



This is a repository copy of *The global historical climate database HCLIM*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/195588/>

Version: Published Version

Article:

Lundstad, E. orcid.org/0000-0002-3443-2480, Brugnara, Y. orcid.org/0000-0001-8427-0064, Pappert, D. et al. (25 more authors) (2023) The global historical climate database HCLIM. *Scientific Data*, 10 (1). 44. pp. 1-16.

<https://doi.org/10.1038/s41597-022-01919-w>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



OPEN

The global historical climate database HCLIM

DATA DESCRIPTOR

Elin Lundstad^{1,2}✉, Yuri Brugnara^{1,2}, Duncan Pappert^{1,2}, Jérôme Kopp^{1,2}, Eric Samakinwa^{1,2}, André Hürzeler^{1,2}, Axel Andersson³, Barbara Chimani⁴, Richard Cornes⁵, Gaston Demarée⁶, Janusz Filipiak⁷, Lydia Gates³, Gemma L. Ives⁸, Julie M. Jones⁹, Sylvie Jourdain¹⁰, Andrea Kiss¹¹, Sharon E. Nicholson¹², Rajmund Przybylak^{13,14}, Philip Jones¹⁵, Daniel Rousseau¹⁶, Birger Tinz³, Fernando S. Rodrigo¹⁷, Stefan Grab¹⁸, Fernando Domínguez-Castro¹⁹, Victoria Slonosky²⁰, Jason Cooper^{12,21}, Manola Brunet^{15,22} & Stefan Brönnimann^{1,2}

There is a growing need for past weather and climate data to support science and decision-making. This paper describes the compilation and construction of a global multivariable (air temperature, pressure, precipitation sum, number of precipitation days) monthly instrumental climate database that encompasses a substantial body of the known early instrumental time series. The dataset contains series compiled from existing databases that start before 1890 (though continuing to the present) as well as a large amount of newly rescued data. All series underwent a quality control procedure and subdaily series were processed to monthly mean values. An inventory was compiled, and the collection was deduplicated based on coordinates and mutual correlations. The data are provided in a common format accompanied by the inventory. The collection totals 12452 meteorological records in 118 countries. The data can be used for climate reconstructions and analyses. It is the most comprehensive global monthly climate dataset for the preindustrial period so far.

Background & Summary

Long-term instrumental meteorological series are crucial for the understanding of interannual-to-decadal variations in climate. Analyzed together with model simulations and climate proxies they may provide new insight into underlying climate mechanisms, such as long-lasting droughts¹, changes in atmospheric circulation², or effects of volcanic eruptions³, and may serve as a basis for the generation of more comprehensive data products in reconstruction⁴ or data assimilation approaches^{5,6}. Long-term instrumental meteorological series also serve as a reference against which human induced climate change can be compared⁷. For instance, Hawkins et al.⁸ suggested using the period 1720–1800 as a preindustrial reference, but only few records from this period are currently available. We define a record as a meteorological time series with one variable at one location.

¹Institute of Geography, University of Bern, Bern, Switzerland. ²Oeschger Centre for Climate Change Research (OCCR), University of Bern, Bern, Switzerland. ³Deutscher Wetterdienst (DWD), Hamburg, Germany. ⁴Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienna, Austria. ⁵National Oceanography Centre, Southampton, United Kingdom. ⁶Royal Meteorological Institute of Belgium, Brussels, Belgium. ⁷Department of Meteorology and Climatology, Institute of Geography, University of Gdansk, Gdansk, Poland. ⁸University of Sheffield, Sheffield, United Kingdom. ⁹Department of Geography, University of Sheffield, Sheffield, United Kingdom. ¹⁰Météo-France, Toulouse, France. ¹¹Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria. ¹²Earth, Ocean, and Atmospheric Sciences (EOAS), Florida State University, Tallahassee, USA. ¹³Department of Meteorology and Climatology, Nicolaus Copernicus University, Toruń, Poland. ¹⁴Centre for Climate Change Research, Nicolaus Copernicus University, Toruń, Poland. ¹⁵Climatic Research Unit (CRU), University of East-Anglia (UEA), Norwich, United Kingdom. ¹⁶Conseil Supérieur de la Météorologie, Toulouse, France. ¹⁷Departamento de Química y Física, Universidad de Almería, La Cañada de San Urbano, Almería, Spain. ¹⁸School of Geography, Archaeology and Environmental Studies, University of Witwatersrand, Johannesburg, South Africa. ¹⁹Departamento de Geografía y Ordenación del Territorio, Universidad de Zaragoza & ARAID foundation, Zaragoza, Spain. ²⁰Centre for Interdisciplinary Research on Montréal, McGill University, Montreal, Québec, Canada. ²¹National Centers for Environmental Information (NCEI), National Oceanic and Atmospheric Administration (NOAA), Asheville, NC, USA. ²²Universidad Rovira i Virgili, Tarragona, Spain. ✉e-mail: elin.lundstad@giub.unibe.ch

