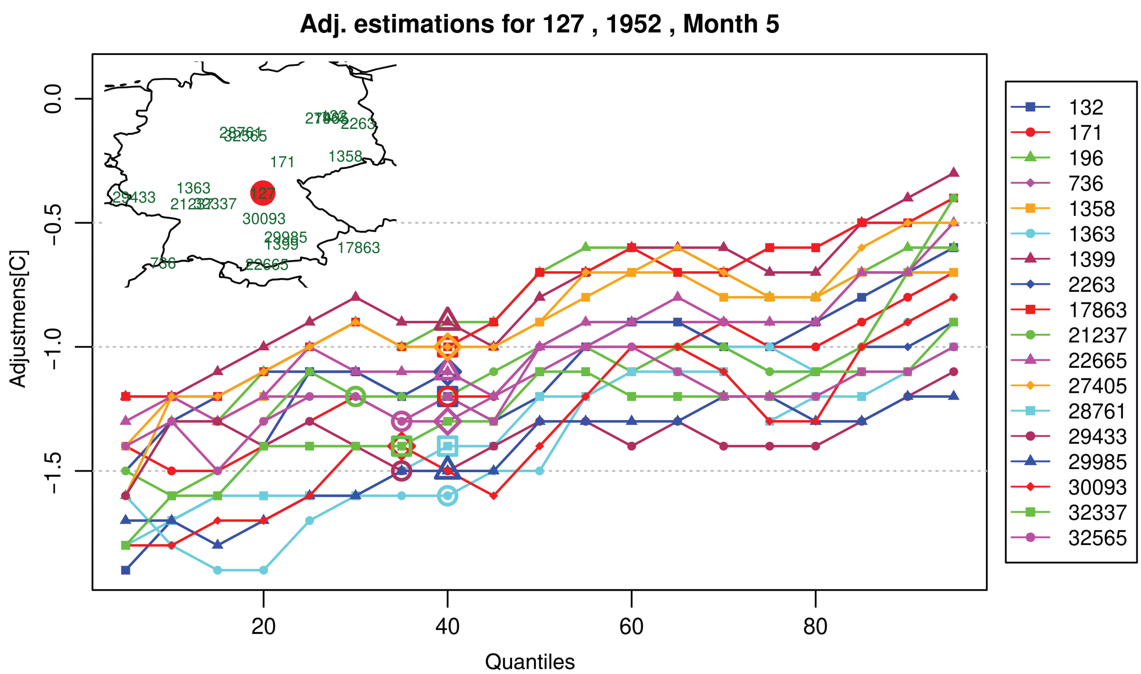

FIGURE 8 Estimation of adjustments for May month, station of Bamberg and break in 1952 after the smoothing process and the negative slope check

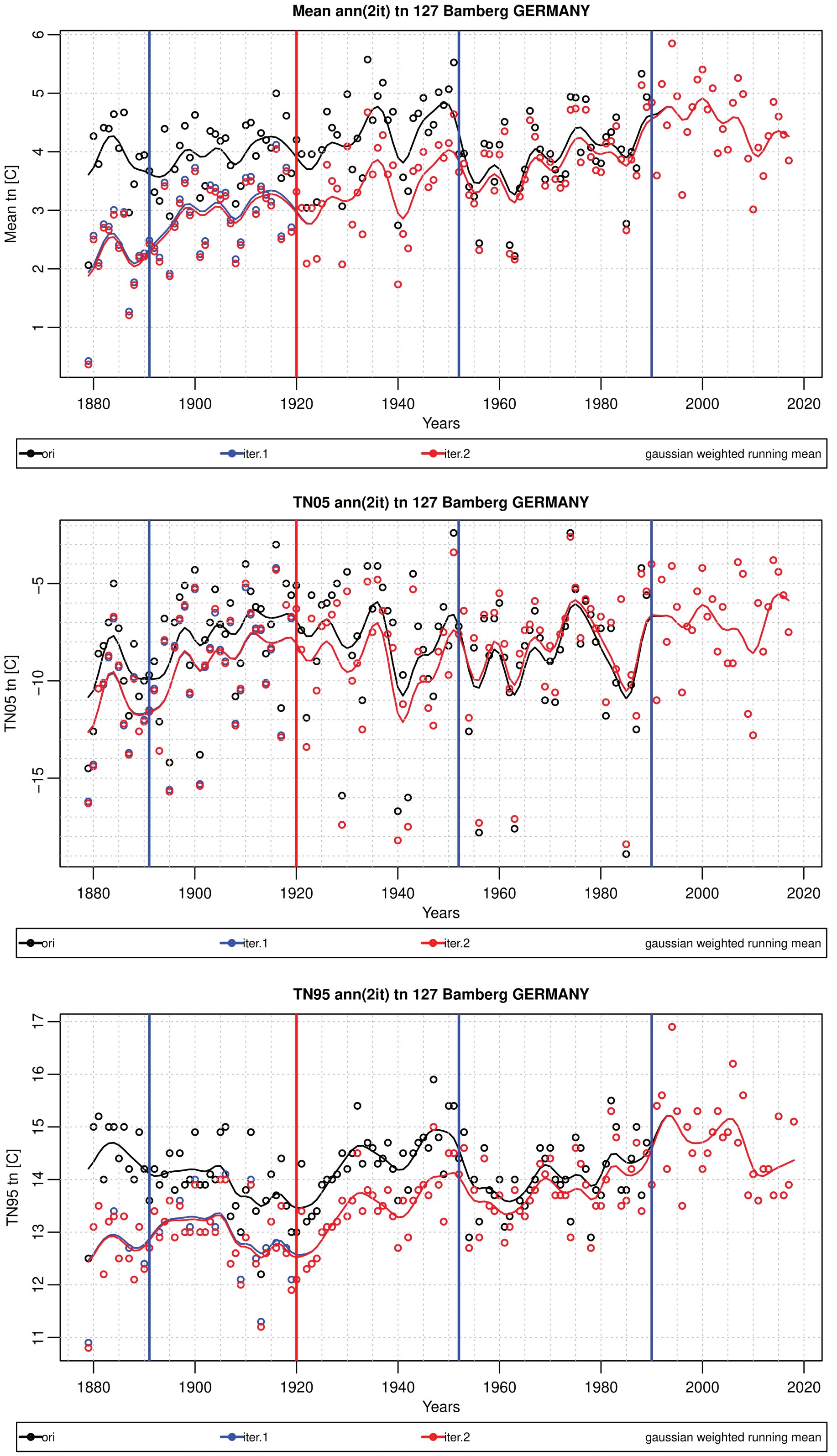

FIGURE 9 Annual mean (top), 5th quantile (centre), 95th quantile (bottom) time series for minimum temperatures in Bamberg. Black line: original series, blue: first iteration result, red: second iteration result. Vertical lines: output of break detection in the first (blue) and second (red) iterations

quantile to 1.1°C for the median, showing the different effect of the break on mean and extreme values. Probability distributions after the two iterations of homogenization almost overlap each other, indicating that a great part of the difference was due to the relocation, while the remaining difference represents actual climatic variability that relates to the surrounding reference series.

