https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

Automated Homogenisation of Early Instrumental Temperature Data at the Global Scale Using the HCLIM Dataset

Elin Lundstad[1,2,3], Andreas Dobler[1], Michael Taylor[4] and Magnus Joelsson[5]

Correspondence to: Elin Lundstad (elinl@met.no)

[1]Norwegian Meteorological Institute (MetNorway), Henrik Mohns Plass 1, 0371 Oslo, Norway
[2]Institute of Geography, University of Bern (UniBe), Hallerstrasse 12, CH-3012 Bern, Switzerland
[3]Oeschger Centre for Climate Change Research (OCCR), University of Bern, Hochschulstrasse 4, CH-3012 Bern, Switzerland
[4]Climatic Research Unit (CRU), Department of Environmental Sciences, University of East Anglia, Research Park, Norwich NR4 7TJ, United Kingdom
[5]Swedish Meteorological and Hydrological Institute (SMHI), 601 76 Norrköping, Sweden

Abstract

Early instrumental observations provide a unique opportunity to investigate climate variability before the establishment of modern meteorological networks. However, their effective use is often hindered by incomplete records, uneven spatial coverage, and non-climatic inhomogeneities. The HCLIM dataset provides a recently compiled global collection of early instrumental temperature records, yet systematic evaluations of automated homogenisation methods for such data remain limited.

Here, we evaluate three widely used homogenisation methods, CLIMATOL, PHA, and BART, applied to the HCLIM dataset. The assessment considers data retention following pre-processing, breakpoint detection in the homogenised series, and agreement with the Twentieth Century Reanalysis version 3 (20CRv3), using French and Southeast Asian station networks as representative case studies.

The pre-processed datasets differ only modestly in overall structure, although BART retains fewer records (80 %) than CLIMATOL (96 %) and PHA (99 %) because of its stricter requirement for a minimum record length of 15 consecutive years. Consequently, BART preserves longer average record lengths and detects substantially more breakpoints, indicating greater sensitivity to potential inhomogeneities. Breakpoint occurrence varies considerably across regions and time periods, reflecting differences in station density and observational coverage. Comparisons with 20CRv3 show generally high agreement for all methods, with stronger consistency in the dense French network than in the sparser Southeast Asian network.

Despite differences in data retention, breakpoint detection, and adjustment strategy, the homogenised datasets show strong agreement with one another and preserve the principal climatic signal. Notably, differences among the homogenisation methods are consistently smaller than the differences between the original and homogenised datasets, indicating that the primary challenge is not selecting a single optimal algorithm but reconstructing temporally coherent climate information from incomplete historical observations.

The homogenised datasets produced in this study constitute HOM-HCLIM, a new global database of homogenised early instrumental temperature records that provides an important resource for climate reconstruction, reanalysis, and long-term climate variability studies.

Keywords: *Early instrumental data, automated homogenisation, homogenisation methods*

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

1. Introduction

Accurate assessment of climate variability and long-term climate change relies on the analysis of long-term early instrumental observations (Hersbach et al., 2020; Wilkinson et al., 2016; Brunet and Jones, 2011). Such datasets are essential for understanding historical climate variability, improving climate reconstructions, and informing future climate projections (Gao et al., 2023; Brönnimann et al., 2019; Slivinski et al., 2019; Solomon et al., 2007). Early instrumental observations are commonly defined as records predating the widespread establishment of national meteorological services, generally before 1850 (Brönnimann et al., 2019), and their availability remains limited. In this context, the HCLIM dataset (Lundstad et al., 2022) provides a global compilation of early instrumental observations of early instrumental observations and represents an important resource for historical climate research, despite substantial gaps and variations in temporal coverage.

However, the reliability of historical climate records is often compromised by non-climatic inhomogeneities, including changes in observing conditions and methodological inconsistencies (Conrad, 1946; WMO, 2018). Examples include changes in (i) site conditions, such as urbanisation, land-use changes, or station relocations; (ii) instrumentation, including calibration, sensitivity, or exposure; and (iii) observation and data processing practices, such as measurement routines, time-averaging methods, or data management procedures (Vincent et al., 2018; Trewin, 2010; Aguilar et al., 2003). These discrepancies may introduce biases that distort the climate signal and compromise data interpretation (Wang et al., 2015; Jones, 2016).

Homogenisation is a well-established approach used to identify and correct inhomogeneities in historical climate data (Ribeiro et al., 2016; Cao and Yan, 2012; Costa and Soares, 2009; WMO, 2011; O'Neill et al., 2022). It generally consists of two main phases: first, the identification of breakpoints, defined as points in time where inhomogeneities occur, and second, the adjustment of the affected series (Aguilar et al., 2003). For early land surface temperature records, homogenisation typically relies on relative statistical methods, in which observations from a given meteorological station are compared with those from neighbouring stations (Gubler et al., 2017). If neighbouring stations are sufficiently close and climatically similar, they are expected to reflect similar climate variability and trends (Menne and Williams, 2009). Ideally, neighbouring records exhibit high correlation, similar variability, and differ only because of scaling factors and random sampling error (Ribeiro et al., 2016). However, homogenisation can be challenging for early instrumental data because neighbouring stations are often sparse or absent (Aguilar et al., 2003). In addition, residual homogenisation errors and uncertainties in climatological normals can introduce systematic biases at individual stations (Osborn et al., 2021; Domonkos, 2023). Identifying and mitigating such biases is therefore essential for robust climate analysis (Changnon and Kunkel, 2006).

A wide range of climatological homogenisation methods is available, and the choice depends on data availability, computational requirements, and the characteristics of the dataset being analysed (WMO, 2008). Previously, the standard normal homogeneity test (SNHT; Alexandersson, 1986) was used to detect breakpoints in the HCLIM dataset (Lundstad et al., 2022, 2023). Originally developed for precipitation series, SNHT identifies potential inhomogeneities through comparison of accumulated anomalies. Although SNHT has been widely applied in climate homogenisation and performs well under many conditions (Venema et al., 2012; Shen et al., 2018), it may produce false break detections, particularly near the beginning and end of time series because of edge effects (Toreti et al., 2011).

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

Alternatively, methods such as CLIMATOL (Guijarro, 2006, 2011, 2018), PHA (Menne and Williams, 2009), and HOMER (HOMogenisation softwarE in R; Mestre et al., 2013), here represented by BART (Joelsson et al., 2022), provide automated approaches that minimise manual intervention during homogenisation. CLIMATOL, developed at the Spanish State Meteorological Agency (AEMET), is a relative homogenisation method designed for monthly climatological series. It is based on SNHT, uses neighbouring stations to construct reference series, and applies iterative breakpoint detection and adjustment procedures. The method is implemented in the R programming language (Guijarro, 2011, 2018). PHA is an automated approach for homogenising climate time series based on pairwise comparisons between neighbouring stations (Menne and Williams, 2009). Since 2011, the Global Historical Climatology Network monthly dataset has been updated using PHA to produce a homogenised dataset (O'Neill et al., 2022). Although PHA has primarily been applied to GHCN, this study extends its application to early instrumental data and assesses its performance for the HCLIM dataset. BART is the digital successor of HOMER; an automated homogenisation tool developed within a European collaborative framework. BART integrates the fast-processing capabilities of HOMER's automatic mode with the flexibility and expert-informed adjustments of its interactive mode (Joelsson, 2021). While automated homogenisation methods offer clear advantages for the reanalysis of early instrumental data, their performance on large and heterogeneous datasets such as HCLIM has not yet been systematically evaluated.

Against this background, this study evaluates the performance of three automated homogenisation methods (CLIMATOL, PHA, and BART) when applied to the global HCLIM dataset. The breakpoint characteristics and homogenised temperature series produced by each method are evaluated through comparisons with the 20CRv3 reanalysis (Slivinski et al., 2019), which provides a global atmospheric reconstruction spanning 1836–2015. The resulting homogenised datasets are then combined to develop HOM-HCLIM, a new global database of homogenised early instrumental temperature records. The objectives are (i) to assess the performance of the three homogenisation methods and (ii) to evaluate the suitability of HOM-HCLIM for climate variability and climate reconstruction studies.

Because early instrumental datasets often contain gaps, incomplete records, and varying temporal coverage, homogenisation should not only be viewed as a correction procedure, but also as a means of reconstructing temporally coherent climate information from incomplete observations. Evaluating how different homogenisation methods perform this task is therefore particularly relevant for historical climate research.

The structure of this paper is as follows: Section 2 describes the data and methods, Section 3 presents the results, Section 4 discusses the findings, and Section 5 concludes.

2. Data and Methods

2.1 Observation-based datasets and homogenisation methods

2.1.1 HCLIM

HCLIM is a global multivariable monthly climate database containing early instrumental observations of several meteorological variables (Lundstad et al., 2023). The database contains series compiled from existing databases that start before 1890 (though continuing to the present), with many of the series subsequently extended using later observations up to 2021 where available (Lundstad et al., 2022, 2023). The inclusion of more recent observations is important for homogenisation, as longer records generally improve the reliability of breakpoint detection and adjustment procedures (Coscarelli et al., 2021; Guijarro, 2018). In the present study, only

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

temperature records are considered. The HCLIM database contains 3,632 temperature series from 125 countries and covers the period 1658–2021.

2.1.2 Twentieth Century Reanalysis (20CRv3)

Near-surface air temperature from the NOAA-CIRES-DOE 20CRv3 reanalysis (Slivinski et al., 2019) is used as an external reference dataset for evaluating the consistency of the homogenised HCLIM records. 20CRv3 provides a global reconstruction of atmospheric conditions spanning the period 1836–2015 and is one of the longest available reanalysis products (Slivinski et al., 2019). Although it relies on relatively sparse observational input data during its early period, it provides a physically consistent representation of historical climate variability and atmospheric conditions over its full temporal extent (Slivinski et al., 2021).
To account for uncertainty, 20CRv3 is generated using an Ensemble Kalman Filter data assimilation system and provides an 80-member ensemble (Slivinski et al., 2021). In this study, monthly mean temperatures from the ensemble mean are used. For each HCLIM station, the nearest 20CRv3 grid cell was selected. Stations located at high elevations or near coastlines may exhibit substantial temperature differences relative to the corresponding grid-cell values owing to the relatively coarse spatial resolution of 20CRv3 ($1° \times 1°$, approximately 110 km). However, such differences are expected to remain relatively constant over time and therefore have only a limited influence on the temporal correlations calculated between 20CRv3 and HOM-HCLIM.

2.1.3 Homogenisation methods

Homogenisation was performed on the HCLIM dataset using three well-established automated methods: CLIMATOL, PHA, and BART. These methods are fully automated, apply relative homogenisation principles based on comparisons with neighbouring stations, and generate statistically derived adjustments to the original series (Peterson et al., 1998). Additional information on the characteristics and implementation of each method is provided in Table S1 of the Supplement.

2.2. Processing chain for homogenisation

The processing chain used for homogenising monthly mean temperature records from the HCLIM dataset is illustrated in Figure 1. In this study, the pre-processed datasets represent the HCLIM records after method-specific screening, exclusions, network construction, and formatting prior to homogenisation. These non-homogenised datasets are referred to as nHom_CLIMATOL, nHom_BART, and nHom_PHA.
The workflow consisted of two main stages. First, data pre-processing was conducted through three sub-steps: (i) record screening and exclusion (Section 2.2.1), (ii) network construction (Section 2.2.2), and (iii) data formatting (Section 2.2.3). Second, breakpoint detection and homogenisation were performed. Breakpoint characteristics in the original HCLIM dataset were evaluated using the Standard Normal Homogeneity Test (SNHT), whereas breakpoint statistics from the homogenised datasets were obtained from the three automated methods (Section 2.2.4). Finally, the resulting homogenised datasets were evaluated against both HCLIM and 20CRv3, including an assessment based on the root mean square deviation (RMSD) (Section 2.2.5).

