Received: 1 April 2022 | Revised: 18 September 2022 | Accepted: 19 September 2022

DOI: 10.1002/joc.7881

RESEARCH ARTICLE

International Journal of Climatology RMetS

Homogenization of Swedish mean monthly temperature series 1860–2021

L. Magnus T. Joelsson | **Erik Engström** | **Erik Kjellström**

Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

Correspondence
L. Magnus T. Joelsson, Swedish Meteorological and Hydrological Institute, Norrköping 601 76, Sweden.
Email: magnus.joelsson@smhi.se

Abstract

The Swedish monthly average 2 m temperature observational data set is homogenized with a new automated version of the homogenization tool HOMogenizaton softwarE in R. Data from 836 individual time series (1860–2021) are merged into 456 time series with a novel automatic merging method. Merging limits the need of interpolation of data and increase the number of long time series without a net loss of data. Twenty-two of the merged time series were found to be homogeneous. For the other time series, the median time per break-point is 17 years. 37% of the detected break-points are supported in meta data. 42% of the data are corrected by $\pm 0.5^\circ$C or less, 3% by $\pm 1^\circ$C or more. On average corrections are negative, larger in the early periods, and larger in the summertime. The average trend 1860–2021 in the resulting merged and homogenized data set is $(0.13 \pm 0.03)^\circ$C decade, which does not significantly differ from that of the raw observational data. Extremely warm months, defined as being outside of three times the standard deviation from the average of the full time series, are most frequent and extremely cold months least frequent in the most recent 30-year period (1991–2020). In the homogenized data set, extremely warm months are even more frequent and extremely cold months even less frequent in 1991–2020, than in the raw observational data set.

KEYWORDS

climate, HOMER, homogenization, merging time series, temperature

1 | INTRODUCTION

Climatological studies require time series data that are both homogeneous and that covers a sufficiently long time period (Venema *et al.*, 2020). Long observational records often have artificial shifts, for example due to changes in location (e.g., Dienst *et al.*, 2017), measurement equipment (e.g., Auchmann and Brönnimann, 2012), or land-use (e.g., Sun *et al.*, 2016). The homogeneity of observational data must therefore be tested, and if required, data need to be homogenized. Several statistical tools are developed for this purpose (Ribeiro *et al.*, 2016; Venema *et al.*, 2020).

At the Swedish Meteorological and Hydrological Institute (SMHI), 35 monthly average 2 m temperature time series were homogenized in 2011 using Standard Normal Homogeneity Test (SNHT, Alexandersson, 1986;

Abbreviations: MORA, Meteorologisk Observationsdatabas för Realtid och Arkiv; OLS, ordinary least squares; SMHI, Swedish Meteorological and Hydrological Institute; SNHT, Standard Normal Homogeneity Test.

Alexandersson and Moberg, 1997). The homogenization was last updated in 2014 to cover the period 1860–2013. More recent observational data has been added to the data set since then, without further homogenization. An application of the homogenized data set is the Swedish temperature climate indicator (SMHI, 2021). This homogenized data set, covering 35 time series, homogenized with the SNHT-method is henceforth called '2014 data set'.

In 2021, SMHI shifted the new climatological normal period 1991–2020 according to WMO (2017), which called for a new homogenization of the monthly mean temperature data. The SNHT method is a reliable but labour-intensive method and the addition of new observations and newly digitized older observations demand for the homogenization data set to be continuously updated. Therefore, an automatic homogenization method is appealing. The use of an automatic method also enables the homogenization of larger data sets. A larger data set can make the homogenization itself more stable as each time series can be compared with more time series from neighbouring weather stations. In addition, it makes the subsequent data analysis more stable, especially for analysis of regional climate features.

A well-established homogenization tool is HOMogenizaton softwarE in R (HOMER, Mestre *et al.*, 2013), used by several meteorological institutes (e.g., Coll *et al.*, 2014; Vertačnik *et al.*, 2015; Kuya *et al.*, 2020). HOMER is a semi-automatic tool, but has recently been develop into a fully automatic tool called Bart (Joelsson *et al.*, 2022). The performance of Bart has been evaluated on synthetic benchmark data (Coscarelli *et al.*, 2021; Joelsson *et al.*, 2022).

In the current study, a new homogenized monthly average temperature dataset from the Swedish observational network is presented. In Section 2, the preparation (Section 2.1) and the homogenization (Section 2.2) processes are described. In Section 3, the resulting merged input (Section 3.1) and homogenized (Section 3.2) datasets are described. In addition, climatological analysis of the dataset are described in Section 3.3. The approach and results are discussed in Section 4, and the study is concluded in Section 5. Future work is suggested in Section 6.

2 | APPROACH

2.1 | Merging of time series

SMHI's digitized observation database MORA (Meteorologisk Observationsdatabas för Realtid och Arkiv, SMHI, 2022a) includes 961 individual series of monthly mean temperature with at least one datum between 1860 and 2021. In the purpose of creating a homogenized data set, it is desirable to use long time series (Venema *et al.*, 2020). HOMER, the homogenization tool used in current study, require at least 15 years of data overlap for two time series to act as references for each other in the homogenization process. A time series with less than 15 years of data can therefore not be included, which leaves 686 of the 961 monthly temperature series in the MORA database to be run through HOMER.

To optimize the use of the available data, time series can be merged (Kuya *et al.*, 2020). At SMHI, time series of groups of stations has traditionally been merged manually, especially for climatological purposes ('Månadens väder och vatten', SMHI, 2022b). The merging scheme of the homogenized 2014 data set is presented in Figures S1–S21 in Supplementary Material 1. Note that the observational time series in MORA occasionally consists of data from several weather stations, similar enough to be regarded as singular weather stations with re-locations. Here, the term 'hard' merging point refer to a month in an observational time series for which the data is observed at a different station than the previous month. 'Applied' merging point refer to a month in a merged time series for which the data is retrieved from a different observational time series than the previous month.

A few principles are used here to merge time series from different stations:

1. Data from time series with high degree of data coverage are preferred.
2. The stations must be situated close to each other.
3. The time series should not be poorly correlated with each other.
4. Stations located on islands, near coastlines, and inland should not be mixed.
5. Data from time series with more recent observations are preferred.
6. The number of change points between different time series should be kept low.
7. As the resulting merged data set here is subject to homogenization, an additional principle is used: The current homogenization method compares parallel time series, assumed to record the same climatology, to find non-climatological shifts in the time series ('break-points', see Section 2.2). Since two time series with enough overlap therefore can be used to find break-points, two time series with long overlap (≥15 years) should not be merged.

A script following the principles stated above is developed to merge suitable time series. The script has two parts: A grouping part where stations are grouped and a merging part where the time series of the different groups