Adjustments applied may be seen in Figure 12, analogous to Figure 8 with the difference that here the slope of the curve in the quantile-adjustment plot is not as

TABLE 4 Available metadata for the station in Salzburg, Austria

January 1863 to December 1883	High school (Gymnasium Altes Borromäum)	z = 424 m
January 1884 to July 1903	High school (Oberrealschule)	z = 419 m
August 1903 to February 28, 1941	Studiengebäude-Lehrerbildungsanstalt	z = 423 m
March 1, 1939 to June 15, 1996	Airport station 1	z = 434 m
Since June 1996	Airport station 2	z = 437 m

steep as in the Bamberg case. The highlighted marks are related to the measurement on 14 May 1938 (4.9°C), whose adjustment will be −0.7°C, with a final value of 4.2°C.

Effects on the series are evident when indices like the annual mean are plotted (Figure 13, top panel). Interesting about this case is that first iteration (blue lines, when not covered by red) corrects the big breaks, such as the break in 1938. On the other hand, the second iteration is able to adjust two early breaks (red vertical lines) that were not detected during the first round because of the lack of long reference series in the early periods. The two new breaks are confirmed by the metadata (Table 4). The amplitude of adjustments in this case is clearly lower, showing that second iteration works as an enhancement of adjustments from the first round. Appendix B shows how in particular cases the second iteration is important for the homogenization of the older part of the series that was not corrected during the first round.

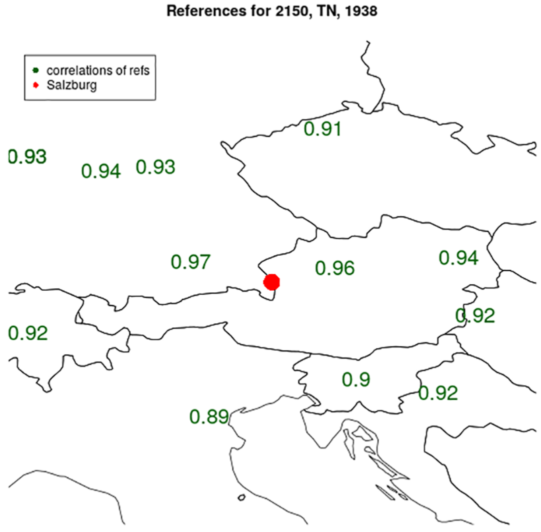

FIGURE 10 Reference series that have been used for the homogenization of the break in 1938 in the station of Salzburg (big dot). Numbers represent correlation calculated between references and basis series. Correlations are calculated using the 20 years following the break [Colour figure can be viewed at wileyonlinelibrary.com]

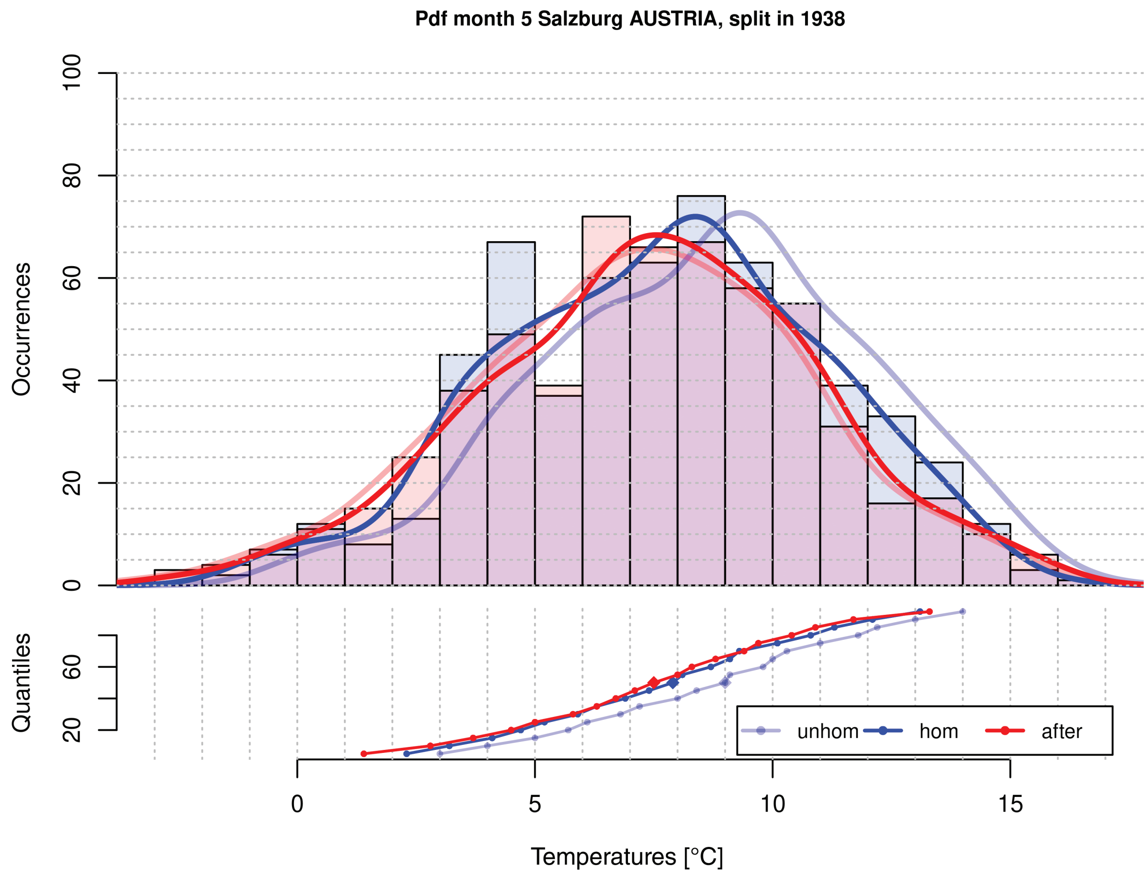

FIGURE 11 (Top) Histograms and probability distribution of adjusted minimum temperature in May for the 20 years before 1938 (blue) and 20 years after that (red). Light blue and red curves represent original distributions. Bottom: quantile functions related to the above distributions, same colour code