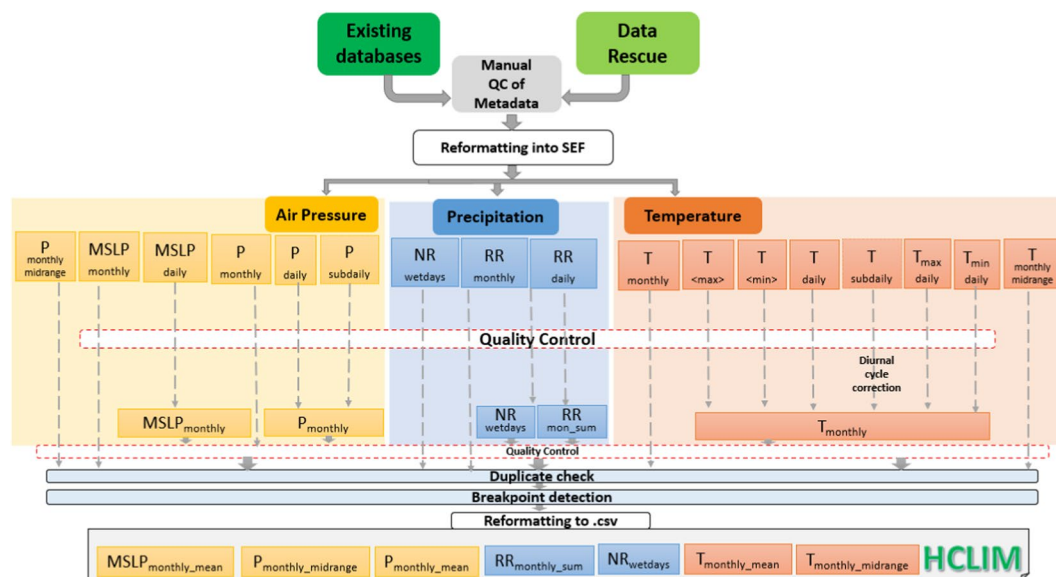


Fig. 1 Schematic diagram, showing the sequence of steps involved in the data collection.

While several global^{9,10}, regional^{11–13}, or national (e.g., Deutscher Wetterdienst (DWD) and the Royal Netherlands Meteorological Institute (KNMI), etc.) climate datasets exist that reach back to the 17th century, each of them has unique records and there is no comprehensive dataset across different variables. Furthermore, coverage of the early instrumental period is not as good as it could be. Brönnimann *et al.*¹⁴ compiled a global inventory of early instrumental meteorological measurements, showing that a large fraction of the series either have not been digitized at all or have not been integrated into global datasets. Here we follow-up on such work and present a comprehensive database of monthly early instrumental climate data.

By integrating digitally available datasets that have not so far been included in other databases and by rescuing a substantial fraction of previously non-digitalized material such as handwritten temperature logs, our database encompasses a much larger volume of early instrumental data than that from previous efforts. Although the database extends to the present, the focus is on the early instrumental periods and therefore all compiled series start before 1890, which is an appropriate year to distinguish early instrumental meteorological observations from more recent meteorological measurements in a global context¹⁴. In total, our database (hereafter HCLIM) comprises 12452 monthly time series of meteorological variables such as temperature (mean, maximum and minimum), air pressure, precipitation and number of wet days compiled from over 28 different source datasets complemented with 1525 newly digitized records.

All data were reformatted to a common format (SEF, described in the Format section under Data Records). Monthly means were formed from daily or subdaily data. All series then underwent a quality control (QC), followed by removing duplicates. The data are presented as monthly files in the database¹⁵.