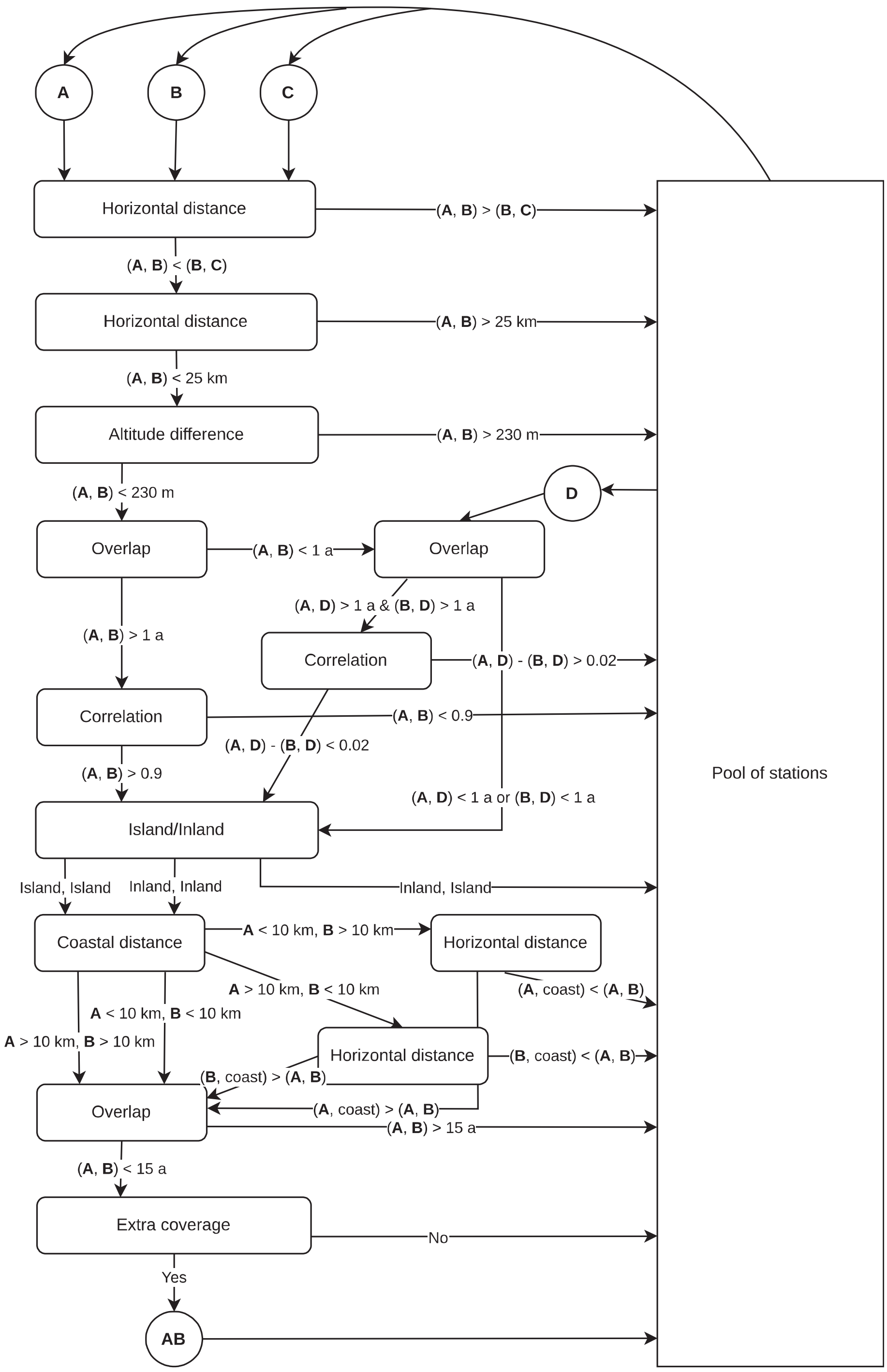

FIGURE 1 The process of grouping stations in the automatic merging scheme. **A** is the station with the longest time series yet not tested in the current grouping round. **B** is the station closest to **A**. **C** is the station closest to **B**. **D** represent a selected station close to **A** and **B**. **AB** is a group, including stations **A** and **B** with a preliminary merged time series. When all stations are tested the next grouping round is started, where the new groups of stations are included. The process is repeated until no new pairs of stations are tested.

of stations are merged. The workflow of the grouping part of the automatic merging script is depicted as a flow chart in Figure 1. The merging part of the script is illustrated in Figure 2.

The longest acceptable distance between stations is 25 km horizontally which corresponds to the distance between Holmön and Holmögadd, the pair of merged stations that are farthest apart in 'Månadens väder och vatten' (SMHI, 2022b). Similarly, the largest altitude difference between a pair of merged stations is approximately 230 m, which is the altitude difference between Gäddede and Gäddede A. Kuya *et al.* (2020) used a horizontal distance threshold value of 10 km and maximum different in altitude of 100 m.

A Pearson correlation coefficient ρ of .9 is the lowest acceptable correlation between two stations. The correlation is only calculated between two time series (A and B) if the overlap is at least 12 months. If the overlap is less than 12 months, but there is a third time series (D) from a relatively close neighbouring station for which both time series has a 12 months overlap, the correlation is tested with the third time series. If the difference between

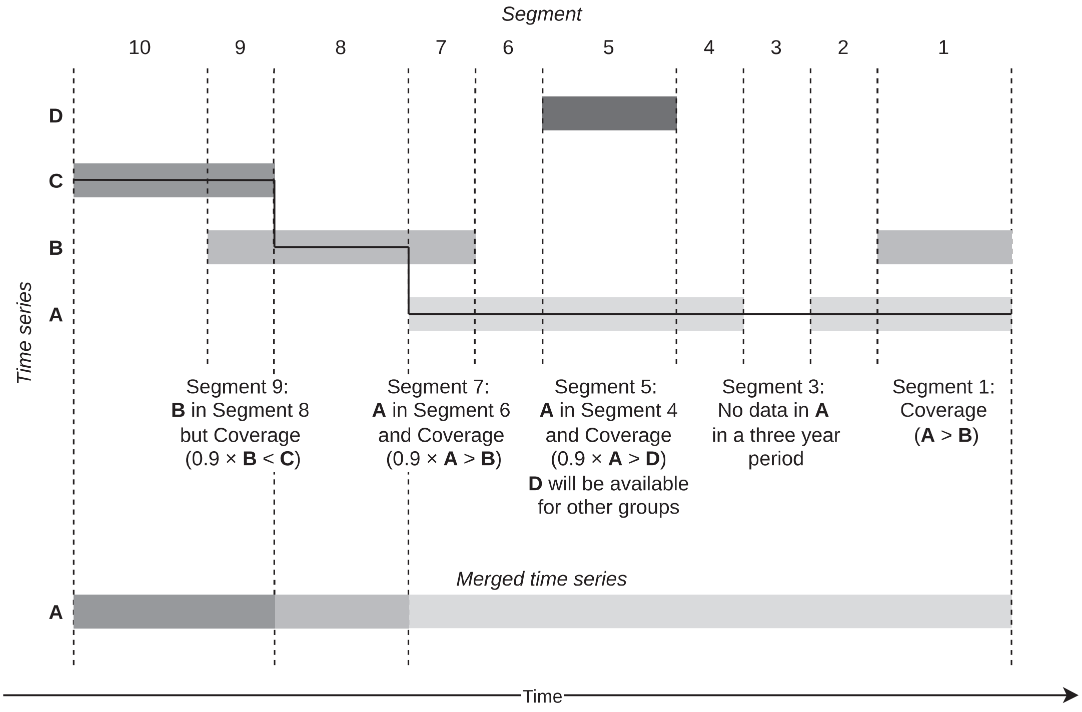

FIGURE 2 The process of merging time series from a fictitious group of stations **ABCD** in the automatic merging scheme. Bars represent occurrence of observations at the stations within the group: **A**, **B**, **C**, and **D**. The dashed lines represent possible change points, dividing the different segments 1–10. A possible change point include the time points before and after a period without data of at least 3 years, segments which include periods without data shorter than 3 years are not divided. The solid line represent the selected data. Data is selected from stations with the highest data coverage ratio (referred to as 'Coverage' in the figure), unless data is selected from another station at later segment, in which case the coverage must be more than 10% higher to make a change. **D** is not included in the final merged time series and will be available for other groups

the correlations is too large ($\rho_{AD} - \rho_{BD} > .02$), the pair of stations fail the correlation test.

Two stations are not paired if one of the stations is defined as a coastal station (closer than 10 km to the coastline), the other station is defined as an inland station (farther than 10 km from the coastline), and the stations are farther from each other than the distance from the coastline to the coastal station (to enable merging of stations that are close to each other but happens to fall on either side of the 10 km-line). In 'Månadens väder och vatten' (SMHI, 2022b), Ronneby (2 km from the coast) and Ronneby-Bredåkra (9 km from the coast) are merged. Two stations cannot be merged if they overlap more than 15 years, according to Principle 2.1. Finally, there must be extra coverage between the two stations, meaning that at least 1 month must have a datum in **A** but not in **B**, and vice versa.

After the stations are grouped according to Figure 1, the group's time series are constructed according to Figure 2.

As opposed to the original time series, a merge time series does not belong to one single station. Here, the merged time series are ascribed to the location of the station with the latest contributed observation to the time series.

2.2 | Homogenization

The homogenization is performed with Bart which is an automated version of the interactive homogenization tool HOMER. Bart uses an array of input parameters to reject or confirm break-points suggested by HOMER.

Two novelties have been introduced in Bart since the version described in Joelsson *et al.* (2022): a reference selection method and an optional maximum reference limit. The time series which is homogenized in a homogenization process is called the 'candidate' time series, the time series used for comparison in order to find a break-point are called 'reference' time series. In order to keep the number of references similar for all candidate time series, a maximum number of references is introduced. The maximum number of references excludes the least suitable time series if there is a surplus of time series that meet the criterion. This measure also reduces the computational time.

In addition to the maximum number of references, a new *hybrid* function for the selection of references is introduced. The hybrid reference selection function uses both proximity and correlation to select references and ensures that there is data coverage in at least a certain number of references for all time points of the candidate time series. In the current homogenization, the hybrid selection method is used. The principles behind selection of reference time series have similarities with the principles behind the merging between time series as described in Section 2.1, such that there is an apparent risk of confusion. The merging of time series and the selection of references should not be confused.

In the observational network's inspection protocols 533 possible break-points are suggested. These possible break-points include changes in the geographical location of the weather station, measurement equipment, and observers. The meta data also includes 723 hard merging points (in the original observational data from the MORA database) and 1,009 applied merging points (from the current automatic merging scheme), described in Section 2.1. In total, 2,265 possible break-points (including break-points suggested in the inspection protocols, hard and applied merging points) are used as meta data input for Bart. A list of the break-points suggested in meta data is included in Supplementary Material 2.

The set of input parameters for Bart are listed in Table S2 in the Supplementary Material 1. The setup is identical to the *HOMER-inter* in Joelsson *et al.* (2022) with the exceptions of the correlation threshold, the reference selection method, and the maximum number of reference series.