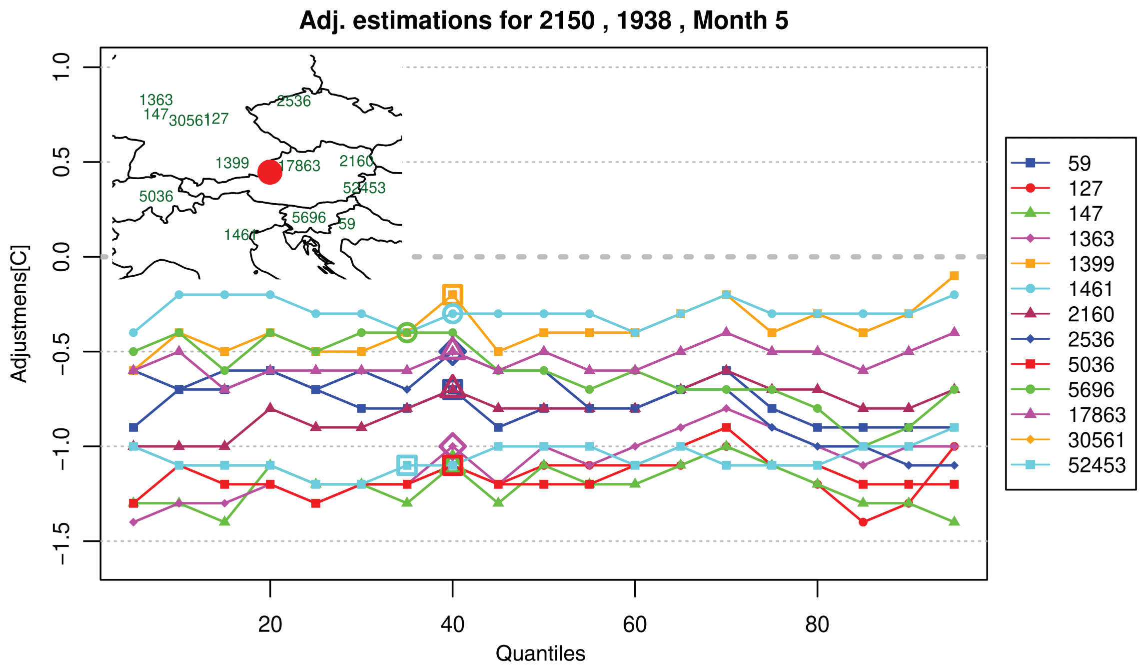

FIGURE 12 Estimation of adjustments for May, station of Salzburg and break in 1938 after the smoothing process and the negative slope check. The inset shows the locations of the series used to calculate the adjustment

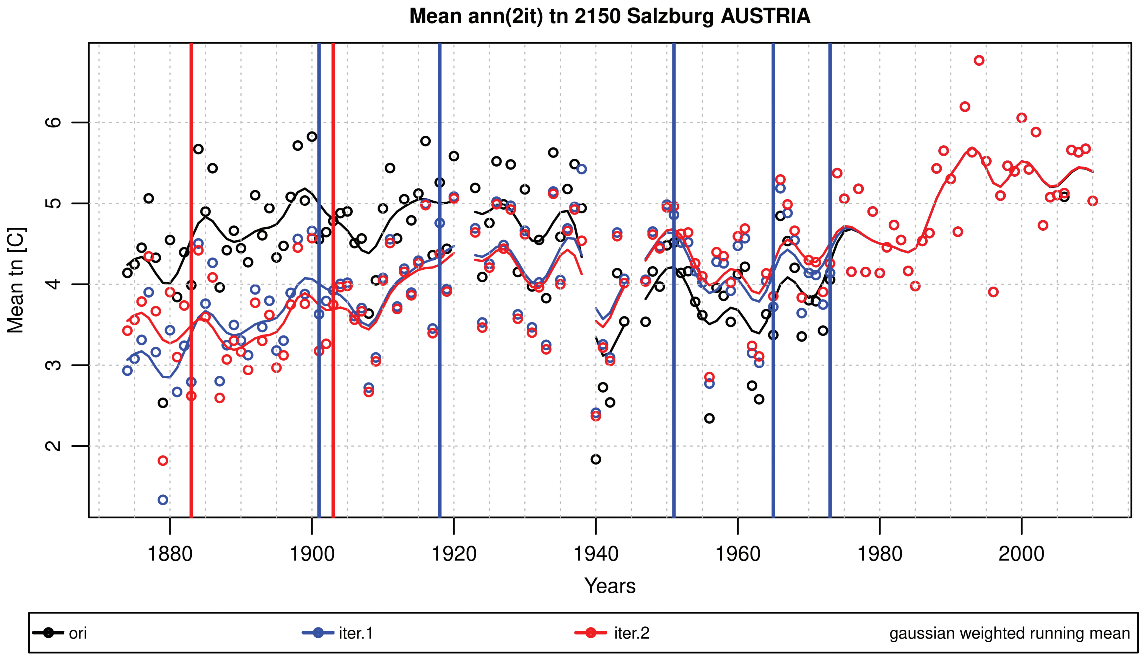

FIGURE 13 Annual mean time series for minimum temperatures in Salzburg. Black line: original series, blue: first iteration result, red: second iteration result. Vertical lines: output of break detection in the first (blue) and second (red) iterations

3.3 | Application to the complete dataset

Figure 14 shows trends in annual mean daily minimum and maximum temperature over the 1961–2010 period, before the homogenization (top panels) and after the homogenization (middle panels). The comparison of these figures shows the removal of several outliers and unrealistic low or high trends values. The trends based on the homogenized series are much more spatially homogeneous. The bottom panels of Figure 14 show the difference in trend values between the non-homogenized and homogenized series which demonstrate that the adjustments go both ways - trends are increased and decreased by this procedure. A coherent spatial pattern of adjustments is not evident from this figure.

The almost complete disappearance of stations with negative and very large trends demonstrates the effectiveness of the method in recognizing and keeping the climate signals that dominates the series and removing outliers trends which are related to artificial signals. Even though the result is the