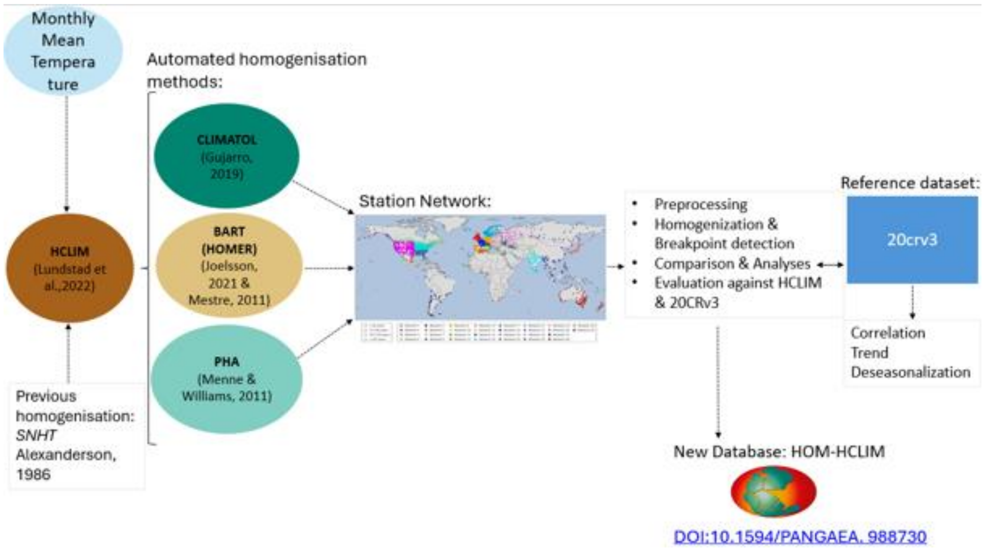

Figure 1. Workflow for the development of the HOM-HCLIM dataset. HCLIM monthly mean temperature data were pre-processed into method-specific input datasets (nHom_CLIMATOL, nHom_BART, and nHom_PHA), homogenized using CLIMATOL, BART, and PHA, and subsequently evaluated against both HCLIM and 20CRv3 before the final HOM-HCLIM dataset was produced.

2.2.1 Exclusions of records

HCLIM records were screened according to the requirements of the three homogenisation methods. Exclusions were based on method-specific criteria, including proximity thresholds, minimum record length, network correlations, and gap tolerance. Together with the subsequent network construction and formatting procedures (Supplementary Material S2–S5), these steps produced the pre-processed, non-homogenised datasets referred to as nHom_CLIMATOL, nHom_BART, and nHom_PHA.

Table 2 summarises the number of HCLIM records removed during pre-processing and the corresponding reasons for exclusion, while Table 3 provides examples of excluded stations (see also Table S3 in the Supplementary Material for the number of records retained at each stage of the homogenisation process). PHA retained nearly all records, excluding only two. In contrast, CLIMATOL and BART applied stricter selection criteria, particularly with respect to correlations with neighbouring stations, spatial distance, and outlier detection. BART additionally required a minimum record length of 15 years. Furthermore, the year 2021 was excluded from all analyses because incomplete years cannot be homogenised reliably.

Table 1. Examples of stations excluded during pre-processing for homogenisation.

Station ID	Station name	Country	M.a.s.l.	Period	Responsive homo-genisation method	Exclusion reason	Excluded from Network
BE-67396	Ben Nevis	UK	1343	1884 - 1903	CLIMATOL	Outlier in temperature (ta)	3
66062	Dawes_Point	Australia	14	1788 - 1853	CLIMATOL & BART	No station close either in time or space	17

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

2790	Kusel-Funfbirken	Germany	296	1881 - 1981	BART	Too short, too many gaps	2
23472	Turukhansk	Russia	38	1880 - 2021	CLIMATOL	Outlier in ta	16
GHCN_ MGXLT954291	Ulaanbaatar	Mongolia	1306	1869 - 2015	CLIMATOL	No neighbour	15

Table 2. Summary of records excluded during pre-processing and the corresponding reasons for exclusion.

Reasons for exclusion	Numbers		
	CLIMATOL	BART	PHA
Total records removed	142	740	2
- Correlation coefficient too low	455		
- Too long distance (defined by the program)	77		
- Record too short		438	
- No neighbours (in distance)		239	
- Outliers identified by the program	4*/182**/ 287***		
- Big gaps	1		
- Other reasons		63	2

* Stations removed ** Station affected *** Observations affected

2.2.2 Network construction

This study initially considered 18 networks, which formed the primary analytical framework, hereafter referred to as Network System A. These networks were defined based on station distribution, geographic proximity, and political boundaries associated with national meteorological services (Figure 2). Network System A served as the common framework for all analyses and comparisons and was applied directly in the PHA method.

For CLIMATOL and BART, an auxiliary network configuration, referred to as Network System B, was introduced to accommodate methodological constraints. Several networks in System A exceeded approximately 1000 km in spatial extent or contained relatively sparse station coverage, which could affect the performance of these methods. Rather than modifying the primary network structure or excluding stations, Network System B was developed as an intermediate processing framework to improve network homogeneity while retaining the full station coverage of System A.

To achieve this, a rule-based network-construction procedure based on Aguilar et al. (2003) was developed to derive smaller and more homogeneous sub-networks from System A. Stations were grouped according to latitude, longitude, and station identifier naming conventions (e.g., GHCN_FR or METFR for France). This approach preserved the overall structure and station composition of System A while improving spatial coherence and consistency with national climatological practices. For example, Köppen-climate-based criteria were applied in Norway following Nordli and Tveito (2008).

Network System B therefore consisted of 30 sub-networks, primarily across Europe, Russia, and Asia, derived from the original 18 networks (Supplementary Material S3 and S4). Following homogenisation, the resulting records were recombined into Network System A, ensuring that all methods could be evaluated and compared within the same primary analytical framework.

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

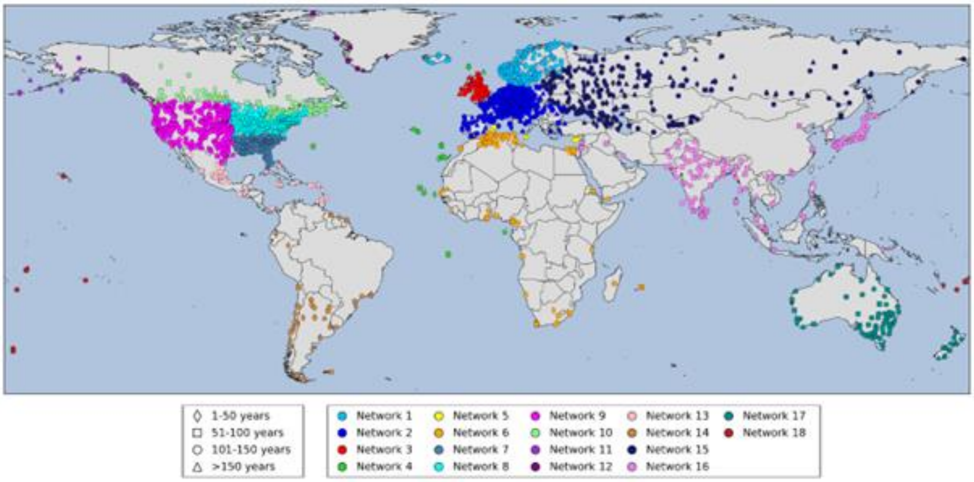

Figure 2: Global distribution of the 18 meteorological networks (Network System A) used to homogenise early temperature records from the HCLIM dataset. Colours indicate individual networks, while symbols represent record length categories. For CLIMATOL and BART, an auxiliary configuration (Network System B) was applied, consisting of 30 sub-networks derived from these 18 networks. Further details are provided in Supplementary Material (Appendix S4).

2.2.3 Data format conversion

HCLIM data were provided in Station Exchange Format (SEF) and reformatted to meet the input requirements of each homogenisation method. CLIMATOL requires CSV input files, with all series combined into a single spreadsheet and accompanied by an .esd metadata file. BART uses ASCII-formatted files containing metadata headers and monthly values arranged row-wise by year, with missing values coded as −999.9. PHA requires separate data and metadata files and was implemented using the original Python and Fortran routines.

Station metadata, including geographic coordinates, elevation, exposure, instrumentation, and observing practices, were harmonised following Lundstad and Tveito (2016). Additional details on data formatting, metadata harmonisation, and method-specific input structures are provided in Table S5 of the Supplementary Material.

2.2.4 Homogenisation procedure

The HCLIM dataset was previously quality controlled and aggregated to monthly means using an 80 % completeness criterion (Lundstad et al., 2023; WMO, 2017). Despite this quality-control procedure, many records remain fragmented or contain periods of missing observations, making homogenisation necessary not only for the adjustment of inhomogeneities but also for the reconstruction of temporally coherent climate series.

Following the pre-processing steps described above, the resulting datasets were homogenised using the CLIMATOL, BART, and PHA methods. Additional standardisation of units and temporal structure was applied where required to meet the input specifications of each method. Each homogenisation method was implemented using its respective breakpoint detection and adjustment procedures.

For comparison, breakpoint characteristics in the original HCLIM dataset were also evaluated using the

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

frequency-based Standard Normal Homogeneity Test (SNHT; Alexandersson, 1986; Lundstad et al., 2023). The resulting breakpoint statistics were subsequently compared with those obtained from the three automated homogenisation methods.

2.2.5 Comparison of the homogenised datasets

The homogenised datasets were evaluated using a three-component framework consisting of (i) comparisons with the original HCLIM dataset, (ii) comparisons with the Twentieth Century Reanalysis (20CRv3), and (iii) an assessment based on the root mean square deviation (RMSD). The France and South-East Asia sub-networks from Network System B were selected for the regional evaluation because they represent contrasting network densities and observational coverage. Their relatively small and spatially coherent configuration provided more favourable conditions for relative homogenisation by increasing station proximity and network homogeneity. The regional analyses are presented in Sections 3.3.1 and 3.3.2, while Paris and Jakarta were selected as representative stations for the station-level comparisons.

3. Results

This section first describes the characteristics of the pre-processed temperature datasets. It then examines the breakpoint characteristics identified by the different homogenisation methods. Finally, the resulting homogenised datasets are evaluated using the three-component framework described in Section 2.2.5.

3.1. Characterization of the pre-processed datasets

Before analysing the breakpoint characteristics and homogenisation results, it is necessary to examine how the pre-processing stage affected the HCLIM dataset. The pre-processing procedures applied by CLIMATOL, BART, and PHA influence record retention, temporal coverage, and data completeness, and may therefore affect the subsequent homogenisation results. This section first examines differences in record length and data availability (Section 3.1.1), followed by an assessment of temporal completeness and gap filling (Section 3.1.2). Finally, the distributions of annual mean temperatures and the preservation of extreme values are evaluated (Section 3.1.3).

3.1.1 Record length and data availability

Figure 3 shows the distribution of the length and number of temperature records for the HCLIM dataset and the pre-processed datasets (nHom_CLIMATOL, nHom_BART, and nHom_PHA). Across all datasets, a large proportion (~71%) of records are long (≥ 100 years). In particular, approximately 1150 records have a length of around 150 years, while about 600 and 400 records span 125 and 175 years, respectively (Fig. 3a).

In contrast, records of intermediate length (50–100 years) and very long records (≥ 200 years) are relatively rare (≤ 200 records). At the same time, a substantial number of short records (≤ 25 years) is present across most datasets. HCLIM and nHom_PHA each include nearly 800 short records, while nHom_CLIMATOL contains around 700. By contrast, nHom_BART retains fewer than 200 short records, reflecting its stricter selection criteria. The resulting U-shaped distribution of record lengths (Fig. 3a) reflects both the abundance of short series and the dominance of century-long records, combined with the exclusion of stations starting after 1890 (Lundstad, 2023).