The method of correction and gap filling is an integral part of HOMER. The estimated values are a sum of a climate effect, common for the reference stations, and a station effect, constant in time within a homogeneous period (Caussinus and Mestre, 2004). The method is called 'ANOVA correction model' and is described in detail in Domonkos (2022).

The observations based unhomogenised data set is henceforth referred to as the 'raw' data set. The raw data set is gap filled with Bart. This version of the raw data set is referred to as the 'gap filled raw data set'.

2.3 | Data analysis

The temperature trend is estimated using Ordinary Least Squares (OLS) linear regression over the average annual means over 35 selected time series included in the 2014 data set. Both the trend over the entire period (1860–2021) and the period 1981–2021 are estimated. 1981 is used as the starting year of the recent global trend estimated in NOAA (2020). Linear trends for the period 1860–2021 are also calculated for all the time series ('National'), for four regions (Northern Norrland, Southern Norrland, Svealand, and Götaland), see Figure 4, and for the different seasons. These regions are the three traditional lands of Sweden: Norrland, Svealand, and Götaland, where, due to its larger size, Norrland is divided in a northern part (including the provinces Lappland, Norrbotten, and Västerbotten) and a southern part (including the provinces Jämtland, Ångermanland, Medelpad, Härjedalen, Hälsingland, and Gästrikland). The regions divide Sweden in four approximately equally large parts from north to south. These regions are familiar and therefore relatable to the Swedish public.

The standardized deviation ($\hat{x}$) from the long-term monthly mean (μ_m) over the entire time period is estimated as:

$$\hat{x}_{i,y,m} = \frac{x_{i,y,m} - \mu_{i,m}}{\sigma_{i,m}}, \tag{1}$$

where i indicate the time series identity, m, the month, and y, the year. The mean value is formulated as:

$$\mu_{i,m} = \frac{\sum_{y=y_1}^{y_N} x_{i,y,m}}{N}, \tag{2}$$

where the start year $y_1 = 1860$, the end year $y_N = 2021$, and N, the number of years. The long-term monthly standard deviation σ_m is formulated as:

$$\sigma_{i,m} = \sqrt{\frac{\sum_{y=y_1}^{y_N} \left(x_{i,y,m} - \mu_{i,m}\right)^2}{N}}. \tag{3}$$

The standardized deviations in each of the different normal periods from 1871–1900 to 1991–2020 are then calculated.

3 | RESULTS

3.1 | Merging of time series

In this section, the results of the automatic merging of the data set are presented. An example of a merging scheme is presented in Figure 3. All the merging schemes are listed in Supplementary Material 1, Table S1 and a list of the stations with information on their locations is included in Supplementary Material 2. The geographical coverage of the merged and the original network are

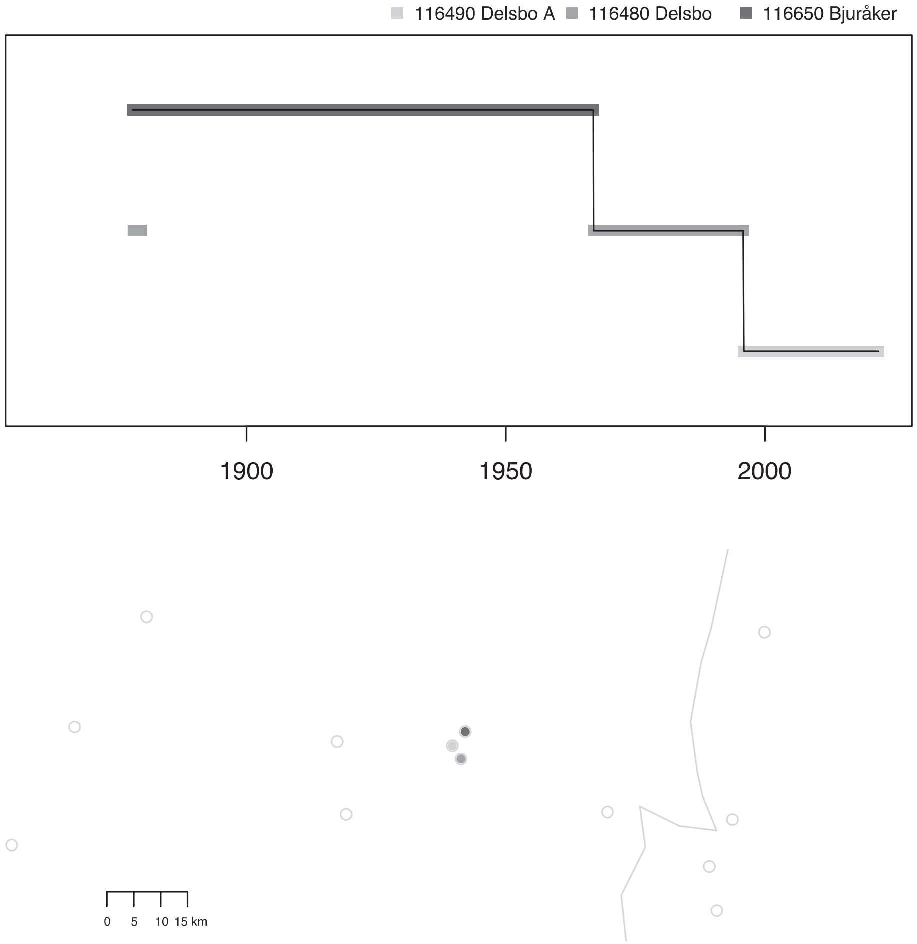

FIGURE 3 Example of a merging scheme. The top panel shows the merging scheme of Delsbo A as a thin black line: 116650 Bjuråker 1878-01 – 1966-12, 116480 Delsbo 1967-01 – 1996-11, and 116490 Delsbo A 1995-12 – 2021-12. The bars indicate observational data. The bottom panel shows a map of Central Sweden with the location of the three stations as dots, Delsbo A, Delsbo, and Bjuråker. The empty circles indicate locations of other stations in the full data set.

presented in Figure 4. The original data set (light grey dots) consists of 961 time series (417,972 data), of which 686 has at least 15 years data coverage (394,161 data), see Table 1. 939,423 values are missing and must therefore be interpolated in order to have a complete data set. The stations with time series of at least 15 years of data (henceforth called the 'Minimum 15 a data set') are indicated by mid-grey circles in Figure 4. The network of stations with red circles correspond to the output network if the original network was homogenized with HOMER. The merged data set (dark grey dots) consists of 456 time series, where data from 836 stations are included (395,674 data, 497,790 missing values). Figure 4 indicate that there is no large difference in geographical distribution between the merged and Minimum 15 a data sets.

The data coverage over the time period in the merged and the original network are shown in Figure 5a. The difference between the merged and the Minimum 15 a data set is largest from the early 2010's to present where there are 20–30 more time series in the merged data set. From the mid-1990's to the early 2000's the merged data set has 15–25 fewer stations than the 'Minimum 15 a data set'. Histograms of the length of the time series for the merged and the original network are shown in Figure 5b. The 'Minimum 15 a data set' corresponds to the red bars. The merged data set has 257 series of its total 456 series that cover at least 60 years of data, which corresponds to the length of two standard normal periods (WMO, 2017). The original data set has 173 series of its total 961 series that covers at least 60 years of data.

As an example of the sensitivity of the merging thresholds, a merging scheme with the thresholds used in Kuya *et al.* (2020) (horizontal distance threshold value of 10 km and maximum different in altitude of 100 m) results in a data set with 535 merged time series from 749 stations with 390,783 data and 649,257 missing values (not shown).

3.2 | Homogenization

In this section, the results of the homogenization of the merged data set are presented. The distribution of the number of break-points in each of the 456-time series is presented as a histogram in Figure 6a. For 22 time series, no break-point is detected. These series are thus found to be homogeneous. The median length of the homogeneous series is 27 years, the corresponding length for all-time series in the data set is 61 years. The longest homogeneous time series is Bjuröklubb, which has 142 years of data. The homogeneous time series has fewer potential break-points listed in the meta data (applied merging points included) compared to all series in the data set: On average two potential break-points are listed in the meta data versus five for all series. The stations with homogeneous series are significantly farther north compared to all stations in the data set. 50% of the stations with homogeneous series are situated north of 63.1°. Correspondingly, 50% of all the stations are situated north of 59.9°. The stations with homogeneous series also have a longer median distance to their reference stations (on average 120 km), compared to the entire data set (on average 96 km), indicating that these stations are located in more sparse regions.