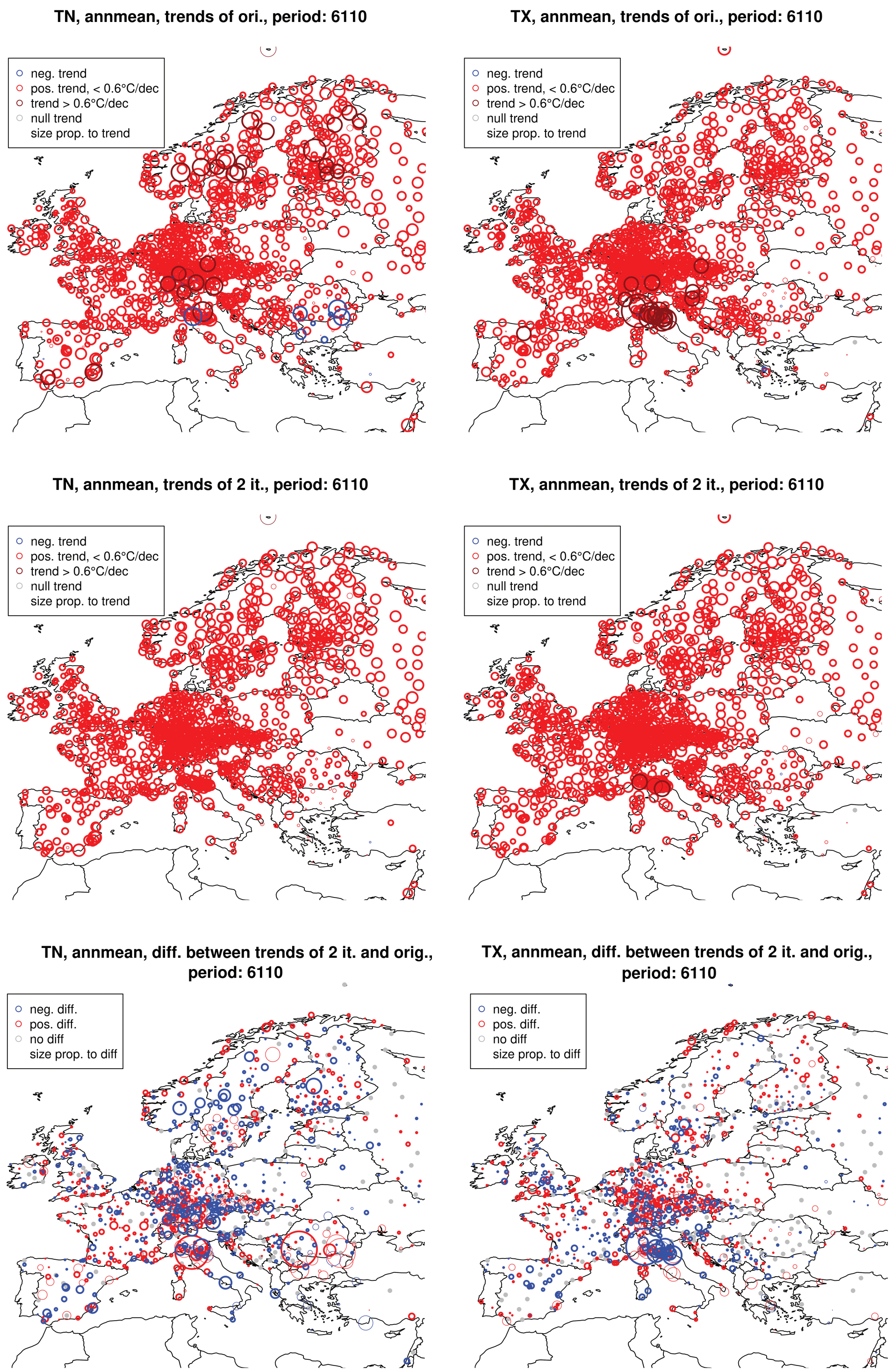

FIGURE 14 Maps of trends of annual mean in the period from 1961 to 2010 of original series (top), homogenized series (middle) and difference between the two (bottom) about minimum (left column) and maximum temperatures (right column). Blue circles indicate negative trends, red circles represent positive trends below 0.6°C/dec, dark red circles represent trends above 0.6°C/dec. Size of the circle is proportional to the amplitude of the trend. Thickness of the circle indicates significance of the trend itself (above 0.95). Code colour is chosen based on the box plots of Figure C1 (blue and brown values lie on the tails of the original distribution)

convergence of the trends of all stations to positive values, it is important to notice that the aim of this process is not the removal of the negative trends. This phenomenon is an indirect effect of the homogenization procedure. Indeed all stations with excessively high trends (i.e., dark red circles) have been adjusted with negative factors, as shown in Figure 14 (bottom row). A further check is to search for stations that still showed a negative trend or a exceptionally high trend exceeding 0.6°C/dec. Figure 15 shows the locations of the stations related to these extreme trends. These are not isolated stations but it is shown that these values are consistent with trends of neighbouring stations. The low value trends are mainly

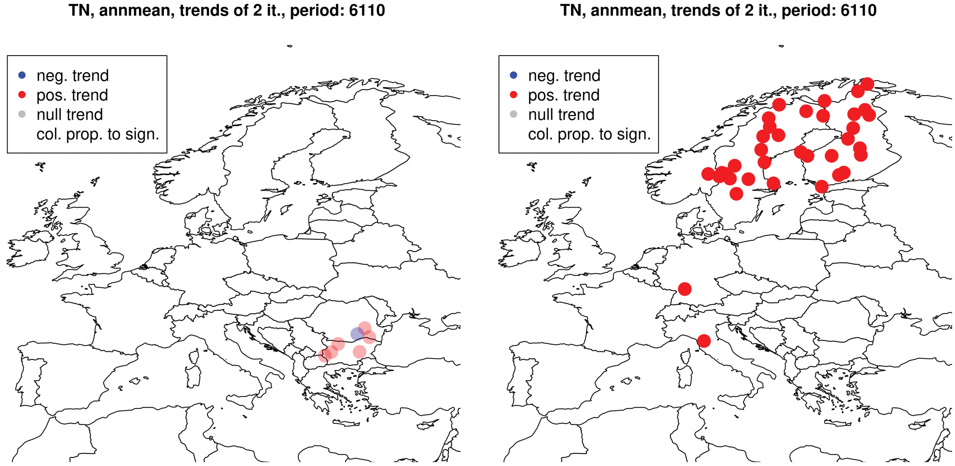

FIGURE 15 Left: Map of series having trend of annual mean of minimum temperatures lower than 0.1°C/dec, transparency of the dots indicate non-significant trend. Right: Map of series having trend of annual mean of minimum temperatures larger than 0.5°C/dec, transparency of the dots indicate non-significant trend