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

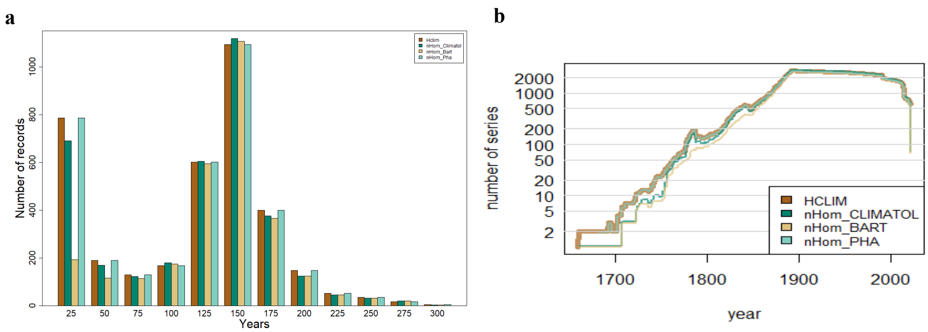

Figure 3. (a) Length and number of temperature records in the HCLIM dataset and the pre-processed datasets derived using the CLIMATOL, BART, and PHA methods (i.e., non-homogenised datasets: nHom_CLIMATOL, nHom_BART, and nHom_PHA). Records are shown up to 2020. (b) Temporal evolution of the number of temperature records for the same datasets (HCLIM, nHom_CLIMATOL, nHom_BART, and nHom_PHA). Note the logarithmic scale.

The number and length of records in the pre-processed datasets vary depending on the homogenisation method applied (Fig. 3a). The nHom_CLIMATOL dataset contains more records in the 100–150-year range than HCLIM, but fewer records between 25 and 75 years. This difference reflects both the large number of short series in HCLIM and the tendency of CLIMATOL to remove duplicated short-term records that are retained in HCLIM.

In contrast, nHom_BART contains substantially fewer short and medium-length records, particularly in the 25–50-year range, reflecting its stricter pre-processing criteria, including requirements for minimum record length, sufficient record overlap, and suitable neighbouring stations (Joelsson, 2021). Consequently, BART retains a relatively high proportion of long records, resulting in the longest average record length (122 years), compared to 102 years for HCLIM and 103–104 years for nHom_CLIMATOL and nHom_PHA. The nHom_PHA dataset closely follows the HCLIM distribution across all record lengths, reflecting its emphasis on the availability of neighbouring stations rather than record length.

Figure 3b illustrates the temporal evolution of record availability. All datasets show a marked increase in the number of records from the 18th century onwards, with a peak around 1890. A distinct increase in the late eighteenth century, associated with the expansion of meteorological observations during the Napoleonic period ("Napoleon bump"), is evident in HCLIM, nHom_CLIMATOL, and nHom_PHA, but largely absent in nHom_BART due to its exclusion of short records.

Throughout the period, nHom_BART consistently contains fewer records than the other datasets, whereas nHom_CLIMATOL and nHom_PHA closely follow the structure of HCLIM. In addition, the pre-processed datasets tend to start later than HCLIM, reflecting the requirement for sufficient neighbouring stations during the pre-processing and network-construction stages. Consequently, early periods with sparse data coverage are excluded, even though they are present in the original HCLIM dataset.

3.1.2 Temporal completeness and gap filling

The homogenisation procedures also affected temporal completeness and record continuity. CLIMATOL generally produced the most complete records and the longest continuous time series, often filling gaps present in the original HCLIM data. In contrast, PHA tended to produce shorter series because of its stronger dependence

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

on neighbouring stations and overlap requirements. While HCLIM contains numerous gaps, all three homogenisation methods were generally able to reconstruct missing values and improve temporal continuity. An exception was observed for some stations in South-East Asia, where PHA occasionally retained gaps due to limited reference station availability.

Bourges, France (GHCN_FR000007255), provides an illustrative example (Fig. 4a). Although the station record spans 1851–2013, the original HCLIM series contains a large gap between 1897 and 1945. The homogenisation methods used neighbouring stations to reconstruct this missing period, thereby producing a substantially more continuous record. A similar pattern is observed for Hanoi, Vietnam (GHCN_VMXLT492315), where a gap between 1921 and 1971 was largely reconstructed despite the station record spanning 1886–2013.

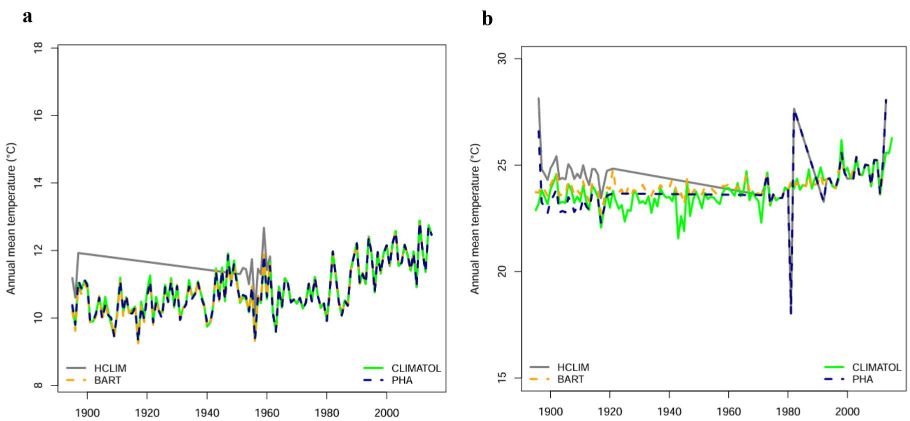

Figure 4. Examples of gap reconstruction and improved temporal continuity following homogenisation. (a) Bourges, France (1851–2013), where a large gap between 1897 and 1945 in the original HCLIM record was reconstructed by the homogenisation methods. (b) Hanoi, Vietnam (1886–2013), where observations missing between 1921 and 1971 were reconstructed. In both cases, CLIMATOL, BART, and PHA substantially improved record completeness relative to the original HCLIM series.

3.1.3 Temperature distributions and extreme values

Figure 5a illustrates the distribution of annual mean temperatures in the HCLIM and pre-processed datasets. Although all datasets originate from the same station base, the original HCLIM dataset exhibits a broader distribution with more pronounced tails. The disappearance of the low-temperature peak in the pre-processed datasets is primarily caused by the removal of records during the initial filtering step. As a result, the pre-processed datasets show more compact and internally consistent distributions, reflecting method-specific data selection and screening procedures. Following homogenisation (Fig. 5b), the distributions converge and show a strong overlap across all methods, while the overall shape is largely preserved. This indicates that the homogenisation procedures reduce non-climatic variability without substantially altering the underlying temperature distribution. Remaining differences are minor and mainly confined to the distribution tails. The strong agreement across methods suggests that, in dense station networks such as France, automated homogenisation preserves the overall temperature distribution despite differences in record retention and pre-processing.

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

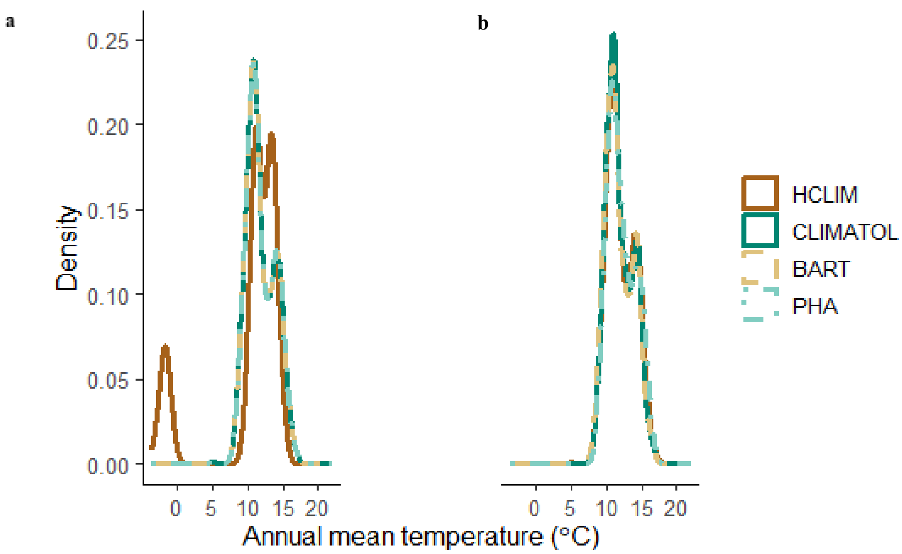

Figure 5. Distribution of annual mean temperatures across datasets before and after homogenisation. (a) Kernel density distributions for the original HCLIM dataset and the pre-processed (non-homogenised) datasets (nHom_CLIMATOL, nHom_BART, nHom_PHA). Differences reflect variations in data availability and selection. (b) Kernel density distributions after homogenisation. The resulting datasets show a high degree of overlap across methods, with similar central tendencies and spread.

While Figure 5 illustrates the overall temperature distributions before and after homogenisation, Table 3 summarises the corresponding extreme values for HCLIM and the pre-processed datasets prior to homogenisation. Overall, a high degree of consistency is observed between HCLIM and the three pre-processed datasets. The most extreme temperature values and their corresponding stations remain nearly identical, indicating that record screening and pre-processing introduce only minor changes to the most extreme temperature values. For mean temperature, the same locations (Massawa and Ustjansk) consistently exhibit the highest and lowest values across all datasets, although small differences in magnitude are present (e.g., −19.4 °C versus −18.3 °C for the lowest mean temperature). Likewise, Verkhoyansk consistently records the lowest absolute minimum temperature, whereas Province Wellesley records the highest absolute minimum temperature across all datasets. Some variation is observed in maximum mean temperatures, where Cape Flora (Franz Josef Land) is replaced by Racine as the location of the lowest maximum mean temperature in certain methods (e.g., CLIMATOL and BART).

Overall, the analyses indicate that the three pre-processing procedures differ substantially in terms of record retention and temporal coverage, particularly for short records. However, the principal characteristics of the temperature distributions and the most extreme temperature values are largely preserved across all datasets. This suggests that the subsequent homogenisation analyses are based on datasets that remain broadly representative of the original HCLIM record despite method-specific screening and selection procedures.

Table 3. Highest and lowest mean, minimum, and maximum temperatures in HCLIM and the pre-processed datasets (nHom_CLIMATOL, nHom_BART, and nHom_PHA).

Dataset/	Mean (highest, lowest °C)	Minimum (lowest, highest °C)	Maximum (highest, lowest °C)

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

Methods			
HCLIM	Massawa (29.9 °C) – Ustjansk (-19.4 °C)	Verhojansk (-55.4 °C) – Province Wellesley (26.1 °C)	Bikaner (40.7 °C) – Cape_Flora_Franz_Josef_Land (1.9 °C)
nHom_CLIMATOL	Massawa (29.9 °C) – Ustjansk (-18.3 °C)	Verhojansk (-55.4 °C) – Province Wellesley (26.1 °C)	Bikaner (40.7 °C) – Racine (2.8 °C)
nHom_PHA	Massawa (29.9 °C) – Ustjansk (-19.4 °C)	Verhojansk (-55.4 °C) – Province Wellesley (26.1 °C)	Bikaner (40.7 °C) – Cape_Flora_Franz_Josef_Land (1.9 °C)
nHom_BART	Massawa (29.9 °C) – Ustjansk (-18.3 °C)	Verhojansk (-55.4 °C) – Province Wellesley (26.1 °C)	Bikaner (40.7 °C) – Racine (2.8 °C)

In the next section, we examine the homogenisation process and the resulting breakpoint characteristics in these datasets.

3.2. Characterization of the breakpoints in the homogenised datasets

This section compares the breakpoint characteristics identified by the three homogenisation methods at both global and regional scales. For reference, breakpoint detection based on the SNHT method applied to the original HCLIM dataset is also included.