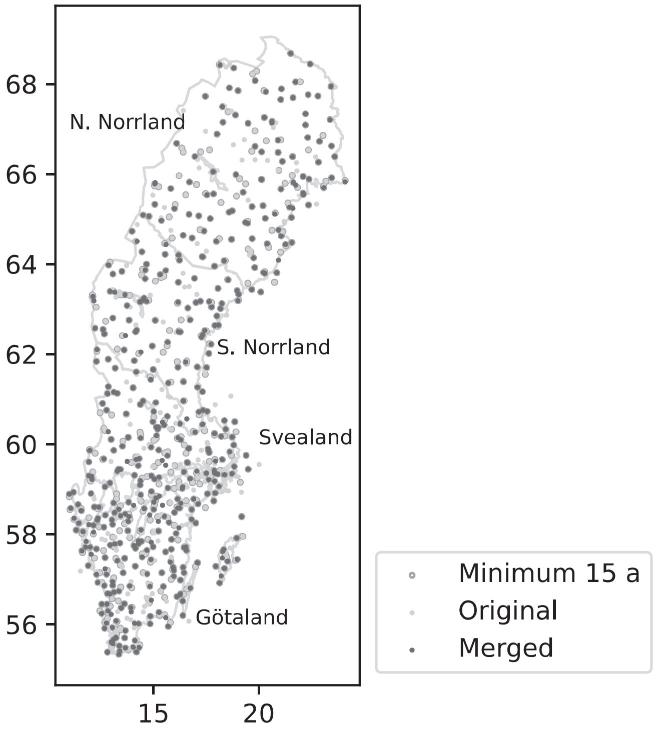

FIGURE 4 Map of the stations, the light grey dots represent the situation of the stations in the original network, the mid-grey circles indicate the stations in the original network with at least 15 years of data and the dark grey dots represent the situation of the head stations in the merged network

The most common number of detected break-points are three break-points, which are detected in 90 time series (21% of all time series). Thirty-three time series have more than seven detected break-points. All the detected break-points are listed in Table S3 in Supplementary Material 1 and are included as a list in Supplementary Material 2.

The mean time between break-points in each time series is presented as a histogram in Figure 6b. Across the series the median time per break-point is 17 years, corresponding to six break-points per 100 years of data. Venema *et al.* (2012) state that there are typically 15–20 years per break-point in European climate records. Two hundred fifteen time series in the current data set (47%) has 11–23 years per break-point.

In the current homogenized data set, the number of detected break-points are 120% of the number of detected break-points in the 2014 data set, for the common time period (1860–2013) and the common 35 stations. 39% of the break-points detected in the 2014 data set are also detected in the new data set (within 2 years). 47% of the break-points detected in the new data set were also detected in the previous data set. The new data set has a smaller share of its break-points early: 25% of the detected break-points are detected before 1939. The corresponding year for the 2014 data set is 1895. Note that the merging schemes of the 2014 and the 2022 homogenized data sets are not identical, see Figures S1–S21 in Supplementary Material 1. The raw data in the 2014 data set is therefore not simply a subset of the raw data in the 2022 data set.

For 34% of the break-points listed in meta data (including merging points) were detected in the new data set. 37% of the break-points detected in the new data are supported by meta data (or the merging scheme).

In the 2022 data set, 60% of the raw data are corrected in the homogenization process, see Figure 7a. 42% has a correction ($T_{\mathrm{hom}} - T_{\mathrm{obs}}$) of $\pm 0.5°$C or smaller, 3% has a correction of $\pm 1°$C or larger. The largest correction is $-3.0°$C in the month of February from 1958 to 1964 in the Gielas time series. The largest positive correction is $+2.9°$C in December 1985 in the Munsvattnet time series. 37% of the raw data has negative correction, 24% positive correction.

TABLE 1 Total number of time series in the original data set, number of time series in the data set longer than 15 years, and the number of times series in the merged data set

Set	Number of time series	Number of data	Number of missing data
Original	961	417,972	1,450,212
Minimum 15 a	686	394,161	939,423
Merged	456	395,674	490,790

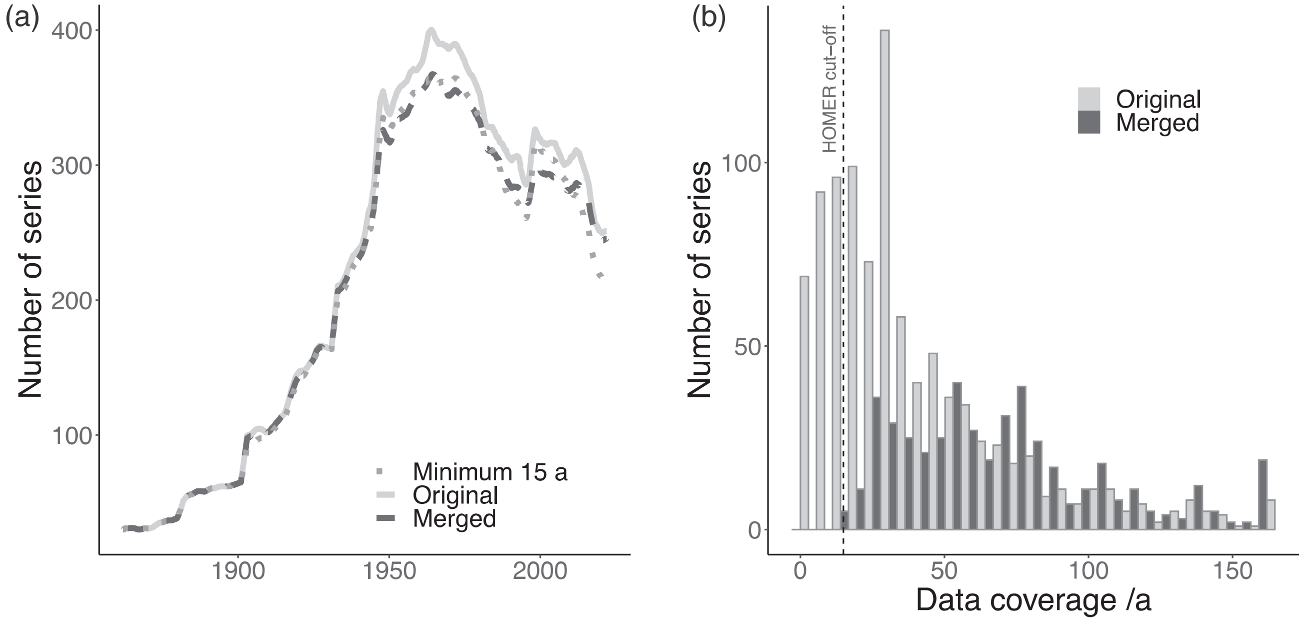

FIGURE 5 Data coverage (1860–2021) of the original data set (light grey) and the merged data set (dark grey). All original time series with at least 15 years of data (mid-grey) are series with minimum length for homogenization with HOMER. (a) The number of time series with data in the months from January 1860 to December 2021, a 25-year running mean filter is applied on the data for readability. (b) Histogram of the data coverage in the time series, the vertical dahed line indicate the shortest time series that can be used in HOMER.

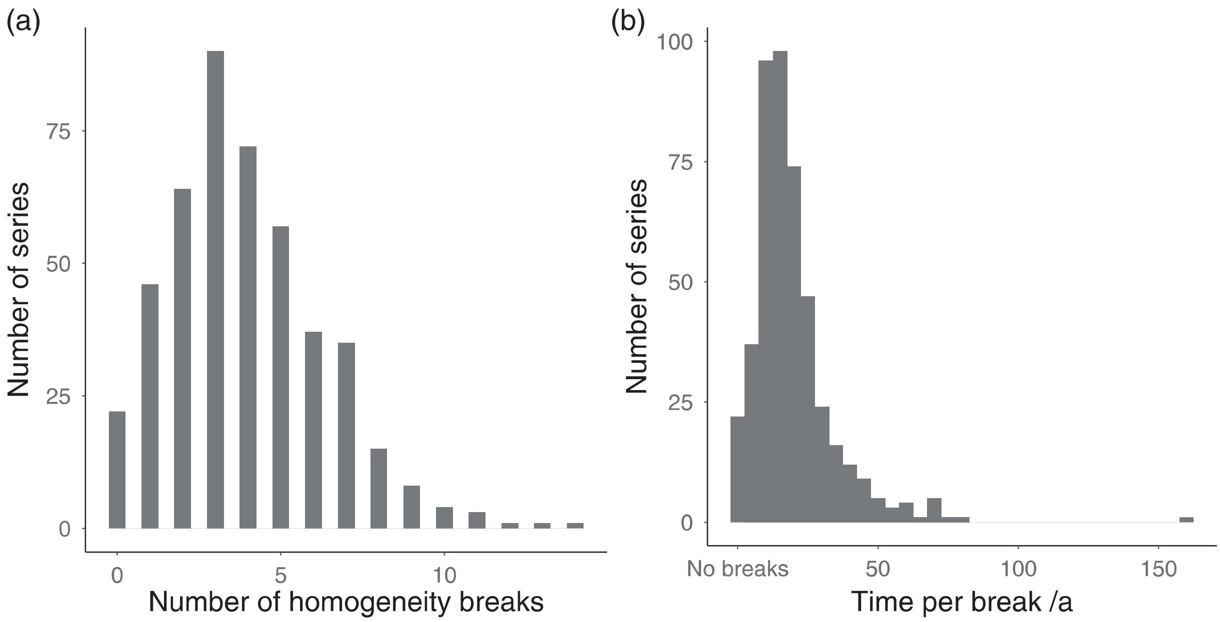

FIGURE 6 Histogram over number of break-points. (a) Number of break-points per time series. (b) Mean time between break-points

In the 2014 data set, 72% of the raw data are corrected in the homogenization process, compared to 53% for the same time period and over the same stations in the 2022 data set. 53% has a correction ($T_{\text{hom}} - T_{\text{obs}}$) of $\pm 0.5°$C or smaller, 4% has a correction of $\pm 1°$C or larger (compared to 38% and 1% respectively in the corresponding part of the 2022 data set). 43% of the raw data has negative correction, 29% positive correction (compared to 34% and 19% respectively in the corresponding part of the 2022 data set).