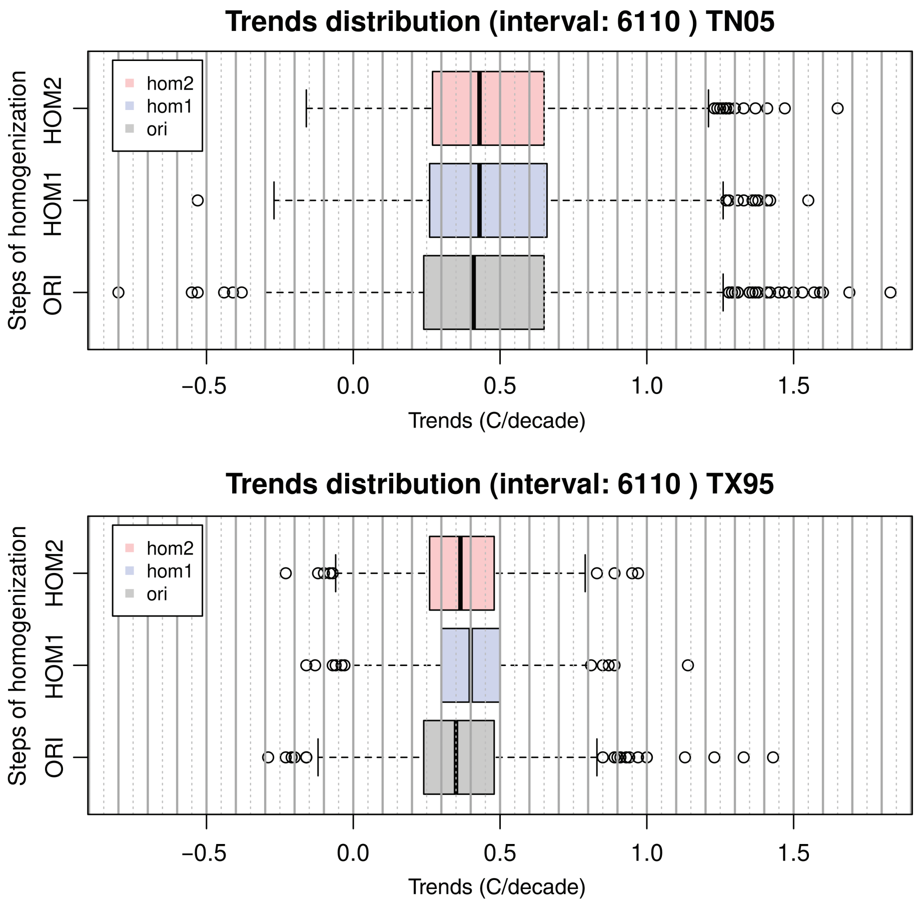

FIGURE 16 Distribution of trends for TN05 (top) and TX95 (bottom). Each plot shows boxplot for original (grey), first iterations' result (blue) and second iterations' result (red)

located in Bulgaria and southern Romania, while the very large trends are mainly in the Northern Baltic area. The second case is likely to be the result of a widespread climatic effect, while the first might be the result of the influence of the series of Bucarest on the neighbours.

In Appendix C box plots show the distribution of the trends in the annual mean of TN and TX for the two successive iterations. These indicate a narrowing of the distribution of the trends together with a significant reduction of outliers.

Figure 16 describes the distribution of the trends on extreme indices (5th percentile of TN and 95th percentile of TX) in the original dataset and after the two stages of homogenization.

TABLE 5 Median trend values of annual averaged daily minimum and maximum temperature, of TN05 and TX95, for the original dataset, after the first iteration and after the second iteration

	TN annual mean (°C/dec)	TX annual mean (°C/dec)	TN05 (°C/dec)	TX95 (°C/dec)
Original	+0.31	+0.35	+0.41	+0.35
First iteration	+0.31	+0.37	+0.42	+0.40
Second iteration	+0.32	+0.37	+0.42	+0.37

Beside the changes in the width of the distributions, changes in the first moment have been observed. The medians show a slight shift to higher values (Table 5) for annual means. TN05 and especially TX95 show a more irregular behaviour where overadjusted trends in the first iteration are refined in the second iteration.

4 | DISCUSSION AND CONCLUSIONS

A fully automatic homogenization method for daily temperature series has been presented and applied to the pan-European dataset ECA&D. The size of the dataset and the absence of detailed metadata for all series and stations require a procedure which is *blind* for metadata and is able to handle a large variety of data quality conditions. These challenges are met by using a combination of the break detection as an evolution of the method by Kuglitsch *et al.* (2012) and the quantile matching method which has been pioneered earlier by Trewin (2013) for use in large datasets. In order to distinguish between climatic signal and artificial signal in the breaks, a network of reference series in the vicinity of the record that needs adjustment is employed. In this study, the reference series are chosen from coordinate boxes of 6° × 6° around the target series, with thresholds on altitude difference. A further selection, using daily raw correlation and length of overlapping period, makes this approach sufficiently flexible to cope with both high and low station density areas. These criteria take inspiration from the nearest neighbour stations approach of Menne and Williams Jr (2009) and Trewin (2013).

The whole procedure is iterated twice. While the first iteration locates and adjusts the largest discontinuities, the second iteration is able to adjust more subtle changes such as earlier breaks, breaks with smaller amplitude or in areas with scarce station density. The more abundant presence of homogeneous sub-series after the first iteration makes this possible. The homogenization has been able to adjust about 2,700 ECA&D temperature series for both TN and TX, while the remaining ones are duplicates or consist in short records. A final round of break detection has shown that only 1,400 series can be considered completely homogenized, while on the rest minor breaks persist. No further iterations are made to remain as close as possible to the original dataset while adjusting for the largest inhomogeneities.

The trends in annual averaged values show a much stronger spatial consistency than before the adjustments. This illustrates the effectiveness in the removal of the artificial signals, thus making the climatic signal dominant. A comparison between trends prior and after the homogenization shows that changes in trends are both ways. The averages of the distributions of European trends of annual means are shifted slightly to warmer values (TN: +3.2%, TX: +5.7%). At the same time the interquantile range of the distributions of these trends are consistently reduced (TN: from 0.16 to 0.10°C/dec, TX: from 0.12 to 0.08°C/dec), indicating a higher uniformity of the values. Similar conclusions are reached when considering indices for extreme values, such as the 5th and 95th quantiles.

The strength of the quantile matching method is that each part of the distribution (i.e., quantile) is considered independently from the others. Previous studies on temperature probability distributions have focused on fitting the probability density functions with sophisticated functions or calculation of variations in the distribution parameters. On the other hand, the quantile matching has a more heuristic approach, aiming at being more versatile and able to adapt to the wide spectrum of signals that artificial activities may lead to the records. In distinction with the Trewin (2013) method no linear interpolation between quantiles (used to obtain adjustments for percentiles between the multiples of 5) is included, reducing the parameterization of the process. Furthermore, the calculation of adjustment is more conservative (use of averaged values and check of negative slopes) and the selection of reference series employs different criteria to give more importance to data availability and correlation.

Nonetheless further studies have been performed and are planned to understand minor controversial aspects of the described method. The dispersion of the reference series has been shown to affect the calculation of the adjustments (Appendix D). Therefore, it is planned to inspect how to lend the reference selection an "angular even distribution", that is, approximately same number of references on the north, south, east and west of the candidate series. The selection of reference series must also take into account the contribution of series with anomalous behaviour, as seen in Figure 15, where the negative trend in a station (Bucarest) might be one of the reasons of the lower trends observed in the surrounding area. The validation and the comparison of the results with other adjustment calculation methods are currently subject of further studies.