3.2.1. Comparison of Breakpoint Characteristics

The initial step in the homogenisation process involved identifying breakpoints in both time and space. The three homogenisation methods applied in this study produced substantially different numbers of breakpoints in the homogenised datasets (hom_CLIMATOL, hom_BART, and hom_PHA). For comparison, breakpoint detection based on SNHT was also applied to the original HCLIM dataset. Among these, hom_BART yielded the highest number of breakpoints (14,851), whereas hom_PHA identified the fewest (2,676) across the global network. The hom_CLIMATOL dataset contained 3,186 breakpoints, while applying SNHT to the original HCLIM dataset resulted in 11,888 breakpoints. Consequently, the sensitivity assessment presented here should be interpreted in the context of these default configurations. It should be noted that both CLIMATOL and BART include adjustable parameters. Consequently, the breakpoint frequencies reported here reflect the parameter settings adopted in this study, and the sensitivity assessment should be interpreted within this configuration.

Table 4 summarises the earliest breakpoint detected in each dataset together with the corresponding station and year. The earliest breakpoint (1662) was identified in the HCLIM dataset using the SNHT method at the Paris station. Among the homogenisation methods, the earliest breakpoint was detected by BART at Delft_Rijnsburg in 1727, followed by CLIMATOL at De Bilt in 1760. In contrast, the earliest breakpoint identified by PHA occurred considerably later, at De Bilt in 1865 for annual data. For monthly data, PHA first detected breakpoints in 1895, including stations such as De Bilt and Hamburg-Deutsche Seewarte (Table 4). Overall, these findings highlight the strong European dominance in early breakpoint detection, consistent with the early development of systematic meteorological observations in Europe (Dobrovolný et al., 2010).

Table 4. Earliest detected breakpoint and corresponding meteorological station in the HCLIM dataset (based on SNHT) and in the homogenised datasets produced by CLIMATOL, PHA, and BART (hom_CLIMATOL, hom_PHA, and hom_BART).

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

Dataset/Methods	Station	Station ID	Year
HCLIM	Paris	EKF400v2t	1662
hom_BART	Delft_Rijnsburg	KNMI-49_Delft_Rijnsburg	1727
hom_CLIMATOL	DeBilt	KNMI_UBERN00575_DeBilt	1760
hom_PHA	DeBilt	KNMI_UBERN00575_DeBilt	1865 (annually)
hom_PHA	Hamburg_Deutsche-Seewarte (+ 27 other stations)	15526 (and many more)	1895 (monthly)

3.2.2. Regional and Temporal Trends in Breakpoint Detection Rates Across Homogenisation Methods

Figure 6 presents a heatmap of breakpoint detection rates (i.e., annual breakpoints per 1000 station-years) across meteorological networks over 40-year periods from 1701 to 2020 for the three homogenisation methods (BART, CLIMATOL, and PHA). This normalisation allows direct comparison across networks with different data availability.

Overall, BART exhibits consistently higher breakpoint detection rates than CLIMATOL and PHA. In particular, during the period 1861–1900, BART shows elevated detection rates across networks in Africa, Australia, the Mediterranean, and the northeastern United States. This pattern is consistent with the fully automated implementation of BART in this study, which differs from its original semi-manual design (e.g. within HOMER; Joelsson et al., 2022). The higher detection rates likely reflect the automated and iterative application of predefined breakpoint detection criteria across the dataset. The absence of BART results for the Alaska and Pacific networks (Fig. 6) reflects cases where the method could not be applied successfully in these regions.

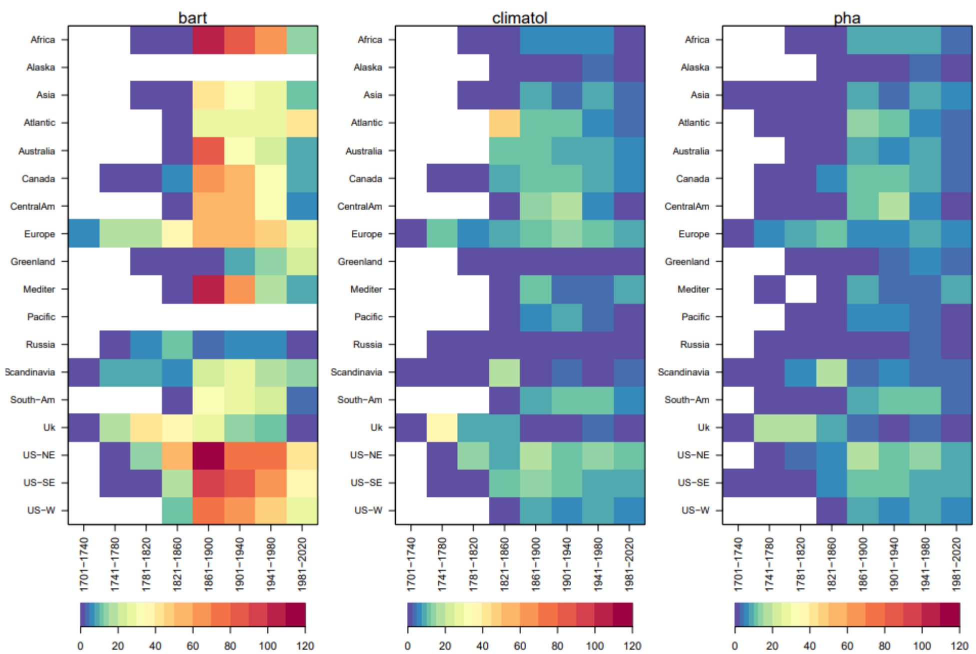

Figure 6. Heatmap of breakpoint detection rates (i.e., annual breakpoint frequency per 1000 station-years) across meteorological networks for 40-year periods between 1701 and 2020, shown for the three homogenisation methods used in this study (BART, CLIMATOL, and PHA; left, middle, and right panels, respectively). Rows represent station networks, and columns represent 40-year intervals. The colour scale is intentionally non-linear to enhance the

visualization of differences across orders of magnitude. Increasing red intensity indicates higher breakpoint detection rates. Corresponding values are provided in Table S3 (Supplementary Material).

Two main regional and temporal patterns emerge. First, prior to approximately 1860, all three methods exhibit relatively low breakpoint detection rates, reflecting the sparse observational coverage during the early instrumental period. During this early period (1701–1740), Europe, Scandinavia, and the UK show comparable breakpoint detection rates. PHA additionally identifies similar detection levels in parts of Asia and the southeastern United States.

Second, after 1900, breakpoint detection rates increase markedly in the United States across all methods, most prominently for BART. This rise coincides with the expansion of meteorological observations associated with agricultural development, transportation, and settlement in newly incorporated territories (Fleming, 1990). In contrast, detection rates decline in the most recent period (1981–2020) across all methods, particularly for BART. This decrease likely reflects increased network stability, improved measurement practices, and reduced station relocations associated with the automation of observations, although reduced algorithm sensitivity may also contribute (WMO-No. 8, 2008; see Supplement S5).

Overall, breakpoint detection increased with the expansion of observational networks but differed substantially among the three homogenisation methods, particularly in periods and regions with contrasting data availability.

3.3. Evaluation of the homogenised datasets

This section presents a comparison of the three homogenised datasets (hom_CLIMATOL, hom_BART, and hom_PHA) against the original HCLIM dataset (Section 3.3.1) and the 20CRv3 reanalysis (Section 3.3.2). In addition, RMSD-based comparisons are presented (Section 3.3.3). Together, these analyses evaluate how different homogenisation methods influence the resulting temperature series and their consistency with external reference datasets. Because HCLIM contains gaps and varying temporal coverage, comparisons with the homogenised datasets should be interpreted as differences between the original and reconstructed records rather than as deviations from a perfect reference series.

3.3.1. Temperature variation among the homogenised datasets and the *HCLIM* dataset

France and South-East Asia represent contrasting observational networks, with France characterised by a long and dense station network and South-East Asia by a sparser and more spatially heterogeneous observational coverage. To assess whether the homogenisation methods introduced systematic differences in annual mean temperature, one-way ANOVA and Tukey HSD tests were applied to the original HCLIM dataset and the three homogenised datasets (hom_CLIMATOL, hom_BART, and hom_PHA) for the common comparison period 1895–2015.

For France, the ANOVA revealed statistically significant differences among the datasets ($F = 8.64$, $p = 1.02 \times 10^{-5}$). Tukey HSD showed that these differences were primarily attributable to contrasts between the original HCLIM dataset and the homogenised datasets (Fig. 7a). All three homogenised datasets exhibited significantly lower mean temperatures than HCLIM, with mean differences ranging from approximately 0.24 to 0.29 °C. In contrast, no statistically significant differences were detected among hom_CLIMATOL, hom_BART, and hom_PHA, indicating strong agreement between the homogenisation methods. No statistically significant differences were detected among hom_CLIMATOL, hom_BART, and hom_PHA.

For South-East Asia, the ANOVA also identified significant differences among the datasets (F = 11.3, $p = 2.16 \times 10^{-7}$). However, the Tukey HSD results indicate that the significance was primarily driven by the CLIMATOL dataset (Fig. 7b). Mean annual temperatures in hom_CLIMATOL were significantly higher than those in HCLIM (+0.323 °C, $p = 0.002$), hom_BART (−0.411 °C, $p < 0.001$), and hom_PHA (−0.427 °C, $p < 0.001$). No significant differences were detected between HCLIM, hom_BART, and hom_PHA. Unlike France, where HCLIM differed systematically from all homogenised datasets, the South-East Asian results suggest that the observed differences are largely associated with CLIMATOL, while BART and PHA remain statistically consistent with the original HCLIM data.

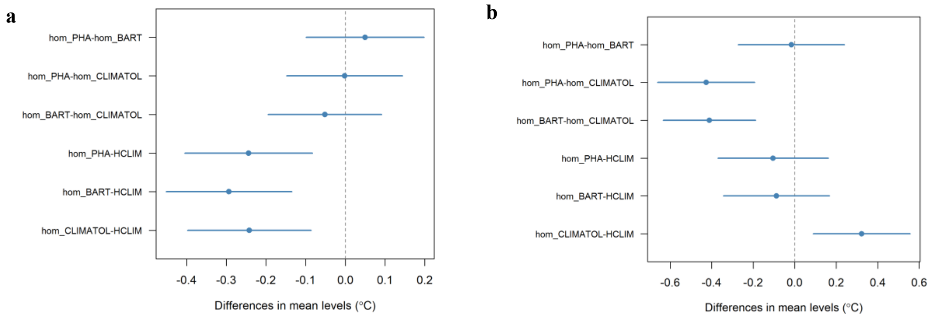

Figure 7. Tukey HSD pairwise comparisons of annual mean temperatures between HCLIM and the homogenised datasets (hom_CLIMATOL, hom_BART, and hom_PHA) during the common period 1895–2015. Points indicate estimated mean differences, and horizontal lines represent 95 % family-wise confidence intervals. The vertical dashed line denotes zero difference. (a) France. Significant differences are observed between HCLIM and all three homogenised datasets, whereas no significant differences occur among the homogenised datasets themselves. (b) South-East Asia. Significant differences are primarily associated with hom_CLIMATOL, while HCLIM, hom_BART, and hom_PHA do not differ significantly from one another.

The Taylor diagrams (Fig. 8; Taylor, 2001) compare the homogenised datasets with the original HCLIM dataset for France and South-East Asia using standard deviation, correlation, and centred RMSD. In both regions, all homogenised datasets exhibit very high correlations with HCLIM ($r \approx 0.98$–0.99), indicating that the homogenisation procedures largely preserve the temporal variability and overall climatic signal of the original records.

For France (Fig. 8a), the three homogenised datasets cluster closely around the HCLIM reference point, exhibiting nearly identical standard deviations (≈ 1.9 °C) and very low centred RMSD values (< 0.3 °C). Only minor differences are apparent, with hom_BART showing a slightly lower standard deviation than the other methods. In South-East Asia (Fig. 8b), correlations remain similarly high, although the datasets display a somewhat larger spread in standard deviation and centred RMSD, reflecting the lower station density and greater spatial heterogeneity of the network. Hom_CLIMATOL most closely reproduces the standard deviation of HCLIM, whereas hom_BART exhibits slightly reduced variability and hom_PHA slightly enhanced variability. Nevertheless, all three methods remain tightly clustered near the HCLIM reference point, demonstrating that the principal characteristics of the climate signal are preserved across methods.