Larger corrections are more frequent in the earlier parts of the data, see Figure 7b. The distribution and the means of the corrections are also shifted towards larger negative corrections in the earlier parts of the data. The mean absolute correction in the period 1871–1900 is $(0.320 \pm 0.004)°$C

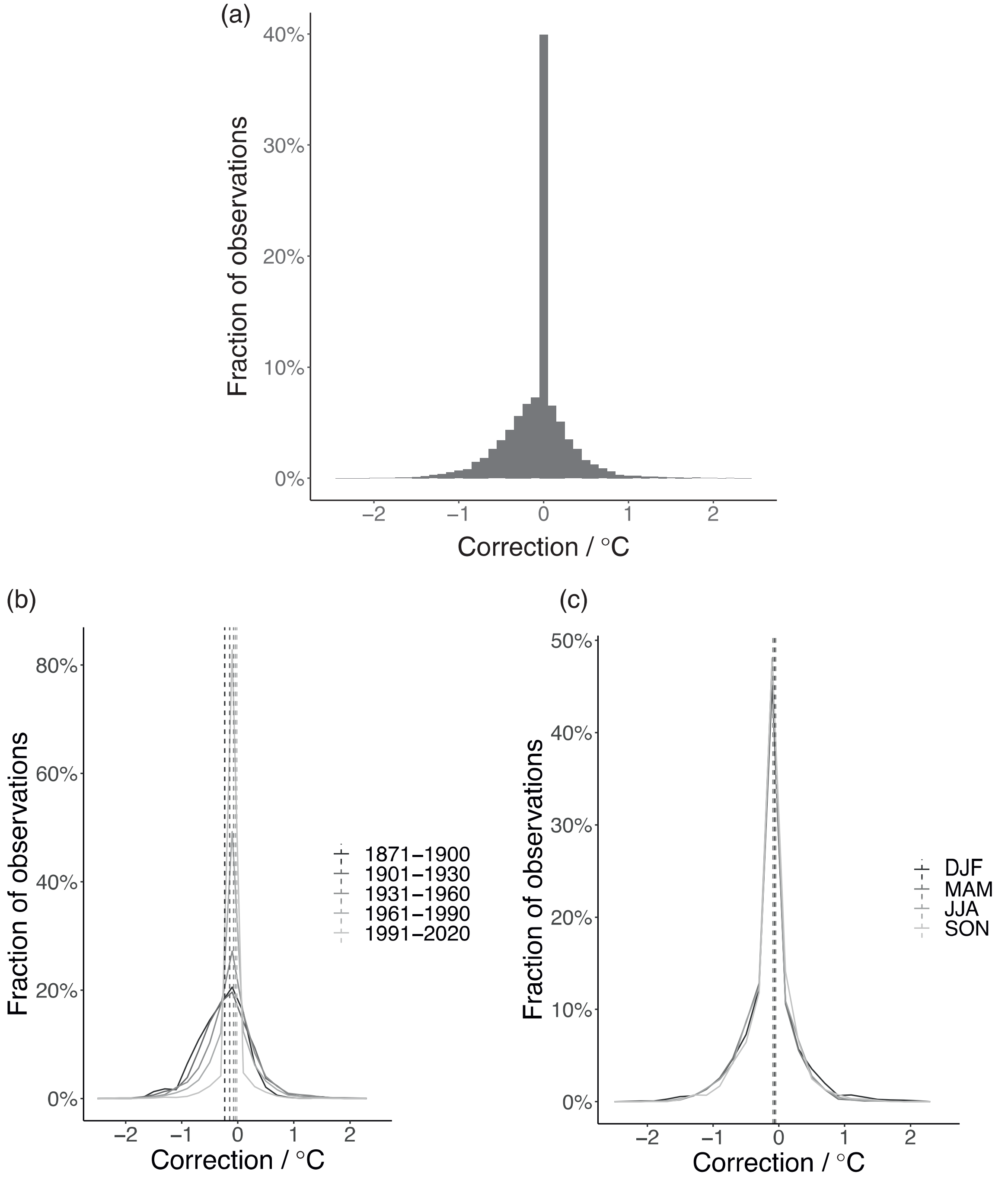

FIGURE 7 Corrections ($T_{\text{hom}} - T_{\text{obs}}$) in the 2022 homogenization data sets. (a) All data. (b) Normal periods. (c) Seasons. Winter: December, January, and February (DJF). Spring: March, April, and May (MAM). Summer: June, July, and August (JJA). Autumn: September, October, and November

and the mean correction −0.213(5)°C. The absolute correction for the 1991–2020 period is (0.030 ± 0.001)°C and the mean correction is (−0.006 ± 0.001)°C.

The largest mean correction of the full period is in the summer months June–August ([−0.050 ± 0.002]°C), the smallest mean correction is in the winter months December–February ([−0.033 ± 0.002]°C), see Figure 7c.

An example of a set of references are presented in Figure 8. An example of a homogenized time series are presented in Figure 9. All-time series are available as plain text files in Supplementary Material 2.

3.3 | Data analysis

In this section, some climatological features of the homogenized data set are presented. Based on the new homogenized data set, the trend for the period 1860

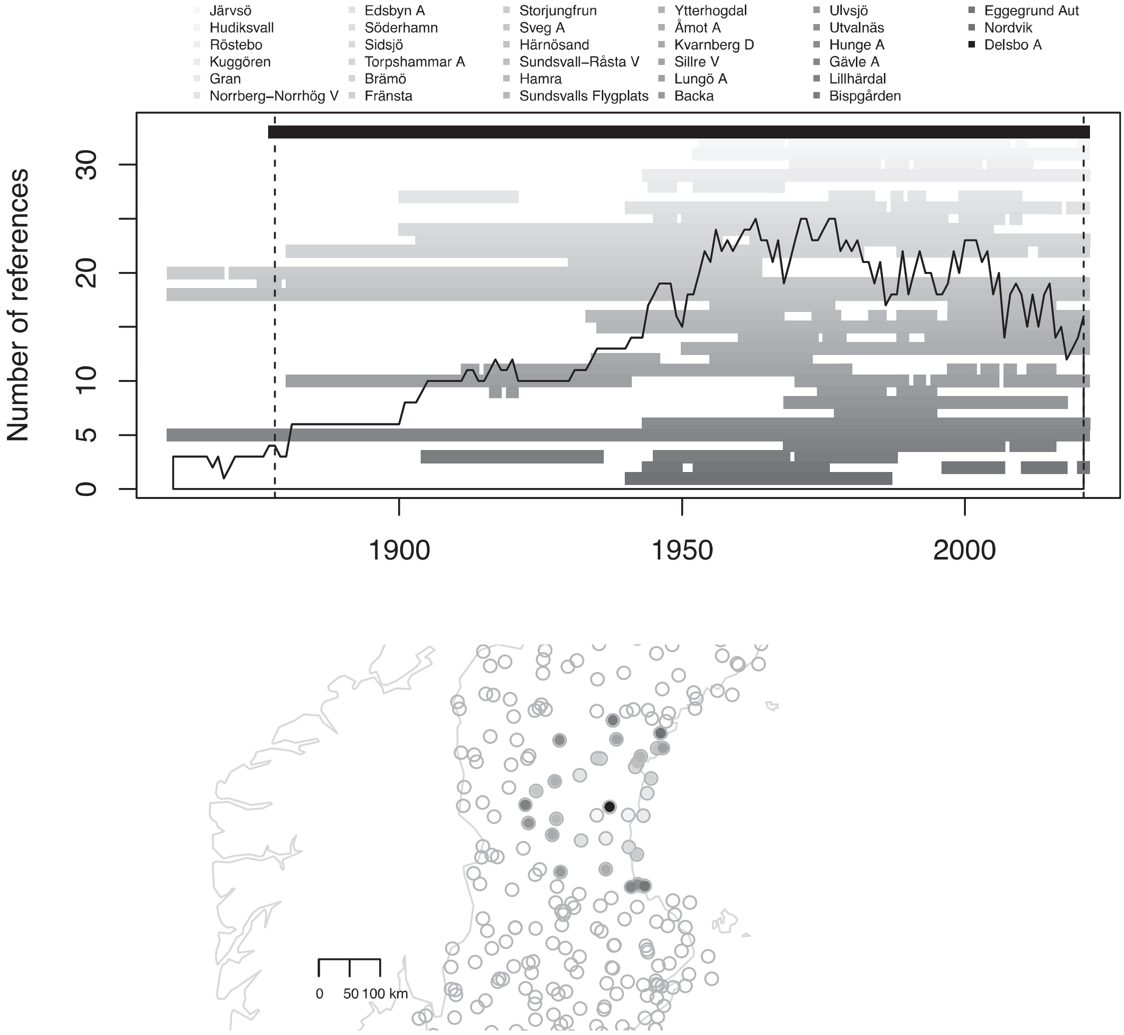

FIGURE 8 Example of a reference series set. The top panel shows the reference series set for the candidate series Delsbo A (black). The bars indicate observational data in the different series. The dashed vertical lines indicate the start and end of the candidate series, the thin black line describe the total number of active stations in the reference set. The bottom panel shows a map of Central Sweden with the location of the stations in the reference set as circles according to the shade of grey of the top panel. The empty circles indicate locations of other stations in the full data set