In conclusion, the method to adjust inhomogeneities in daily temperature discussed here is a purely statistical method. While the use of quantile matching favours the differentiation of adjustments for low and high daily extremes and have a seasonal cycle, these adjustments do not consider the existing meteorological circumstances. Future work using a physical approach to calculate the adjustments, in which actual weather contributes to the size of the adjustment, will give an alternative estimate of the homogeneity adjustment.

ACKNOWLEDGEMENTS

We acknowledge the data providers in the ECA&D project (www.ecad.eu). Funding has been received from the EU FP7 Collaborative Project UERRA (Uncertainties in Ensembles of Regional ReAnanalysis), grant agreement 607193, and the EU H2020 EUSTACE Project, grant agreement 640171.

ORCID

Antonello A. Squintu https://orcid.org/0000-0001-9011-4455

Gerard van der Schrier https://orcid.org/0000-0001-7395-8023

Yuri Brugnara https://orcid.org/0000-0001-8427-0064

Albert Klein Tank https://orcid.org/0000-0002-6275-2406

REFERENCES

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C. and Wieringa, J. (2003) *Guidance on metadata and homogenization*, WMO TD No. 1186, p. 53.

Alexandersson, H. (1986) A homogeneity test applied to precipitation data. *International Journal of Climatology*, 6(6), 661–675.

Alexandersson, H. and Moberg, A. (1997) Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends. *International Journal of Climatology*, 17(1), 25–34.

Begert, M., Schlegel, T. and Kirchhofer, W. (2005) Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology*, 25(1), 65–80.

Brandsma, T. and Van der Meulen, J. (2008) Thermometer screen intercomparison in De Bilt (the Netherlands)—Part II: description and modeling of mean temperature differences and extremes. *International Journal of Climatology*, 28(3), 389–400.

Brunetti, M., Maugeri, M., Monti, F. and Nanni, T. (2006) Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series. *International Journal of Climatology*, 26(3), 345–381.

Caussinus, H. and Mestre, O. (2004) Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 53(3), 405–425.

Della-Marta, P. and Wanner, H. (2006) A method of homogenizing the extremes and mean of daily temperature measurements. *Journal of Climate*, 19(17), 4179–4197.

Domonkos, P. (2013) Efficiencies of inhomogeneity-detection algorithms: comparison of different detection methods and efficiency measures. *Journal of Climatology*, 2013, 1–15.

Hartmann, D., Tank, A. and Rusticucci, M. (2013) *Climatic Change: Working Group I contribution to the IPCC fifth assessment report*, pp. 31–39.

Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Milate, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., van Engelen, A.F.V., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio López, J., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L. and Petrovic, P. (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. *International Journal of Climatology*, 22(12), 1441–1453.

Klok, E.J. and Klein Tank, A.M.G. (2008) Updated and extended European dataset of daily climate observations. *International Journal of Climatology*, 29, 1182–1191. https://doi.org/10.1002/joc.1779.

Kuglitsch, F.-G., Auchmann, R., Bleisch, R., Brönnimann, S., Martius, O. and Stewart, M. (2012) Break detection of annual swiss temperature series. *Journal of Geophysical Research - Atmospheres*, 117(D13), 1–12.

Lawrimore, J.H., Menne, M.J., Gleason, B.E., Williams, C.N., Wuertz, D.B., Vose, R.S. and Rennie, J. (2011) An overview of the global historical climatology network monthly mean temperature data set, version 3. *Journal of Geophysical Research - Atmospheres*, 116(D19), 1–18.

Lindau, R. and Venema, V. (2013) On the multiple breakpoint problem and the number of significant breaks in homogenization of climate records. *Idojaras: Quarterly Journal of the Hungarian Meteorological Service*, 117(1), 1–34.

Menne, M.J. and Williams, C.N., Jr. (2005) Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate*, 18(20), 4271–4286.

Menne, M.J. and Williams, C.N., Jr. (2009) Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, 22(7), 1700–1717.

Mestre, O., Gruber, C., Prieur, C., Caussinus, H. and Jourdain, S. (2011) Splidhom: a method for homogenization of daily temperature observations. *Journal of Applied Meteorology and Climatology*, 50(11), 2343–2358.

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. (1998) Homogeneity adjustments of *in situ* atmospheric climate data: a review. *International Journal of Climatology*, 18, 1493–1517.

ECA&D Project Team. (2012) *European Climate Assessment & Dataset Algorithm Theoretical Basis Document* (*ATBD*), version 10.5, De Bilt, the Netherlands.

Štěpánek, P., Zahradníček, P. and Skalák, P. (2009) Data quality control and homogenization of air temperature and precipitation series in the area of the Czech Republic in the period 1961–2007. *Advances in Science and Research*, 3(1), 23–26.

Thorne, P.W., Parker, D.E., Christy, J.R. and Mears, C.A. (2005) Uncertainties in climate trends: lessons from upper-air temperature records. *Bulletin of the American Meteorological Society*, 86(10), 1437–1442.

Thorne, P., Menne, M., Williams, C., Rennie, J., Lawrimore, J., Vose, R., Peterson, T.C., Durre, I., Davy, R., Esau, I., Klein-Tank, A.M.G. and Merlone, A. (2016) Reassessing changes in diurnal temperature range: a new data set and characterization of data biases. *Journal of Geophysical Research - Atmospheres*, 121(10), 5115–5137.

Toreti, A., Kuglitsch, F.G., Xoplaki, E. and Luterbacher, J. (2012) A novel approach for the detection of inhomogeneities affecting climate time series. *Journal of Applied Meteorology and Climatology*, 51(2), 317–326.

Trewin, B. (2013) A daily homogenized temperature data set for Australia. *International Journal of Climatology*, 33(6), 1510–1529.

Venema, V.K., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C.N., Menne, M.J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P. and Brandsma, T. (2013) Benchmarking homogenization algorithms for monthly data. *AIP Conference Proceedings*, 1552(1), 1060–1065.

Vincent, L.A. (1998) A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate*, 11, 1094–1104.

Vincent, L.A., Zhang, X., Bonsal, B. and Hogg, W. (2002) Homogenization of daily temperatures over Canada. *Journal of Climate*, 15(11), 1322–1334.

Wang, X.L., Wen, Q.H. and Wu, Y. (2007) Penalized maximal *t* test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology*, 46(6), 916–931.

World Meteorological Organization (2007) *Manual on the Global Telecommunication System*, WMO-No. 386. Geneva: WMO.

World Meteorological Organization (2011) *Guide on Climatological Practices*, WMO-No. 100. Geneva: WMO.

How to cite this article: Squintu AA, van der Schrier G, Brugnara Y, Klein Tank A. Homogenization of daily temperature series in the European Climate Assessment & Dataset. *Int J Climatol*. 2019;39:1243–1261. https://doi.org/10.1002/joc.5874

APPENDIX A: CHECK OF NEGATIVE SLOPES IN THE ADJUSTMENT SEQUENCES

During the application of the quantile matching method it might happen that the rank of measurements is not preserved. This occurs if the adjustment of a high quantile is smaller than that of a lower quantile.