This consistency is particularly evident for CLIMATOL and BART, which generally produce longer and more complete temperature series than the original HCLIM dataset. CLIMATOL typically generates the most continuous records, while BART also reconstructs missing periods but applies stricter selection criteria that limit the inclusion of shorter records. Although such reconstructions require additional methodological assumptions, the strong agreement among the homogenised datasets suggests

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

that the principal climatic characteristics are reproduced consistently despite differences in record completeness and homogenisation strategy.

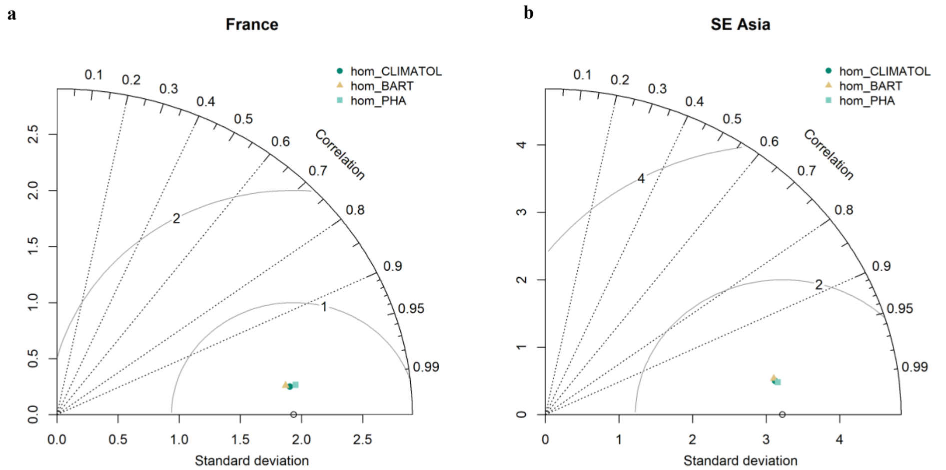

Figure 8. Taylor diagrams comparing the homogenised datasets (hom_CLIMATOL, hom_BART, and hom_PHA) with the original HCLIM dataset for the common period 1895–2015. The diagrams summarise agreement in terms of correlation, standard deviation, and centred root mean square difference (RMSD). The open circle on the x-axis represents the HCLIM reference dataset. Datasets located closer to the reference point indicate stronger agreement. (a) France, representing a dense and well-connected observational network. (b) South-East Asia, representing a sparser and more spatially heterogeneous network. In both regions, correlations are very high ($r \approx 0.98$–0.99), while the homogenised datasets exhibit only minor differences in variability and RMSD relative to HCLIM. The slightly larger spread in South-East Asia reflects greater network sparsity and regional heterogeneity.

3.3.2 Temperature variations among the homogenised datasets and 20CRv3 dataset

Based on STL decomposition, both deseasonalised and trend components were analysed to evaluate agreement between HCLIM, the homogenised datasets, and 20CRv3. Figure 9 provides insights into the effects of homogenisation on temperature records at the Paris and Jakarta station. At the Paris station, the HCLIM dataset provides the longest record, spanning 1658–2018 (360 years), whereas the homogenised datasets show shorter coverage. These differences reflect methodological requirements, particularly the need for sufficient overlap and reliable reference series, which constrain the effective length of homogenised records. Both hom_BART and hom_CLIMATOL begin in 1757 (263 years), while hom_PHA has the shortest record, starting in 1895, reflecting more restrictive data requirements (Menne and Williams, 2009). The 20CRv3 dataset begins in 1836, providing intermediate coverage (Fig. 9a). Despite the shorter record length, homogenisation enhances temporal consistency by reducing short-term variability and revealing clearer long-term trends. In particular, hom_CLIMATOL shows a smoother evolution compared to the original HCLIM dataset, while hom_BART indicates a slight temperature increase towards the end of the record. At the Jakarta station, the relative dataset coverage differs substantially from Paris. Here, 20CRv3 provides the longest record (1836–2012, 176 years), followed by hom_CLIMATOL (1863–2020, 157 years). In contrast, HCLIM covers a shorter period (1866–1970, 104 years), similar to hom_BART, while hom_PHA again shows the shortest coverage (1895–1969, 73 years) (Fig. 9b).

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

Overall, the homogenised, deseasonalised, and smoothed datasets show good agreement at the Paris station, with only minor differences between methods. In contrast, larger discrepancies are observed at the Jakarta station. In particular, a sudden temperature increase of approximately 1°C is evident in the HCLIM dataset around 1953 in the deseasonalised series (Fig. 9b), likely reflecting a non-climatic breakpoint associated with station relocation, instrumentation changes, or changes in observing practices. Such discontinuities are common in early instrumental records and highlight the importance of homogenisation for ensuring temporal consistency (Aguilar et al., 2003; Menne and Williams, 2009; Venema et al., 2012). This feature is consistent with documented inhomogeneities in the Jakarta Observatory record, related to transitions between observational sites and changes in station exposure, including the shift from Batavia Observatory to Kemayoran station (Siswanto et al., 2015). The absence of this discontinuity in the hom_CLIMATOL dataset suggests that CLIMATOL has successfully detected and corrected the inhomogeneity, likely through comparisons with neighbouring reference stations (Guijarro, 2018; Venema et al., 2012).

Despite these differences, all datasets at both stations exhibit a consistent long-term warming trend over recent decades, in agreement with previous studies (Brönnimann, 2022).

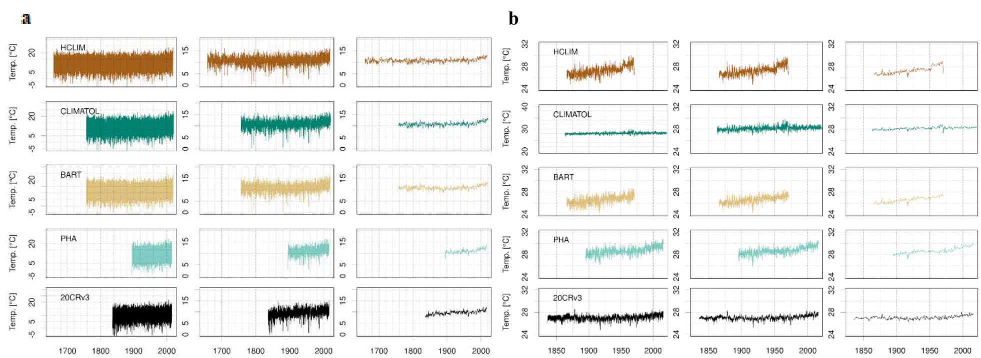

Figure 9. Temperature records (°C) for the Paris (a) and Jakarta (b) stations. The left column shows, from top to bottom, the original HCLIM dataset, the homogenised datasets (hom_CLIMATOL, hom_BART, and hom_PHA), and the 20CRv3 dataset. The middle column shows the corresponding deseasonalised series, and the right column shows the trend component (i.e. the smoothed series after removal of the seasonal cycle).

To further assess dataset quality relative to 20CRv3, correlations were computed based on STL-derived trend and deseasonalised components, as shown in Figure 10, for the homogenised datasets (hom_CLIMATOL, hom_BART, and hom_PHA) and HCLIM across 24 stations in the French and South-East Asian networks. Strong correlations are observed across the French network. For both the trend and deseasonalised time series (Fig. 10a,b), the Paris station shows high agreement with 20CRv3, with correlation values of 0.92, 0.92, 0.90, and 0.89 for the trend component, and 0.96, 0.96, 0.95, and 0.95 for the deseasonalised series (HCLIM, hom_CLIMATOL, hom_BART, and hom_PHA, respectively). Overall, CLIMATOL and HCLIM show the highest agreement with 20CRv3 in the French network, with mean correlations of 0.92 for the trend component and 0.96 for the deseasonalised series.

In contrast, correlations are lower and more variable across the South-East Asian network (Fig. 10c,d), reflecting the sparser and more heterogeneous observational conditions. At the Jakarta station, correlations for the trend component are moderate, with median values ranging from approximately 0.45 to 0.60, and

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

hom_BART showing the highest agreement with 20CRv3. For the deseasonalised series, correlations increase across all datasets, indicating improved agreement once the seasonal cycle is removed. In this case, hom_BART and hom_PHA show the highest correlations, although substantial variability remains across methods. These results indicate that homogenisation improves consistency with reanalysis data, particularly for deseasonalised temperature series, while differences between methods remain more pronounced in data-sparse regions such as South-East Asia.

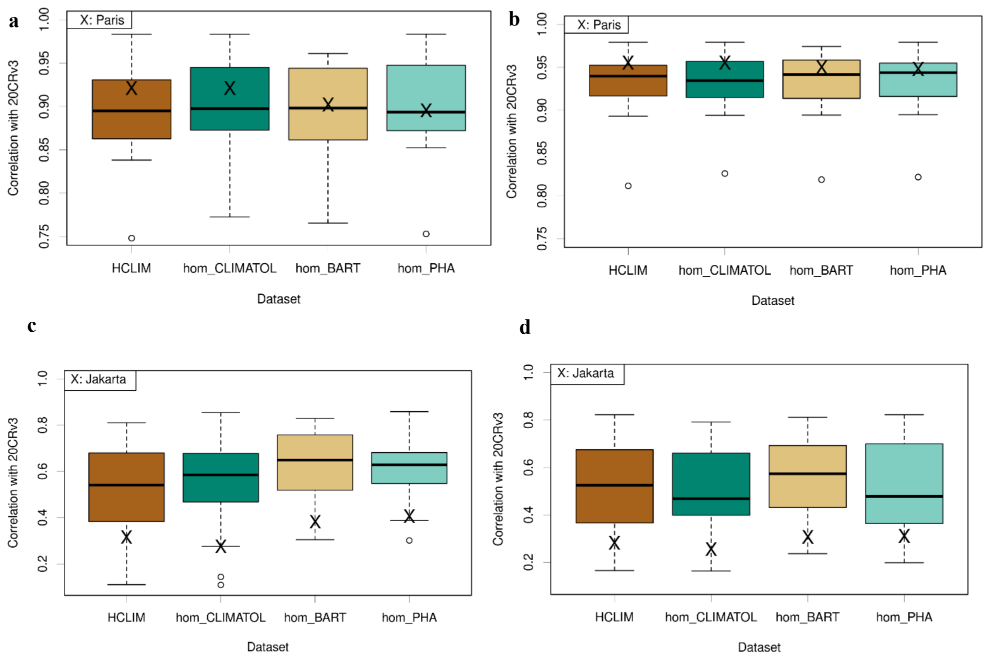

Figure 10. Boxplots of correlations of the trend (a, c) and the deseasonalised (b, d) components of the *HCLIM* and the homogenised datasets using BART, CLIMATOL, and PHA homogenisation methods (*i.e.*, hom_CLIMATOL, hom_BART, and hom_*PHA*) against the 20CRv3 dataset for 20 stations in the French (a, b) as well as the South-East Asian (c, d) networks. The marker (X) in panels a and b represent the Paris station, while the X in c and d represents the Jakarta stations.

3.3.3 Differences between the homogenised datasets and HCLIM

To evaluate the impact of homogenisation, the root mean squared deviation (RMSD) was calculated between the homogenised datasets (hom_CLIMATOL, hom_BART, and hom_PHA) and the original HCLIM dataset at eight representative stations: Paris, Turin, Trondheim, Charleston, Funchal, Rio de Janeiro, Calcutta, and Sydney (Figure 11). These stations were selected because they provide long and continuous temperature records across different climatic regions, including well-documented European stations with dense observational networks. In this study, RMSD quantifies the magnitude of differences between the homogenised datasets and the original HCLIM dataset, providing an integrated measure of how strongly each method modifies the series in response to detected inhomogeneities.