through 2021 is estimated at $(0.13 \pm 0.03)^{\circ}$C decade, see Figure 10. The corresponding raw data has an estimated trend of $(0.11 \pm 0.03)^{\circ}$C decade and 2014 data set $(0.12 \pm 0.03)^{\circ}$C decade. The uncertainty is given as the 95% confidence interval. As a comparison, an OLS linear regression of the HadCRUT5 global annual mean temperature data (Morice *et al.*, 2021) 1860–2021 gives a trend of $(0.07 \pm 0.01)^{\circ}$C decade.

At the precision of the uncertainty level, the estimated trend and its standard deviation for the entire data set and the 35-time series corresponding to those included in the previous homogenized data set are identical, both for the homogenized and the raw gap filled data.

Since 1981, used as the starting point of a late warming trend by NOAA (2020), the 2022 data set and the 2014 data set has estimated trends of $(0.49 \pm 0.22)^{\circ}$C decade and $(0.50 \pm 0.22)^{\circ}$C decade respectively. Correspondingly, the raw gap filled data shows a trend of $(0.47 \pm 0.22)^{\circ}$C decade. Again, comparing with the global mean, the HadCRUT5 data has a trend of $(0.20 \pm 0.02)^{\circ}$C decade in the period 1981–2021. The confidence intervals of the trend estimates are given as the 95% confidence interval level.

The regional and seasonal linear trends (1860–2021) are presented in Table 2. For the longer trend (1860–2021), the rate of warming in the spring months (March, April, and May) is significantly higher compared to the annual trend. There is no significant difference in the rate of warming between the different regions. For the short trend (1981–2021), the rate of warming in winter is largest, but the variability is large.

The frequency of deviations, as defined in Equation (1) in the raw gap filled and homogenized 2022 data sets as they occur in the different normal periods are presented as a box plot in Figure 11. (54 ± 4) % of cold extremes (≤−3 SD) in the homogenized data occurred in 1871–1900, whilst (6 ± 1) % occurred in 1991–2020. Corresponding values in the raw gap filled data is (47 ± 3) % and (8 ± 1) % respectively. (1 ± 1) % of warm extremes (≥3 SD) in the homogenized data occurred in 1871–1900, whilst (91 ± 5) % occurred in 1991–2020. Corresponding values in the raw gap filled data is (3 ± 1) % and (86 ± 5) % respectively. Confidence intervals of frequency of deviations are given at 95% level and are calculated by the bootstrap method (Efron, 1992).

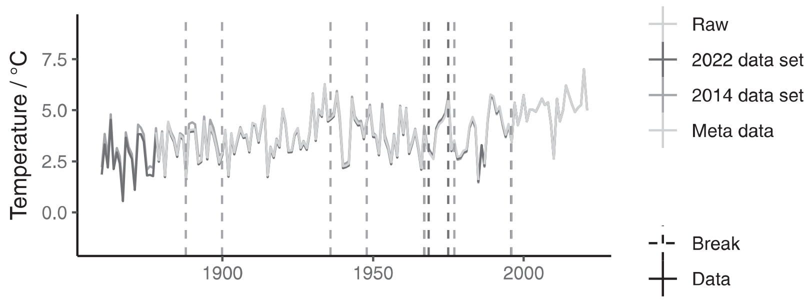

FIGURE 9 Example of a homogenized series (Delsbo A). The solid lines represent homogenized (dark grey: 2022 data set and mid-grey: 2014 data set) and raw (light grey) annual mean temperature values. The vertical dashed lines represent detected break-points (dark grey: 2022 data set and mid-grey: 2014 data set) and suggested break-points (light grey: Meta data and applied merging points)

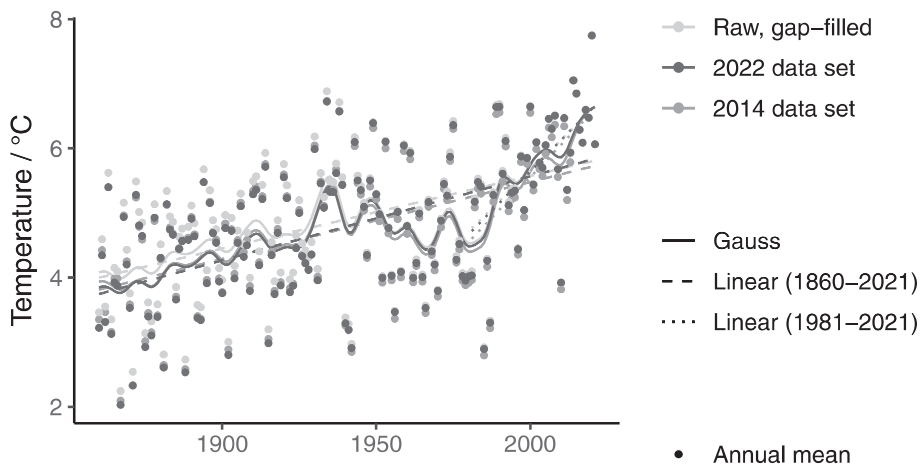

FIGURE 10 Raw gap filled and two sets of homogenized mean annual temperatures, averaged over 35 selected time series. Linear annual mean temperature trend from 1860 through 2021 (dashed) and from 1981 through 2021 (dot-dashed), a Gaußian fit of the data are represented as a solid line and the annual mean temperature are represented as dots

TABLE 2 Regional and seasonal linear trends 1860–2021 (upper rows) and 1981–2021 (lower rows), uncertainties are gives as 95% confidence interval

	/°C decade				
	National	**Northern Norrland**	**Southern Norrland**	**Svealand**	**Götaland**
Annual	0.13 ± 0.03	0.12 ± 0.04	0.13 ± 0.03	0.14 ± 0.03	0.12 ± 0.03
	0.47 ± 0.22	0.48 ± 0.23	0.47 ± 0.23	0.48 ± 0.23	0.47 ± 0.21
Winter	0.13 ± 0.08	0.13 ± 0.09	0.15 ± 0.09	0.14 ± 0.08	0.11 ± 0.07
	0.68 ± 0.60	0.77 ± 0.68	0.77 ± 0.66	0.67 ± 0.63	0.58 ± 0.56
Spring	0.18 ± 0.04	0.18 ± 0.05	0.18 ± 0.04	0.20 ± 0.04	0.17 ± 0.04
	0.39 ± 0.26	0.37 ± 0.29	0.36 ± 0.26	0.40 ± 0.27	0.40 ± 0.27
Summer	0.10 ± 0.03	0.08 ± 0.04	0.10 ± 0.03	0.11 ± 0.03	0.10 ± 0.03
	0.43 ± 0.23	0.43 ± 0.23	0.38 ± 0.25	0.44 ± 0.25	0.45 ± 0.25
Autumn	0.11 ± 0.04	0.10 ± 0.05	0.11 ± 0.04	0.12 ± 0.04	0.11 ± 0.03
	0.44 ± 0.27	0.42 ± 0.34	0.41 ± 0.29	0.44 ± 0.28	0.46 ± 0.25