This possible setback, that involves approximately 0.5% of the adjustment calculation, requires a constraint in order to keep the rank of data when the sequence has a negative slope in the adjustment—quantile plane. By definition, a quantile sequence including the result of the adjusting process ($\tilde{\mathbf{s}}_{j,q,m}$) must have a nonnegative slope. This implies that for any q:

$$\bar{\mathbf{s}}_{j,q+5,m}-\bar{\mathbf{s}}_{j,q,m}\geq 0. \quad (A.1)$$

For each q, elements of the adjusted quantile sequence are calculated as:

$$\bar{\mathbf{s}}_{j,q,m}=\mathbf{s}_{j,q,m}+\bar{\mathbf{a}}_{j,q,m}. \quad (A.2)$$

Thus

$$\begin{aligned}&\tilde{\mathbf{s}}_{j,m,q+5}-\tilde{\mathbf{s}}_{j,m,q}=\\&=\left(\mathbf{s}_{j,q+5,m}+\bar{\mathbf{a}}_{j,q+5,m}\right)-\left(\mathbf{s}_{j,q,m}+\bar{\mathbf{a}}_{j,q,m}\right)=\\&=\left(\mathbf{s}_{j,q+5,m}-\mathbf{s}_{j,q,m}\right)+\left(\bar{\mathbf{a}}_{j,q+5,m}-\bar{\mathbf{a}}_{j,q,m}\right)\geq 0.\end{aligned}$$

And finally,

$$\left(\bar{\mathbf{a}}_{j,q+5,m}-\bar{\mathbf{a}}_{j,q,m}\right)\geq-\left(\mathbf{s}_{j,q+5,m}-\mathbf{s}_{j,q,m}\right). \quad (A.3)$$

This constraint is implemented fixing the adjustment related to the median and checking the two tails quantile by quantile. In case of a too negative slope, the value is corrected moving it to the closest acceptable value.

For instance, for the right tail of the distribution, if:

$$\mathbf{a}\left(\tilde{j},55,\tilde{m}\right)-\mathbf{a}\left(\tilde{j},50,\tilde{m}\right)<-\mathbf{s}\left(\tilde{j},55,\tilde{m}\right)+\mathbf{s}\left(\tilde{j},50,\tilde{m}\right) \quad (A.4)$$

then the corrected adjustment is set to:

$$\bar{\mathbf{a}}\left(\tilde{j},55,\tilde{m}\right)=\mathbf{a}\left(\tilde{j},50,\tilde{m}\right)-\mathbf{s}\left(\tilde{j},55,\tilde{m}\right)+\mathbf{s}\left(\tilde{j},50,\tilde{m}\right) \quad (A.5)$$

APPENDIX B: MUNICH

The need and the utility of the second iteration with the break detection and homogenization can be appreciated when looking data from the station of Munich (Germany). Metadata reports a set of breaks (Table B1) including one in the 1920s.

Here the big difference between red and blue lines in Figure B1 indicates that the first iteration was not able to correct the breaks in 1912 and 1926. Nevertheless, the higher availability of data and long homogeneous segments, together with a better signal to noise ratio has allowed to adjust the earliest part of the series on the second run of the software.

TABLE B1 Available metadata regarding the station in Munich, Germany

January 1879 to July 1954	Munich (Botanic Garden Nynphenburg)	z = 515 m
August 1954 to March 1999	Munich (Nynphenburg residential area)	z = 515 m
April 1999 to present	No location metadata (probably not changed; changes in measuring times in 2001–2004)	z = 515 m

APPENDIX C: TREND ASSESSMENTS ON ANNUAL MEANS

Assessment of trends before and after the homogenization has been computed on annual means, showing a relevant narrowing of the distribution, especially between original

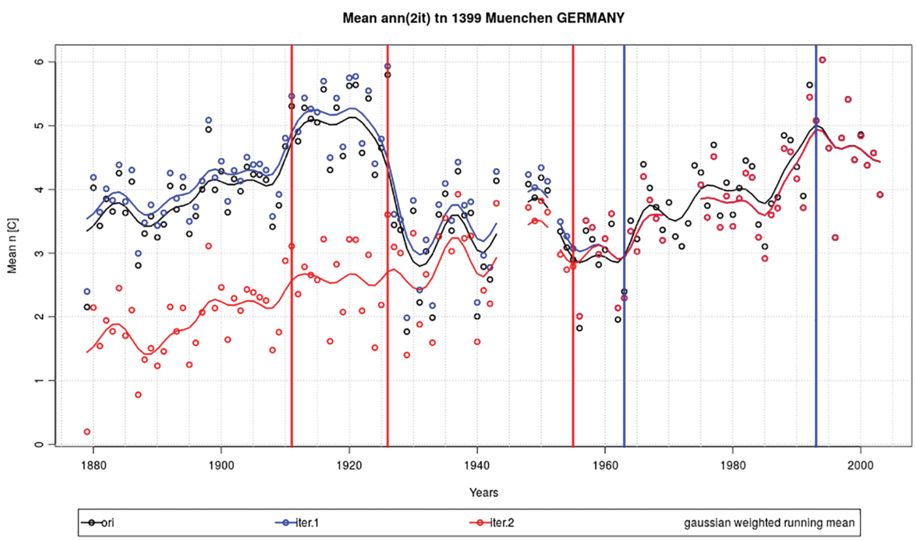

FIGURE B1 Annual mean time series for minimum temperatures in Munich. Same colour code as previous figures and first iteration, while the second iteration acts more like a refining of the result (see Figure C1).

APPENDIX D: GEOGRAPHICALLY INDUCED PATTERNS ON ADJUSTMENTS

Adjustment calculations may depend on the geographical distribution of reference series around the target series. The homogenization of the break in 1938 of series of Salzburg (Austria) has been checked considering separately the reference series that lie on the north (south, west and east), see Figure D1, top panels. Both pairs north–south and east–west present some differences, more evident in the case east–west. Reason of this difference is highlighted in the adjustments sequences for the month of May of western and eastern series, where it is clear that four series in the western dataset introduce larger (negative) adjustments. This series correspond to the four north-western German stations, which differ for topographical characteristics with respect to the rest of the stations.

Therefore, the extremely high variety of European topography and climate features require to perform an accurate choice of the references for each target series, reasoning on its locations, surroundings, etc. Furthermore, in some cases the sparseness of stations density does not allow to be able to select the best stations. In further versions of the software an even angular distribution of the references around the target will be implemented in case of regions with high density of stations.