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

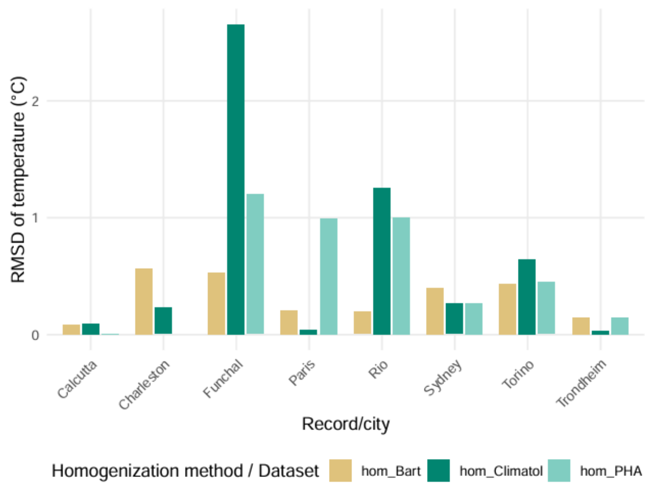

Figure 11. Root mean squared deviation (RMSD) between the homogenised datasets (hom_CLIMATOL, hom_BART, and hom_PHA) and the original HCLIM dataset for eight representative meteorological stations (i.e., Calcutta[1], Charleston[2,] Funchal[3], Paris[4], Rio de Janeiro[5], Sydney[6], Torino[7], and Trondheim[8]). While the hom_BART dataset generally has smaller RMSD values compared to hom_CLIMATOL, the Funchal station has the highest RMSD in the hom_CLIMATOL dataset.

RMSD values varied considerably among both the homogenisation methods and the eight stations (Figure 11). At Funchal, hom_CLIMATOL produced a very high RMSD value (2.65), whereas considerably lower values were obtained for hom_BART (0.53) and hom_PHA (1.20). Similarly, relatively large RMSD values were observed for hom_CLIMATOL at Rio de Janeiro (1.25), while hom_BART produced a much lower value (0.19). In contrast, several stations showed relatively low RMSD values across all homogenisation methods. At Calcutta, the RMSD values were 0.09, 0.08, and 0.002 for hom_CLIMATOL, hom_BART, and hom_PHA, respectively. Comparable values were observed at Sydney (0.26, 0.39, and 0.26) and Trondheim (0.03, 0.14, and 0.14). At Paris, hom_CLIMATOL produced the lowest RMSD value (0.036), whereas higher values were obtained for hom_BART (0.21) and hom_PHA (0.99).

4. Discussion

This study compared the homogenised datasets derived from the early instrumental HCLIM dataset using three automatic homogenisation methods. The pre-processed datasets show minor differences. The main distinction among these datasets is their retention rate: nHom_*BART* preserved about 80% of HCLIM records, while

[1] PALAEO-RA_Global-Datasets_Dove-441_181601-184312_ta_monthly_merged
[2] BERKELEY_EARTH-HadleyCentreCRUpartialdatacollection_BE-67539_182301-201105_ta_monthly_merged
[3] GHCN_v4_qcf_GHCN_PO000008522_186501-202112_ta_monthly_merged
[4] Hclim_EKF400_EKF400v2t_002_165801-201812_ta_monthly_merged
[5] BERKELEY_EARTH-HadleyCentreCRUpartialdatacollection_BE-64963_183201-199108_ta_monthly_merged
[6] GHCN_v4_qcf_GHCN_ASN00066062_185901-202012_ta_monthly_merged
[7] ISTI_00122_Torino_1760_2009_ta_monthly_merged
[8] BERKELEY_EARTH-HadleyCentreCRUpartialdatacollection_BE-19012_176201-201105_ta_monthly_merged

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

nHom_CLIMATOL and nHom_PHA contained 96% and 99%, respectively. CLIMATOL and PHA demonstrated greater tolerance for shorter series, which resulted in denser spatial coverage. This can be an advantage, particularly in data-sparse regions, as it improves the representation of small-scale variability. However, shorter records may reduce the reliability of the resulting adjustments. Despite these differences, all three methods demonstrated strong consistency in retaining the most valuable records, specifically the longest time series and the most extreme temperature values. The pre-processed datasets retained most extreme temperature values from the original HCLIM records, indicating that the initial homogenisation step did not obscure climatic extremes. This is a critical outcome, as pre-processed records often serve as anchors for understanding past climate variability and extremes (Jones et al., 2002; Brázdil et al., 2010; Brönnimann et al., 2015, 2018). An important aspect of the present study is that the homogenised datasets are not being evaluated against a complete and error-free reference record. HCLIM represents a unique collection of early instrumental observations, but many of the underlying series contain gaps, varying temporal coverage, and periods of limited observational continuity. Consequently, differences between HCLIM and the homogenised datasets should not automatically be interpreted as homogenisation artefacts. Rather, they often reflect the ability of the homogenisation algorithms to reconstruct a temporally coherent climate signal from incomplete observations. CLIMATOL generally produced the longest and most continuous temperature series, whereas PHA was more conservative and retained shorter records and BART retained fewest records (≈80 %) because of its stricter selection criteria. The broad agreement among the homogenised datasets despite these differences suggests that the major climatic characteristics are reproduced consistently, even when substantial data reconstruction is required. This interpretation is consistent with the Tukey HSD and Taylor diagram results. These findings suggest that automated homogenisation can substantially enhance the value of fragmented early instrumental observations for climate research. This is particularly important because many seventeenth- to nineteenth-century records remain underutilised in climate reconstructions owing to gaps and non-climatic discontinuities.

BART detected more breakpoints than CLIMATOL and PHA, reflecting its iterative adjustment procedure and greater sensitivity to potential inhomogeneities. This increased the likelihood of identifying subtle artificial shifts but may also increase the risk of false positives and over-adjustment in sparse networks with limited reference series (Lindau and Venema, 2015). In contrast, CLIMATOL and PHA applied more conservative criteria, reducing the risk of excessive corrections but potentially overlooking smaller discontinuities, resulting in possible under-homogenisation (Domonkos et al., 2021; Lindau et al., 2015). Method performance is further influenced by factors such as network density, natural climate variability, and methodological design, including the degree of iteration and automation (Menne and Williams, 2009; Gubler et al., 2017; Venema et al., 2012). Methods using more iterative adjustment procedures tend to be more sensitive, whereas conservative approaches prioritise stability and are less likely to adjust weaker signals. Consequently, the choice of homogenisation method requires careful consideration of the trade-off between over- and under-correction (Costa and Soares, 2009). The spatial distribution of the retained records also reflects the historical development of meteorological observing networks. The predominance of European records in the homogenised datasets is consistent with Europe's long tradition of systematic meteorological measurements and relatively strong institutional support for data collection and archiving. In many regions outside Europe, early instrumental records remain sparse, fragmentary, or poorly documented, limiting their usefulness for homogenisation analyses (Parker et al., 2000; Jones, 2001; Camuffo et al., 2023; Brönnimann et al., 2019).

The case of De Bilt, one of Europe's longest continuous temperature series, illustrates how metadata and statistical methods can converge on the same breakpoint, in this case as early as 1727, thereby increasing confidence in the homogenised record (Brandsma and Können, 2006; Brönnimann et al., 2019). The pronounced increase in breakpoint detection after the 1860s coincides with the rapid expansion of meteorological networks in Europe and North America, including developments such as the U.S. Army Signal Corps and the establishment of the U.S. Weather Bureau (Harper, 2008; Easterling et al., 2016). While these expanding networks improved spatial coverage, they also introduced new sources of inhomogeneity.

Conversely, the decline in breakpoint detections after 1980 likely reflects increasing automation, instrument standardisation, and improved observing practices. Overall, the observed breakpoint patterns highlight the combined influence of historical data availability, network development, and methodological sensitivity. As meteorological observing systems evolve, breakpoints remain an inevitable feature of long-term climate records, making their identification and correction essential for producing reliable climate datasets (Aguilar et al., 2003; Venema et al., 2012; Menne and Williams, 2009; Trewin, 2010).

RMSD analyses indicated that all methods generally preserved the large-scale climate signal, although station-specific differences remained. Higher RMSD values at Funchal and Rio de Janeiro likely reflect complex local conditions and sparse reference networks. BART generally produced the lowest RMSD values, whereas CLIMATOL showed greater regional variability. PHA performed best at selected stations, particularly Charleston and Calcutta. However, no single method consistently outperformed the others across all regions and stations.

Comparisons with 20CRv3 revealed clear regional differences. Correlations were generally higher and more consistent in the dense French network than in the sparser South-East Asian network, reflecting the strong influence of network density and data availability on both homogenisation performance and reanalysis agreement (Menne and Williams, 2009; Gubler et al., 2017). These findings are consistent with previous studies showing that reanalyses perform best in well-observed regions, whereas sparse observational networks are more strongly affected by observational uncertainty and natural variability (Compo et al., 2011; Brönnimann et al., 2018; Rayner et al., 2003).

Although BART generally showed the strongest agreement with 20CRv3, no single homogenisation method proved universally superior. Method performance depended strongly on network density, temporal coverage, and station availability, confirming earlier benchmarking studies (Venema et al., 2012; Gubler et al., 2017; Domonkos et al., 2021). Consistent with the Taylor diagrams, all three methods converge towards a similar large-scale climatic signal despite differences in breakpoint detection and adjustment procedures.

Methodological differences become more apparent in data-sparse regions. Several limitations should be acknowledged. First, the spatial coverage of early instrumental observations remains highly uneven, with Europe being substantially better documented than many tropical and high-latitude regions. Second, the analysis relied exclusively on automated homogenisation methods; incorporating metadata-based or semi-automated approaches could further improve results for historically complex series. Third, only monthly temperature data and a single reanalysis product (20CRv3) were evaluated. Future work should therefore include daily observations, additional reanalyses such as ERA5 and JRA-55, and further exploration of machine-learning and data-assimilation approaches for breakpoint detection and adjustment.

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

More importantly, the objective is not to identify which method most closely reproduces the original HCLIM dataset. Because HCLIM itself contains gaps and varying temporal coverage, the more relevant question is how effectively different methods reconstruct a climatically plausible and temporally coherent climate signal. The homogenised datasets show broad agreement in their representation of historical temperature variability. Overall, the results demonstrate that large-scale automated homogenisation is feasible even for highly heterogeneous early instrumental datasets such as HCLIM. While methodological differences remain, particularly in data-sparse regions, robust climatic signals can nevertheless be recovered from incomplete historical observations. This represents an important step towards the systematic integration of early instrumental records into climate reconstructions and reanalysis frameworks. These findings demonstrate that automated homogenisation provides a practical pathway for incorporating fragmented early instrumental observations into future climate reconstructions and reanalysis products, thereby extending the observational basis for understanding long-term climate variability.

5 Conclusions

In this study, three automated homogenisation methods (CLIMATOL, PHA, and BART) were applied to the HCLIM dataset, and the resulting homogenised datasets were evaluated using breakpoint analyses, comparisons with the 20CRv3 reanalysis, and RMSD-based assessments. The pre-processed datasets showed broadly consistent characteristics, with CLIMATOL and PHA retaining nearly all HCLIM records (96–99 %), whereas BART retained fewer records (≈80 %). Breakpoint analyses indicated that BART detected the largest number of breakpoints, whereas CLIMATOL and PHA produced fewer breakpoint detections overall. Comparisons with 20CRv3 and RMSD analyses further demonstrated that the methods differ in their adjustment behaviour, with outcomes varying according to network density, temporal coverage, and data availability.