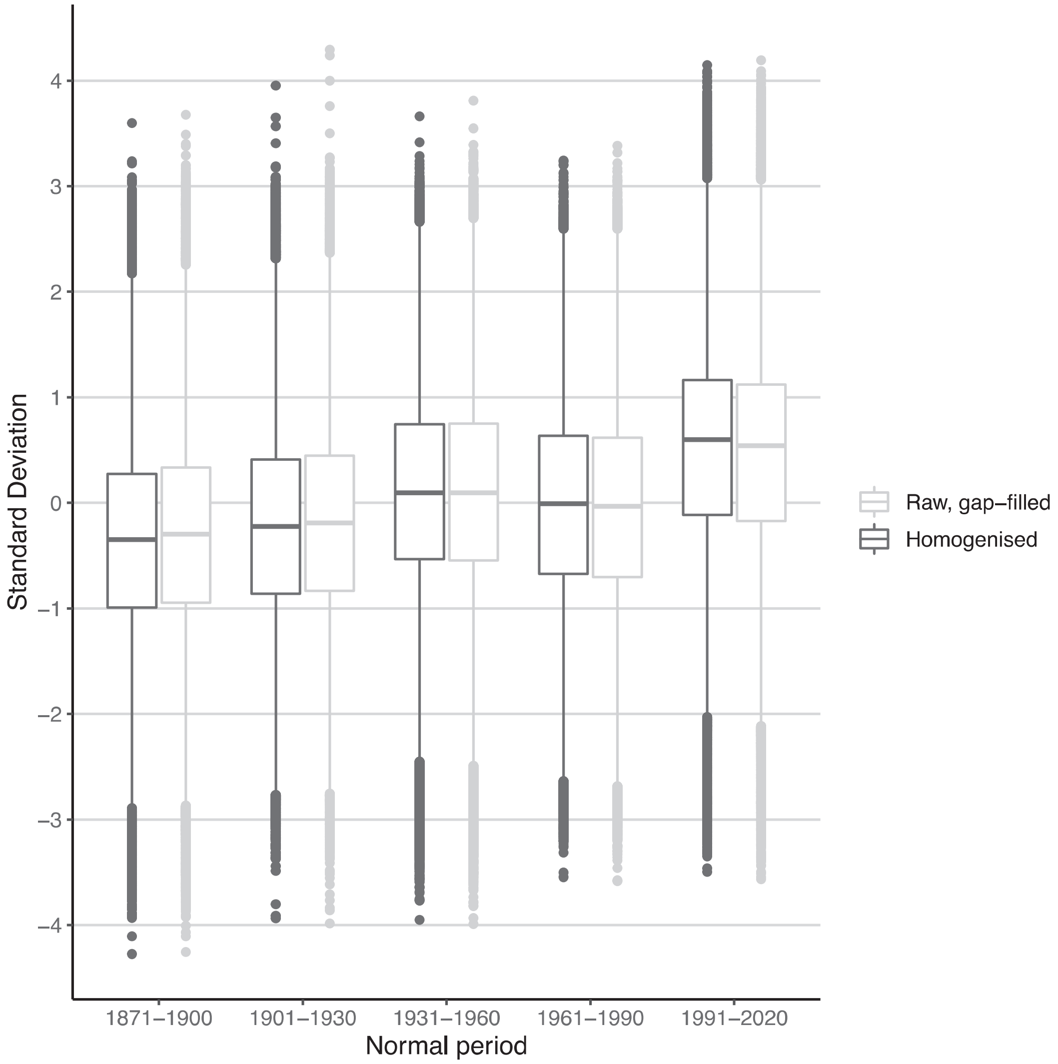

FIGURE 11 Frequency of standardized deviation in all time series from the mean of specific months over the entire data sets, the middle of the boxes represent the median of the data whilst the edges represent the first and third quartiles respectively, the vertical line indicate 1.5 times the interquartile range above or below the first and third quartile respectively, and the dots represent values above or below the range of the vertical line

The average monthly standard deviation in the raw gap filled and the homogenized data sets are 2.10°C and 2.12°C respectively.

4 | DISCUSSION

The merging of data series enables the use of a larger part of the observational data set, since about 30% of the time series in the observational data set includes less than 15 years of data. On the other hand, overlapping data in merged time series are discarded. In the current merged scheme, there is a small gain in data in the merging of time series (about 0.4%). More prominent is the decrease in the fraction of missing data, from 70% in the Minimum 15 a data set to 55% in the merged data set. Missing data must be interpolated in the homogenization process. There is also an increase of long time series, for example, the number of time series that cover at least 60 years of data increases from 173 to 257 when merging of time series is applied.

The optimal sensitivity of the homogenization method is an open question. Joelsson *et al.* (2022) concluded that a more sensitive method will detect a larger portion of the actual break-points ('true positives'), but also misinterpret more of the natural variations as break-points ('false positives') and thereby infer false corrections. A less sensitive method will, on the other hand, detect both fewer true and false positives. The sensitivity of the current method is higher than the previous SNHT-based method (applied in the 2014 data set) with 20% more break-points. Inconsistency of homogenization is discussed in O'Neill *et al.* (2022), where the inclusion of meta data in homogenization methods is encouraged. Meta data is currently an input to the Bart tool used here, but much of the weather station inspection protocols remains to be digitized. Efforts to digitize more of these documents, albeit labour intensive, could potentially improve the consistency of future homogenizations. The fraction of detected break-points supported by meta data is similar to the results of O'Neill *et al.* (2022), although the time windows when break-points are considered to be supported are not identical.

The frequency of break-point (one break-point every 17 years) correspond well to the typical frequency of European climate records (one break-point every 15–20 years) as stated by Venema *et al.* (2012) and therefore gives confidence to the sensitivity of the method.

The series found to be homogeneous are relatively short and have relatively few potential break-points in the meta data (including merging points), indicating that the apparent homogeneity is true homogeneity rather than shortcomings of the homogenization method. On the other hand, these stations seem to be located in sparse regions (farther north and with longer distance to their reference stations), which could indicate that the homogenization method struggles with detecting break-points in these regions.

Compared to the 2014 data set the number of corrections of data are fewer, especially large corrections (>1°C) are less frequent. The 2022 data set has a larger negative bias in its corrections compared to the 2014 data set, which are consistent with the slightly steeper warming trend in the 2022 data set, compared to the 2014 data set.

The correction has larger amplitude in the earlier parts of the series than in the more recent parts and are on average negative (observations are warmer than the assumed real value). The larger corrections in the earlier parts are expected since the homogenization method fixes the time series to the last observed point. The shift towards larger negative corrections could be due to inadequate measurement techniques in the early parts of the data sets. Inadequate measurement techniques of temperature tend to return especially too warm summertime measurements (Moberg *et al.*, 2003; Böhm *et al.*, 2010). In the current data set the average correction is significantly more negative in the summertime than in the wintertime.

The estimated shift in trend when data is homogenized is not significant, especially the estimated shift in trend over the last 40 years (from 1981) is very small compared to the uncertainty. The trend itself is significant in both the homogenized and the raw data, both over the entire period and over the last 40 years. The estimated rate of warming is about double the global trend which are consistent with the estimated Arctic amplification for the region (Smith *et al.*, 2019). The rate of warming in the more recent period (1981–2021) is almost four times as high as the rate of the entire period. This is consistent with the estimate represented in Gulev *et al.* (2021).

The stronger trend in winter and spring compared to summer and autumn is consistent with a recent report of climate change in the Baltic region (Meier *et al.*, 2022). Kjellström *et al.* (2022), who estimated the mean seasonal temperature in the normal periods 1961–1990 and 1991–2020 from ERA5 data, found a similar seasonal pattern. They argue that the increased wintertime temperatures from 1961 partly can be attributed to generally more zonal conditions.

There is no significant difference in trends between the regions. Meier *et al.* (2022) divided the Baltic region into two parts; north and south of the 60th parallel. They report no difference in annual mean temperature trends between the northern and the southern parts. In seasonal trends, however, there are some differences reported, albeit no confidence intervals are given.

Cold extremes are more common and warm extremes are less common in the earliest period (1871–1900) than in the most recent period (1991–2020). There is a small but significant shift in the frequency of extremes when the data is homogenized. The shift is consistent with the shift in trend and the distribution of corrections where the homogenization process finds more of the observations to be too warm than too cold.

5 | CONCLUSIONS

Monthly average temperature data (1860–2021) from 836 Swedish weather stations are merged into 456 time series. The data set is homogenized using a recently developed automatic version of the HOMER homogenization tool. Twenty-two of the merged times series were found to be homogeneous. The median number of years of data per break-point is 17 years.

The trend of the mean annual temperatures averaged over the entire data set for the period 1860 through 2021 is (0.13 ± 0.03)°C decade. The homogenization does not significantly change the trend. The rate of warming is larger in spring and winter than in summer and autumn. There is no significant regional difference. The rate of warming since 1981 is almost four times the average rate of the entire period.

Extremely warm months (≥3 SD) are more frequent in the most recent period than in the earliest period, extremely cold months less frequent. Homogenization increases these differences.

6 | OUTLOOK

Future work includes the homogenization of monthly averages of daily maximum and minimum temperature series, expansion of the monthly data to daily data, and homogenization of precipitation time series.

Using the recently released SMHI GridClim (Andersson *et al*., 2021), a gridded data set of several of meteorological parameters over the Swedish domain, it could be possible to check the area for which the climatology in a certain point is representative and thereby improve of the vertical and horizontal thresholds in the merging method.