The paper is structured as follows: Section 2 describes the method: compilation, data rescue, and processing; Section 3 summarizes the HCLIM database; Section 4 presents a technical validation and Section 5 provides the usage notes.

Methods

The processing chain is illustrated in Fig. 1. It starts with compiling data from existing databases, and rescued data. Subdaily data is processed to monthly means and then a quality control procedure (QC) is performed. Finally, the data are checked for duplicates and the detection of inhomogeneities.

Compilation of existing data. Data were compiled from 28 available databases listed in Tables 2, 3 and 4 and in addition 16 other small data sources (Supplementary Table 1). The datasets comprise databanks with a global scope providing 27905 records to our compilation. Regional (1800), national (538) and other (61) databanks provided further records. The numbers refer to those data series compiled, i.e., only data series reaching back before 1890 that are at least one year in length. In a later process (Sect. Removing duplicates), about two thirds of the records were identified as duplicates and removed. Abbreviations given in the tables (Tables 2, 3 and 4) onwards are explained here in Table 1.

Global climate databases. We used six global climate databases (Table 2). They are compiled by UK and US institutions. Some of them are global temperature datasets (e.g., ISTI), whilst others have sea level pressure (e.g., ISPD) or all meteorological parameters (e.g., GHCN). Some measurements are over 300 years in age. Overview of the geographical distribution of the records can be found in Supplementary Fig. 1.

Abbreviation	Parameter	Unit
Ta	Temperature, mean	°C (Celsius)
Tmin	Temperature, minimum	°C
Tmax	Temperature, maximum	°C
P	Air pressure, mean	hPa (Hectopascal)
MSLP	Mean sea Level Pressure	hPa
RR	Precipitation	mm (Millimetre)
NR	Number of wet days	Number

Table 1. Explanation of abbreviations in the Tables 2, 3 and 4.

Name	Acronyms	Spatial Coverage	Parameter (Number of records)	Resolution	Total number of records compiled	Number of records in the HCLIM dataset after removing duplicates	References	Data
Berkeley Earth Data	BEST	Global	ta (3835)	Monthly	6798	1767	10	46
			Tmax (1452)					
			Tmin (1511)					
Climatic Research Unit	CRU	Global	ta (1418)	Monthly	1418	1391	47	48
			P (83)		83	82		
Global Historical Climatology Network	GHCN	Global	ta (2870)	Monthly	3630	1049	9	49
			Tmax (372)					
			Tmin (388)					
			RR (2840)		2840	2654		
			P (24)		24	24		
A monthly global paleo-reanalysis (input)	EKF400	Global	ta (79)	Monthly	77	18	5	50
			P (101)		101	91		
			MSLP (39)		39	30		
			RR (76)		76	76		
The International Surface Pressure Databank	ISPD	Global (mostly Europe)	P (163)	Subdaily	163	118	51	52
			MSLP (258)		258	159		
International Surface Temperature Initiative	ISTI*	Global	Ta (7129)	Monthly	12398	2526	53	54
			Tmax (2634)					
			Tmin (2635)					

Table 2. Overview of Global Climate Databases. *Supplementary Fig. 2.

Regional and thematic databases. Regional and thematic databases have a narrower spatial focus (Table 3). For example, there are several sources related to the ACRE project (The international Atmospheric Circulation Reconstructions over the Earth)¹⁶. In addition, Supplementary Table 1 shows an overview of other smaller datasets or individual stations that were incorporated into HCLIM. Overview of the geographical distribution of these records can be found in Supplementary Fig. 3.

National weather institutes. Most countries have their own National Weather Service. These provide weather forecasts for civilian and military purposes and conduct research in meteorology, oceanography, and climatology. Many have developed good climate databases, from which data can be extracted. Table 4 lists those that we integrated into our compilation.

Data rescued. In addition to the data collected from various sources, we transcribed and digitized a large number of early instrumental records that were hitherto not available in digital form. Figure 2 provides maps of all the rescued records that have been digitized, categorized by sources (Fig. 2a), and start year (Fig. 2b). Figure 2c shows the time evolution of the number of digitized records and Fig. 2d shows the histogram of the length of the records that have been digitized.