The results highlight important contrasts between dense and sparse observational networks. The French network, characterised by high station density and relatively homogeneous observing practices, yielded more consistent results across all methods. In contrast, the South-East Asian network exhibited greater spatial heterogeneity, larger observational uncertainties, and stronger methodological differences. Nevertheless, all three homogenisation methods preserved the large-scale climate signal and converged towards similar representations of historical temperature variability despite differences in breakpoint detection and adjustment strategies. Overall, the results demonstrate that large-scale automated homogenisation is feasible even for highly heterogeneous early instrumental datasets such as HCLIM. While methodological differences remain, particularly in data-sparse regions, robust climatic signals can be recovered from incomplete historical observations. These findings demonstrate that automated homogenisation provides a practical framework for improving the usability of fragmented early instrumental observations, thereby supporting their integration into future climate reconstructions and reanalysis products.

Practical implications

More broadly, these results highlight the value of comparing multiple homogenisation approaches when evaluating historical climate records. Cross-validation across methods reduces methodological uncertainty, increases confidence in long-term trends, and provides a more robust basis for analysing climate variability and change across regions and centuries. The strong agreement among the homogenised datasets further suggests that

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

robust climatic signals can be reconstructed consistently despite differences in breakpoint detection and adjustment strategies.

Our findings also emphasise the importance of homogenised station datasets for studies of historical climate variability. Early instrumental observations often contain gaps, inconsistencies, and changes in observational practice, and homogenisation plays a critical role in reconstructing temporally coherent climate records. While reanalyses such as 20CRv3 provide invaluable large-scale spatial coverage and physically consistent climate fields, their quality remains dependent on the availability and distribution of assimilated observations, particularly during the early instrumental period. The Jakarta case study illustrates how sparse observational coverage can limit the reliability of reanalysis products at regional scales.

Consequently, homogenised station datasets such as HOM-HCLIM; https://doi.pangaea.de/10.1594/PANGAEA.988345, dataset in review) (Lundstad et al., in review) should be regarded as an essential foundation for historical climate analyses whenever observational records are available. At the same time, reanalyses provide important complementary information by placing local observations into a broader spatial context. The combined use of homogenised observations and reanalysis products therefore offers a particularly powerful framework for reconstructing past climate variability and long-term climate change.

Code and data availability

The R code used for formatting, quality control, breakpoint detection, and homogenisation is publicly available at: https://github.com/elinlun/hom-HCLIM. The homogenisation software CLIMATOL is openly accessible via CRAN (CLIMATOL): https://cran.r-project.org/web/packages/climatol/index.html (CRAN, 2022). Information about other codes can be obtained from the corresponding author upon request. The STL codes: https://github.com/hafen/stlplus. The instrumental series are stored in monthly SEF files in HCLIM database: The best homogenised dataset can be downloaded here: https://doi.pangaea.de/10.1594/PANGAEA.988345.

The 20CRv3 data set is available from:
https://www.psl.noaa.gov/data/gridded/data.20thC_ReanV3.html?utm_source=chatgpt.com

Supplement link

S1. Table
S2. Table
S3. Table
S4. Figure and list for Station Network

Author contributions

E.L. designed and wrote the paper series and finalized all the figures with contributions from all co-authors. A.D. contributed to paper with knowledge about 20crv, M.T. contributed with the knowledge in the PHA method and experienced advice in the homogenisation method, M.J. contributed with the knowledge in BART method.

Competing interests

The corresponding author declares that none of the authors has any competing interests.

Disclaimer

Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in the text, published maps, institutional affiliations, or any other geographical representations in this paper. While

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgments

This study is connected to and partly funded by the ERC project PALAEO-RA, initiated and led by Stefan Brönnimann from the Oeschger Centre for Climate Change Research at the University of Bern.
The authors gratefully acknowledge José A. Guijarro for sharing the homogenisation program CLIMATOL, assistance with its installation, and continued technical support.
Special thanks go to Jianquan Dong from the University of Bern for technical assistance, to Peter Stucki, from the Oeschger Centre for Climate Research for proofreading, guidance on figure preparation (especially Fig. 6) and to Svein Olav Krøgli from Nibio, Norway for concrete advice with the statistical methods in this paper.
This work would not have been possible without the practical support and many productive exchanges with the colleagues from the Norwegian Meteorological institute, namely Cristian Lussana, Anita Dyrdal, Inger Hansen Bauer, Franceso Cavalleri, Herman Fæhn Fuglestvedt, Reidun Gangstø, Ketil Isaksen, Elinah Khasandi Kuya, Karianne Ødemark and Ole Einar Tveito. All reconstructions and analyses presented in this data descriptor were conducted using free and open-source software, primarily R and Python. The authors acknowledge the use of ChatGPT (OpenAI GPT-5.5) for language editing and proofreading. The authors are solely responsible for the scientific content of the manuscript. Support for the Twentieth Century Reanalysis version 3 dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), the National Oceanic and Atmospheric Administration Climate Program Office, and the NOAA Physical Sciences Laboratory.

Financial support

This work was supported by the Norwegian Meteorological Institute through payroll funding. The study builds on earlier and ongoing research supported by the European Research Council under the European Union's Horizon 2020 programme (grant agreement No. 787574, PALAEO-RA), the Swiss National Science Foundation (project WeaR 188701), the Copernicus Climate Change Service (C3S) contract 311a Lot 1, and the Federal Office of Meteorology and Climatology MeteoSwiss within the framework of GCOS Switzerland (project "Long Swiss Meteorological Series").

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., and Wieringa, J. *Guidelines on Climate Metadata and Homogenization.* WMO-TD No. 1186, WCDMP No. 53. World Meteorological Organization, Geneva, Switzerland, 2003.

Alexandersson, H.: A homogeneity test applied to precipitation data, Journal of Climatology, 6, 661–675, https://doi.org/10.1002/joc.3370060607, 1986.

Brandsma, T. and Können, G. P.: Application of nearest-neighbor resampling for homogenizing temperature records on a daily to sub-daily level, International Journal of Climatology, 26, 75–89, https://doi.org/10.1002/joc.1236, 2006.

Brunet, M. and Jones, P.: Data rescue initiatives: Bringing historical climate data into the 21st century, Climate Research, 47, 29–40, https://doi.org/10.3354/cr00960, 2011.

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

Brázdil, R., Dobrovolný, P., Luterbacher, J., Moberg, A., Pfister, C., Wheeler, D., and Zorita, E.: European climate of the past 500 years: New challenges for historical climatology, Climatic Change, 101, 7–40, https://doi.org/10.1007/s10584-009-9783-z, 2010.

Brönnimann, S.: Climatic Changes Since 1700, vol. 55 of Advances in Global Change Research, Springer, Cham, https://doi.org/10.1007/978-3-319-19042-6, 2015.

Brönnimann, S., Brugnara, Y., Allan, R. J., Brunet, M., Compo, G. P., Crouthamel, R. I., Jones, P. D., Jourdain, S., Luterbacher, J., Siegmund, P., Valente, M. A., and Wilkinson, C. W.: A roadmap to climate data rescue services, Geoscience Data Journal, 5, 28–39, https://doi.org/10.1002/gdj3.56, 2018.

Brönnimann, S., Allan, R., Ashcroft, L., Baer, S., Barriendos, M., Brázdil, R., Brugnara, Y., Brunet, M., Brunetti, M., Chimani, B., Cornes, R., Domínguez-Castro, F., Filipiak, J., Founda, D., Herrera, R. G., Gergis, J., Grab, S., Hannak, L., Huhtamaa, H., Jacobsen, K. S., Jones, P., Jourdain, S., Kiss, A., Lin, K. E., Lorrey, A., Lundstad, E., Luterbacher, J., Mauelshagen, F., Maugeri, M., Maughan, N., Moberg, A., Neukom, R., Nicholson, S., Noone, S., Nordli, Ø., Ólafsdóttir, K. B., Pearce, P. R., Pfister, L., Pribyl, K., Przybylak, R., Pudmenzky, C., Rasol, D., Reichenbach, D., Řezníčková, L., Rodrigo, F. S., Rohr, C., Skrynyk, O., Slonosky, V., Thorne, P., Valente, M. A., Vaquero, J. M., Westcott, N. E., Williamson, F., and Wyszyński, P.: Unlocking pre-1850 instrumental meteorological records a global inventory, *Bulletin of the American Meteorological Society*, 100, S1–S36, https://doi.org/10.1175/BAMS-D-19-0040.1, 2019.

Brönnimann, S.: Historical Observations for Improving Reanalyses, Front. Clim. 4:880473. https://doi.org/10.3389/fclim.2022.880473, 2022.

Camuffo, D., della Valle, A., and Becherini, F.: Instrumental and Observational Problems of the Earliest Temperature Records in Italy: A Methodology for Data Recovery and Correction, Climate, 11, https://doi.org/10.3390/cli11090178, 2023.

Cao, L.-J. and Yan, Z.-W.: Progress in research on homogenization of climate data, Advances in Climate Change Research, 3, 59–67, https://doi.org/10.3724/SP.J.1248.2012.00059, 2012.

Changnon, S. A. and Kunkel, K. E.: Changes in instruments and sites affecting historical weather records: A case study, Journal of Atmospheric and Oceanic Technology, 23, 825–828, https://doi.org/10.1175/JTECH1876.1, 2006.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin,P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H.-T., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century625 Reanalysis Project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28, https://doi.org/10.1002/qj.776, 2011.

Conrad, V.: Methods in Climatology, Weather, 1, 84–84, https://doi.org/10.1002/j.1477-8696.1946.tb00041.x, 1946.

Coscarelli, R., Caroletti, G. N., Joelsson, M., Engström, E., and Caloiero, T.: Validation metrics of homogenization techniques on artificially inhomogenized monthly temperature networks in Sweden and Slovenia (1950–2005), Scientific Reports, 11, https://doi.org/10.1038/s41598-021-97685-7, 2021.

Costa, A. C. and Soares, A.: Homogenization of climate data: Review and new perspectives using geostatistics,

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

https://doi.org/10.1007/s11004-008-9203-3, 2009.
Dobrovolný, P., Moberg, A., Brázdil, R., Pfister, C., Glaser, R., Wilson, R., van Engelen, A., Limanówka, D., Kiss, A., Halíčková, M., Macková, J., Riemann, D., Luterbacher, J., and Böhm, R. Monthly, seasonal and annual temperature reconstructions for Central Europe derived from documentary evidence and instrumental records since AD 1500. *Climatic Change*, 101, 69–107. https://doi.org/10.1007/s10584-009-9724-x
Domonkos, P.: Combination of using pairwise comparisons and composite reference series: A new approach in the homogenization of climatic time series with ACMANT, Atmosphere, 12, https://doi.org/10.3390/atmos12091134, 2021.
Domonkos, P.: Time Series Homogenization with ACMANT: Comparative Testing of Two Recent Versions in Large-Size Synthetic Temperature Datasets, Climate, 11, 224, https://doi.org/10.3390/cli11110224, 2023.
Easterling, D. R., Kunkel, K. E., Wehner, M. F., and Sun, L.: Detection and attribution of climate extremes in the observed record, Weather and Climate Extremes, 11, 17–27, https://doi.org/10.1016/j.wace.2016.01.001, 2016.
Fleming, J. R.: Meteorology in America, 1800–1870, Johns Hopkins University Press, Baltimore, https://doi.org/https://doi.org/10.56021/9780801839580, 1990.
Gao, X., Sokolov, A., and Schlosser, C. A.: A Large Ensemble Global Dataset for Climate Impact Assessments, Scientific Data, 10, https://doi.org/10.1038/s41597-023-02708-9, 2023.
Gubler, S., Hunziker, S., Begert, M., Croci-Maspoli, M., Konzelmann, T., Brönnimann, S., Schwierz, C., Oria, C., and Rosas, G.: The influence of station density on climate data homogenization, International Journal of Climatology, 37, 4670–4683, https://doi.org/10.1002/joc.5114, 2017.
Guijarro, J. A.: Homogenization of climatic series with Climatol, Report State Meteorological Agency. Balearic Islands Office Spain climatol2.0.tar.gz, StateM eteorologicalAgency(AEM ET), 2011.
Guijarro, J. A.: Homogenization of climatic series with Climatol, https://doi.org/10.13140/RG.2.2.27020.41604, 2018.
Guijarro, J. A. P.: Software libre para la depuración y homogeneización de datos climatológicos, Tech. rep., 2004.
Hafen, R. (2016). *stlplus: Enhanced Seasonal Decomposition of Time Series by Loess*. R package version 0.5.1. Available at: https://CRAN.R-project.org/package=stlplus
Hafen, M. R.: Package 'stlplus' Type Package Title Enhanced Seasonal Decomposition of Time Series by Loess, Tech. rep., 2025.
Hannart, A., Mestre, O., and Naveau, P.: An automatized homogenization procedure via pairwise comparisons with application to Argentinean temperature series, International Journal of Climatology, 34, 3528–3545, https://doi.org/10.1002/joc.3925, , 2014.
Harper, K. C.: Weather by the Numbers: The Genesis of Modern Meteorology, MIT Press, Cambridge, MA, 2008.
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049. https://doi.org/10.1002/qj.3803, 2020.