ACKNOWLEDGEMENTS

We would like to thank Olivier Mestre, Peter Domonkos, Jose Guijarro, Enric Aguilar, and Brigitte Dubuisson for the wonderful work of HOMER. We would especially like to thank Olivier Mestre for the permission to use and alter the HOMER-script. Futhermore, we would like to thank Christophe Sturm, Johan Södling, Lennart Wern, Sverker Hellstöm, Anna Jansson, and Thomas Carlund for insightful advices and support during this project.

ORCID

L. Magnus T. Joelsson https://orcid.org/0000-0002-0287-4962
Erik Engström https://orcid.org/0000-0002-6207-6460
Erik Kjellström https://orcid.org/0000-0002-6495-1038

REFERENCES

Alexandersson, H. (1986) A homogeneity test applied to precipitation data. *Journal of Climatology*, 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: A Journal of the Royal Meteorological Society*, 17, 25–34.

Andersson, S., Bärring, L., Landelius, T., Samuelsson, P. and Schimanke, S. (2021) SMHI gridded climatology. *Report RMK 118*, SMHI.

Auchmann, R. and Brönnimann, S. (2012) A physics-based correction model for homogenizing sub-daily temperature series. *Journal of Geophysical Research: Atmospheres*, 117, 1–13.

Böhm, R., Jones, P.D., Hiebl, J., Frank, D., Brunetti, M. and Maugeri, M. (2010) The early instrumental warm-bias: a solution for long central european temperature series 1760–2007. *Climatic Change*, 101, 41–67.

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, 405–425.

Coll, J., Curley, C., Walsh, S. and Sweeney, J. (2014) Ireland with homer. In: *Proceedings of the 8th Seminar for Homogenization and Quality Control in Climatological Databases and 3rd Conference on Spatial Interpolation in Climatology and Meteorology, Climate Data and Monitoring WCDMP No*, (Vol. 84, pp. 23–45). Geneva, Switzerland: WMO.

Coscarelli, R., Caroletti, G.N., Joelsson, M., Engström, E. and Caloiero, T. (2021) Validation metrics of homogenization techniques on artificially inhomogenized monthly temperature networks in Sweden and Slovenia (1950–2005). *Scientific Reports*, 11, 1–18.

Dienst, M., Lindén, J., Engström, E. and Esper, J. (2017) Removing the relocation bias from the 155-year Haparanda temperature record in northern europe: removing the relocation bias from a 155-year temperature record. *International Journal of Climatology*, 37, 4015–4026.

Domonkos, P. (2022) ACMANT v.4: Scientific content and operation of the software.

Efron, B. (1992) Bootstrap methods: another look at the jackknife. In: *Breakthroughs in Statistics* (pp. 569–593). New York, NY: Springer.

Gulev, S.K., Thorne, P.W., Ahn, J., Dentener, F.J., Domingues, C. M., Gerland, S., Gong, D., Kaufman, D.S., Nnamchi, H.C., Quaas, J., Rivera, J.A., Sathyendranath, S., Smith, S.L., Trewin, B., von Schuckmann, K. and Vose, R.S. (2021) Changing state of the climate system. In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R. and Zhou, B. (Eds.) chap. 2*Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY, USA: Cambridge University Press, pp. 287–422.

Joelsson, L.M.T., Sturm, C., Södling, J., Engström, E. and Kjellström, E. (2022) Automation and evaluation of the interactive homogenization tool homer. *International Journal of Climatology*, 42, 2861–2880.

Kjellström, E., Hansen, F. and Belušić, D. (2022) Contributions from changing large-scale atmospheric conditions to changes in scandinavian temperature and precipitation between two climate normals. *Tellus. Series A, Dynamic meteorology and oceanography*, 74, 204–221.

Kuya, E.K., Gjelten, H.M. and Tveito, O.E. (2020) Homogenization of Norway's mean monthly. *Network*, 2, 1–55.

Meier, H., Kniebusch, M., Dieterich, C., Gröger, M., Zorita, E., Elmgren, R., Myrberg, K., Ahola, M.P., Bartosova, A., Bonsdorff, E., Börgel, F., Capell, R., Carlén, I., Carlund, T., Carstensen, J., Christensen, O.B., Dierschke, V., Frauen, C., Frederiksen, M., Gaget, E., Galatius, A., Haapala, J.J., Halkka, A., Hugelius, G., Hünicke, B., Jaagus, J., Jüssi, M., Käyhkö, J., Kirchner, N., Kjellström, E., Kulinski, K., Lehmann, A., Lindström, G., May, W., Miller, P.A., Mohrholz, V., Müller-Karulis, B., Pavón-Jordán, D., Quante, M., Reckermann, M., Rutgersson, A., Savchuk, O.P., Stendel, M., Tuomi, L., Viitasalo, M., Weisse, R. and Zhang, W. (2022) Climate change in the Baltic Sea region: a summary. *Earth System Dynamics*, 13, 457–593.

Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Lebarbier, E., Böhm, R., Aguilar, E., Guijarro, J.A., Vertacnik, G., Klancar, M., Dubuisson, B. and Stepanek, P. (2013) Homer: a homogenization software–methods and applications. *Időjárás: Quarterly Journal of the Hungarian Meteorological Service*, 117, 47–67.

Moberg, A., Alexandersson, H., Bergström, H. and Jones, P.D. (2003) Were southern swedish summer temperatures before 1860 as warm as measured? *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 23, 1495–1521.

Morice, C.P., Kennedy, J.J., Rayner, N.A., Winn, J., Hogan, E., Killick, R., Dunn, R., Osborn, T., Jones, P. and Simpson, I. (2021) An updated assessment of near-surface temperature change from 1850: the hadcrut5 data set. *Journal of Geophysical Research: Atmospheres*, 126, e2019JD032361.

NOAA. (2020) State of the climate: Global climate report for 2020. In: Tech. Rep. Washington, DC: National Oceanic and Atmospheric Administration. Available at: https://www.ncdc.noaa.gov/sotc/global/202013/supplemental/page-3 [Cited 14 March 2022].

O'Neill, P., Connolly, R., Connolly, M., Soon, W., Chimani, B., Crok, M., de Vos, R., Harde, H., Kajaba, P., Nojarov, P., Przybylak, R., Rasol, D., Skrynyk, O., Skrynyk, O., Štěpánek, P., Wypych, A. and Zahradníček, P. (2022) Evaluation of the homogenization adjustments applied to european temperature records in the global historical climatology network dataset. *Atmosphere*, 13, 285.

Ribeiro, S., Caineta, J. and Costa, A.C. (2016) Review and discussion of homogenization methods for climate data. *Physics and Chemistry of the Earth, Parts A/B/C*, 94, 167–179.

SMHI. (2021) Klimatindikator - temperatur [homepage on the Internet]. Available at: https://www.smhi.se/klimat/klimatet-da-och-nu/klimatindikatorer/klimatindikator-temperatur-1.2430 [Cited 14 March 2022].

SMHI. (2022a) Ladda ner meteorologiska observationer [homepage on the Internet]. Available at: https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer [Cited 23 March 2022].

SMHI. (2022b) Månadens väder och vatten i Sverige [homepage on the Internet]. Available at: https://www.smhi.se/klimat/klimatet-da-och-nu/manadens-vader-och-vatten-sverige [Cited 14 March 2022].

Smith, D.M., Screen, J.A., Deser, C., Cohen, J., Fyfe, J.C., García-Serrano, J., Jung, T., Kattsov, V., Matei, D., Msadek, R., Peings, Y., Sigmond, M., Ukita, J., Yoon, J.H. and Zhang, X. (2019) The polar amplification model intercomparison project (pamip) contribution to cmip6: investigating the causes and consequences of polar amplification. *Geoscientific Model Development*, 12, 1139–1164.

Sun, Y., Zhang, X., Ren, G., Zwiers, F.W. and Hu, T. (2016) Contribution of urbanization to warming in China. *Nature Climate Change*, 6, 706–709.

Venema, V., Trewin, B., Wang, X., Szentimrey, T., Lakatos, M., Aguilar, E., Auer, I., Guijarro, J., Menne, M., Oria, C., Louamba, W. and Rasul, G. (2020) *Guidelines on Homogenization, 2020 edition, wmo-no. 1245*. Geneva: World Meteorological Organization.

Venema, V.K., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Muller-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., Duran, M.P., Likso, T., Esteban, P. and Brandsma, T. (2012) Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8, 89–115.

Vertačnik, G., Dolinar, M., Bertalanič, R., Klančar, M., Dvoršek, D. and Nadbath, M. (2015) Ensemble homogenization of slovenian monthly air temperature series. *International Journal of Climatology*, 35, 4015–4026.

WMO. (2017) WMO guidelines on the calculation of climate normals.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Joelsson, L. M. T., Engström, E., & Kjellström, E. (2022). Homogenization of Swedish mean monthly temperature series 1860–2021. *International Journal of Climatology*, 1–15. https://doi.org/10.1002/joc.7881