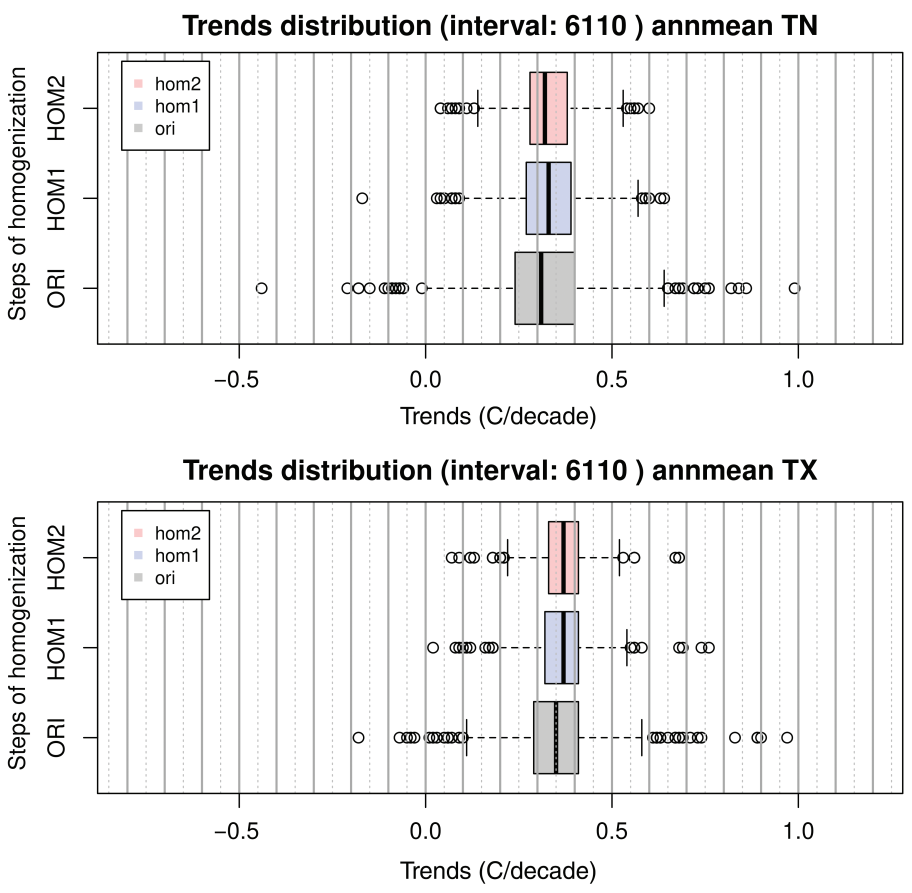

FIGURE C1 Distribution of trends for annual mean of minimum (top) and maximum temperature. Each plot shows boxplot for original (grey), first iterations' result (blue) and second iterations' result (red)

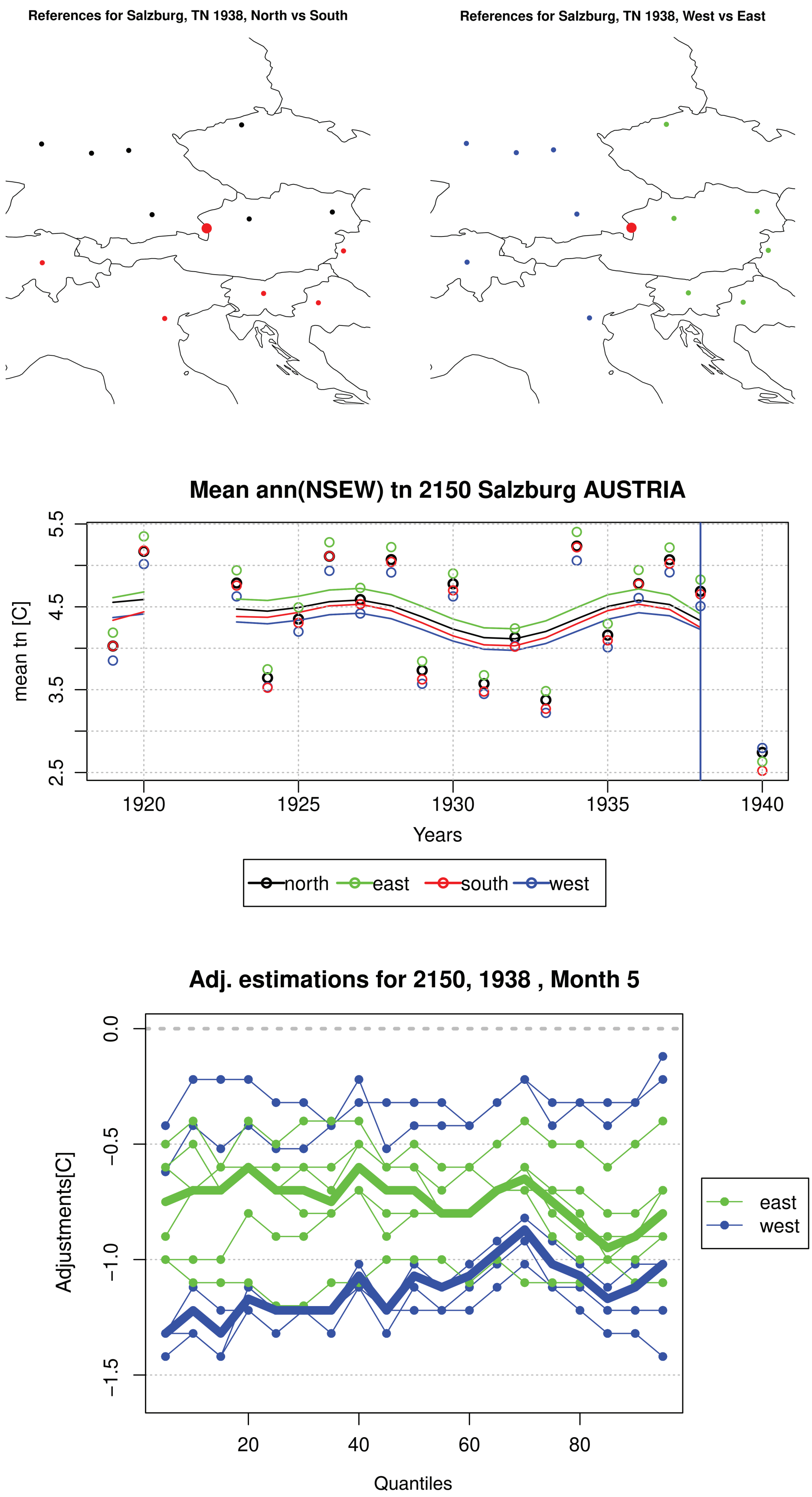

FIGURE D1 Annual mean of minimum temperatures in Salzburg after homogenization using four different sets of reference series: northern (black), eastern (blue), southern (red), western set (green)