Joelsson, L. M. T., Sturm, C., Södling, J., Engström, E., and Kjellström, E.: Automation and evaluation of the interactive homogenization tool HOMER, International Journal of Climatology, 42, 2861–2880, https://doi.org/10.1002/joc.7394, 2022.

Joelsson, L. M. T., Engström, E., and Kjellström, E.: Homogenization of Swedish mean monthly temperature series 1860–2021, International Journal of Climatology, 43, 1079–1093, https://doi.org/10.1002/joc.7881, 2023.

Jones, P.: The reliability of global and hemispheric surface temperature records, https://doi.org/10.1007/s00376-015-5194-4, 2016.

Jones, P. D. and Lister, D. H.: The Daily Temperature Record for St. Petersburg (1743–1996), International Journal of Climatology, 22, 1717–1726, https://doi.org/10.1002/joc.816, 2002.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General specifications and basic characteristics, Journal of the Meteorological Society of Japan, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Lindau, R.: Estimation of Break and Noise Variance and the Maximum Distance of Climate Stations Allowed in Relative Homogenisation of Annual Temperature Anomalies, International Journal of Climatology, 45, https://doi.org/10.1002/joc.8724, 2025.

Lindau, R. and Venema, V.: A new method to study inhomogeneities in climate records: Brownian motion or random deviations?, International Journal of Climatology, 39, 4769–4783, https://doi.org/10.1002/joc.6105, 2019.

Lindau, R. and Venema, V.: The uncertainty of break positions detected by homogenization algorithms in climate records, International Journal of Climatology, 36, 576–589, https://doi.org/10.1002/joc.4366, 2016.

Lundstad, E. and Tveito, O. E.: Homogenization of Daily Mean Temperature in Norway, Tech. Rep. 6, Norwegian Meteorological Institute, Oslo, Norway, 2016.

Lundstad, Elin; Brugnara, Y. B. S.: Global Early Instrumental Monthly Meteorological Multivariable Database (HCLIM), https://doi.org/10.1594/PANGAEA.940724, 2022.

Lundstad, E., Brugnara, Y., Pappert, D., Kopp, J., Samakinwa, E., Hürzeler, A., Andersson, A., Chimani, B., Cornes, R., Demarée, G., Filipiak, J., Gates, L., Ives, G. L., Jones, J. M., Jourdain, S., Kiss, A., Nicholson, S. E., Przybylak, R., Jones, P., Rousseau, D., Tinz, B., Rodrigo, F. S., Grab, S., Domínguez-Castro, F., Slonosky, V., Cooper, J., Brunet, M., and Brönnimann, S.: The global historical climate database HCLIM, Scientific Data, 10, https://doi.org/10.1038/s41597-022-01919-w, 2023.

Menne, M. J. and Williams, C. N.: Homogenization of temperature series via pairwise comparisons, Journal of Climate, 22, 1700–1717, https://doi.org/10.1175/2008JCLI2263.1, 2009.

Menne, M. J., Williams, C. N., and Vose, R. S.: The U.S. Historical Climatology Network monthly temperature data, version 2, Bulletin of the American Meteorological Society, 90, 993–1007, https://doi.org/10.1175/2008BAMS2613.1, 2009.

Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Lebarbier, E., Böhm, R., Aguilar, E., Guijarro, J., Vertachnik, G., Klancar, M., Dubuisson, B., and Stepanek, P.: HOMER : a homogenization software-methods

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

and applications, Tech. Rep. 117, Hungarian Meteorological Service, http://hdl.handle.net/20.500.11765/1494, issue: 1 Pages: 47-67 Publication Title: ID ˝OJÁRÁS Quarterly Journal of the Hungarian Meteorological Service Volume: 117, 2013.

Nordli, Ø. and Tveito, O. E.: Calculation of Monthly Mean Temperature by Köppen's Formula in the Norwegian Station Network, Tech. Rep. 18, Norwegian Meteorological Institute, Oslo, Norway, 2008.

Osborn, T. J., Jones, P. D., Lister, D. H., Morice, C. P., Simpson, I. R., Winn, J. P., Hogan, E., and Harris, I. C.: Land Surface Air Temperature Variations Across the Globe Updated to 2019: The CRUTEM5 Data Set, Journal of Geophysical Research: Atmospheres, 126, https://doi.org/10.1029/2019JD032352, 2021.

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., Štěpánek, P., Wypych, A., and Zahradníček, P.: Evaluation of the Homogenization Adjustments Applied to European Temperature Records in the Global Historical Climatology Network Dataset, Atmosphere, 13, https://doi.org/10.3390/atmos13020285, 2022.

Parker, D. E.: Climate Observations – The Instrumental Record, technical report, 2000.

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullet, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Førland, E. J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., and Parker, D.: Homogeneity adjustments of in situ atmospheric climate data: A review, International Journal of Climatology, 18, 1493–1517, https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T, 1998.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002jd002670, 2003.

Ribeiro, S., Caineta, J., and Costa, A. C.: Review and discussion of homogenisation methods for climate data, https://doi.org/10.1016/j.pce.2015.08.007, 2016.

Shen, L., Lu, L., Hu, T., Lin, R., Wang, J., and Xu, C.: Homogeneity test and correction of daily temperature and precipitation data (1978-2015) in North China, Advances in Meteorology, 2018, https://doi.org/10.1155/2018/4712538, 2018.

Siswanto, S., van Oldenborgh, G. J., van der Schrier, G., Jilderda, R., and van den Hurk, B.: Temperature, extreme precipitation, and diurnal rainfall changes in the urbanized Jakarta city during the past 130 years, International Journal of Climatology, 36, 3207–3225, https://doi.org/10.1002/joc.4548, 2016.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Blancq, F. L., Lee, T. C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. W., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszy´nski, P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Quarterly Journal of the Royal Meteorological Society, 145, 2876–2908, https://doi.org/10.1002/qj.3598, 2019.

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., Brohan, P., Yin, X., Smith, C. A., Spencer, L. J., Vose, R. S., Rohrer, M., Conroy, R. P., Schuster, D. C., Kennedy, J. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Kubota, H., Lee, T. C., Lorrey, A. M., Luterbacher, J., Mock, C. J., Przybylak, R. K., Pudmenzky, C., Slonosky, V. C., Tinz, B., Trewin, B., Wang, X. L., Wilkinson, C., Wood, K., and Wyszynski, P.: An evaluation of the performance of the twentieth century reanalysis version 3, Journal of Climate, 34, 1417–1438, https://doi.org/10.1175/JCLI-D-20-0505.1, 2021.

Solomon, S. et al. (Eds.). (2007). *Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2007.

Sturm, C.: Spatially aggregated climate indicators over Sweden (1860-2020), Part 1: Temperature, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-582, 2024.

Taylor, K. E. *Summarizing multiple aspects of model performance in a single diagram*. Journal of Geophysical Research: Atmospheres, 106(D7), 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Toreti, A., Kuglitsch, F. G., Xoplaki, E., Della-Marta, P. M., Aguilar, E., Prohom, M., and Luterbacher, J.: A note on the use of the standard normal homogeneity test to detect inhomogeneities in climatic time series, International Journal of Climatology, 31, 630–632, https://doi.org/10.1002/joc.2088, 2011.

Trewin, B.: Exposure, instrumentation, and observing practice effects on land temperature measurements, Wiley Interdisciplinary Reviews: Climate Change, 1(4), 490–506, https://doi.org/10.1002/wcc.46, 2010.

Valler, V., Franke, J., Brugnara, Y., Samakinwa, E., Hand, R., Lundstad, E., Burgdorf, A. M., Lipfert, L., Friedman, A. R., and Brönnimann, S.: ModE-RA: a global monthly paleo-reanalysis of the modern era 1421 to 2008, Scientific Data, 11, https://doi.org/10.1038/s41597-023-74502733-8, 2024.

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., Duran, M. P., Likso, T., Esteban, P., and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, Climate of the Past, 8, 89–115, https://doi.org/10.5194/cp-7508-89-2012, 2012.

Vincent, L. A., Milewska, E. J., Wang, X. L., and Hartwell, M. M.: Uncertainty in homogenized daily temperatures and derived indices of extremes illustrated using parallel observations in Canada, International Journal of Climatology, 38, 692–707, https://doi.org/10.1002/joc.5203, 2018.

Wang, F., Ge, Q., Wang, S., Li, Q., and Jones, P. D.: A new estimation of urbanization's contribution to the warming trend in China, Journal of Climate, 28, 8923–8938, https://doi.org/10.1175/JCLI-D-14-00427.1, 2015.

Wilkinson, C., Brönnimann, S., Jourdain, S., Roucaute, E., Crouthamel, R., Team, I., Brohan, P., Valente, A., Brugnara, Y., Brunet, M., and Compo, G. P.: Best Practice Guidelines for Climate Data Rescue, https://doi.org/10.25607/OBP-1513, guidelines for the rescue, digitisation, and quality control of historical climate data, 2016.

World Meteorological Organization: *Proceedings of the Sixth Seminar for Homogenization and Quality Control in Climatological Databases*, WMO-No. 1575, World Meteorological Organization, Geneva, Switzerland, 2008.

World Meteorological Organization: *Guide to Climatological Practices*, WMO-No. 100, 2011 edn., World Meteorological Organization, Geneva, Switzerland, 2011.

https://doi.org/10.5194/egusphere-2026-3936
Preprint. Discussion started: 31 July 2026

World Meteorological Organization: *WMO Guidelines on the Calculation of Climate Normals*, WMO-No. 1203, World Meteorological Organization, Geneva, Switzerland, 2017.

World Meteorological Organization: *Guide to Instruments and Methods of Observation*, WMO-No. 8, World Meteorological Organization, Geneva, Switzerland, 2018.

World Meteorological Organization: *Guidelines on Homogenization*, WMO-No. 1245, World Meteorological Organization, Geneva, Switzerland, 2020.

Internet pages:

Station Exchange Format (SEF): https://climate.copernicus.eu/data-rescue-service).

Web site of the Task Team on HOMOGENIZATION: https://www.climatol.eu/tt-hom/ [reading date: 24th January 2024]

Guijarro, J. A. User's guide of the Climatol R Package. https://www.climatol.eu/climatol4-en.pdf [reading date: 6th Novembre 2023]

Venema's page: https://variable-variability.blogspot.com/2017/10/history-homogenization-climate-station-data.html [reading date: 5th Novembre 2023]

WMO: https://community.wmo.int/en/climate-data-homogenization [reading date: 2nd Octobre 2023]

WMO: https://public-old.wmo.int/en/resources/bulletin/critical-role-of-observations-informing-climate-science-assessment-and-policy [reading date: 24th November 2023]

Wikipedia: https://en.wikipedia.org/wiki/Homogenization_(climate) [reading date: 20.10.2023]

R Documentation for Seasonal Decomposition in R:
https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/stl
https://cran.r-project.org/web/packages/stlplus/stlplus.pdf