The bulk of the rescued data comes from early data collections published by prominent meteorologists such as Heinrich Wilhelm Dove¹⁷ (1803–1879) or from early networks of weather stations such as those organized in the 18th century by the Royal Society in London, Société Royale de Médecine¹⁸ (Royal Society of Medicine) in Paris and the Societas Meteorologica Palatina¹⁹ (Palatine Meteorological Society) in Mannheim. The main data sources are listed in Table 5. The reason for the peaks in the number of digitized records around 1780 and 1850 can be found in Table 5. The first peak is due to the many records belonging to the networks of the Royal Society of Medicine and the Palatine Meteorological Society); the second peak arises from the Dove¹⁷ and Weselowskij collections. The first peak is a bit misleading because it does not represent a large increase in spatial coverage, being most of the stations located in France and Germany. The vast majority of rescued data are short records (<20 years, see Fig. 2b) that were overlooked in previous digitization efforts.

Name	Acronyms	Spatial Coverage	Monthly Resolution	Daily Resolution	References	Data
			Parameter (Number of records)	Parameter (Number of records)		
Historical climate observations in Canada: 18 th and 19 th century	ACRE	Canada		Ta (25)	55,56	57
				Tmax (9)		
				Tmin (12)		
				P (17)		
				MSLP (17)		
ACRE South Africa	ACRE	South-Africa	Ta (8)	Ta (4)	42	16
			Tmax (6)	Tmax (5)		
			Tmin (5)	Tmin (7)		
				RR (3)		
				P (12)		
Early instrumental meteorological observations in Switzerland	CHIMES	Switzerland		Ta (66)	58	20
				Tmax (12)		
				Tmin (15)		
				RR (14)		
				P (63)		
Deutscher Wetterdienst	DWD	Overseas data	Ta (13)		59	60
			P (13)			
			RR (1)			
European Climate Assessment & Dataset	ECAD	Europe	Ta (152)		61	62
Early meteorological records from Latin-America and the Caribbean during the 18 th and 19 th centuries	EMERLAC	Latin-America and the Caribbean	Ta (18)	Ta (2)	63	63
			Tmax (11)	Tmax (1)		
			Tmin (11)	Tmin (1)		
			MSLP (4)	P (1)		
			RR (15)			
Improved Understanding of Past Climatic Variability from Early Daily European Instrumental Sources	IM-PROVE	Europe		Ta (5)	64	65
				P (5)		
Historical Instrumental Climatological Surface Time Series of the greater Alpine Region	HISTALP	Alps in Europe	Ta (104)		66	32
Japan-Asia Climate Data Program	JCDP	Japan	Ta (6)		44	67
			MSLP (5)			
C3-EURO4M-MEDARE Mediterranean historical climate data	MEDARE	Mediterranean region	Ta (1)		68	69
				Tmax (20)		
				Tmin (20)		
				RR (20)		
				P (20)		
HARD 2.0 - Historical Arctic Database	NCU	Arctic	Ta (58)		70	71
			P (51)			
				Ta (16)		
				Tmax (2)		
				Tmin (2)		
SEA-historical climate data	SEA	Australia	Ta (9)	Ta (16)	72	73
			Tmax (3)	Tmax (2)		
			Tmin (3)	Tmin (2)		
			RR (8)	RR (3)		
			MSLP (2)	MSLP (2)		
			P (12)	P (4)		
	NR (7)					
Southeast Asian Climate Assessment & Dataset	SACA&D	Central & South-Asia	Ta (19)		74	75
Two-Century Precipitation Dataset for the Continent of Africa		Africa	RR (233)		11	76
Forts and Volunteer Observer Database	US FORTS	United States		Ta (288)	77,78	79

Table 3. Overview of other climate regional Databases.

Name of Institute	Acronyms	Spatial Coverage	Parameter (Number of records)	Web
Deutscher Wetterdienst	DWD	Germany	Ta (155)	45
Koninklijk Nederlands Meteorologisch Instituut	KNMI	Netherlands	Ta (25)	80
MétéoFrance	MétéoFrance	Mainland France	Ta (58)	81
MétéoFrance	MétéoFrance	French overseas territories	Ta (3)	82
			RR (39)	
Norsk Meteorologisk Institutt	MetNorway	Norway	Ta (71)	83
			RR (30)	
			P (16)	
Roshydromet Всероссийский научно-исследовательский институт	RIHMI-WDC	Russia	Ta (75)	84
Sveriges meteorologiska och hydrologiska institut	SMHI	Sweden	Ta (66)	85

Table 4. Overview of national weather institutions from which we have only monthly data.

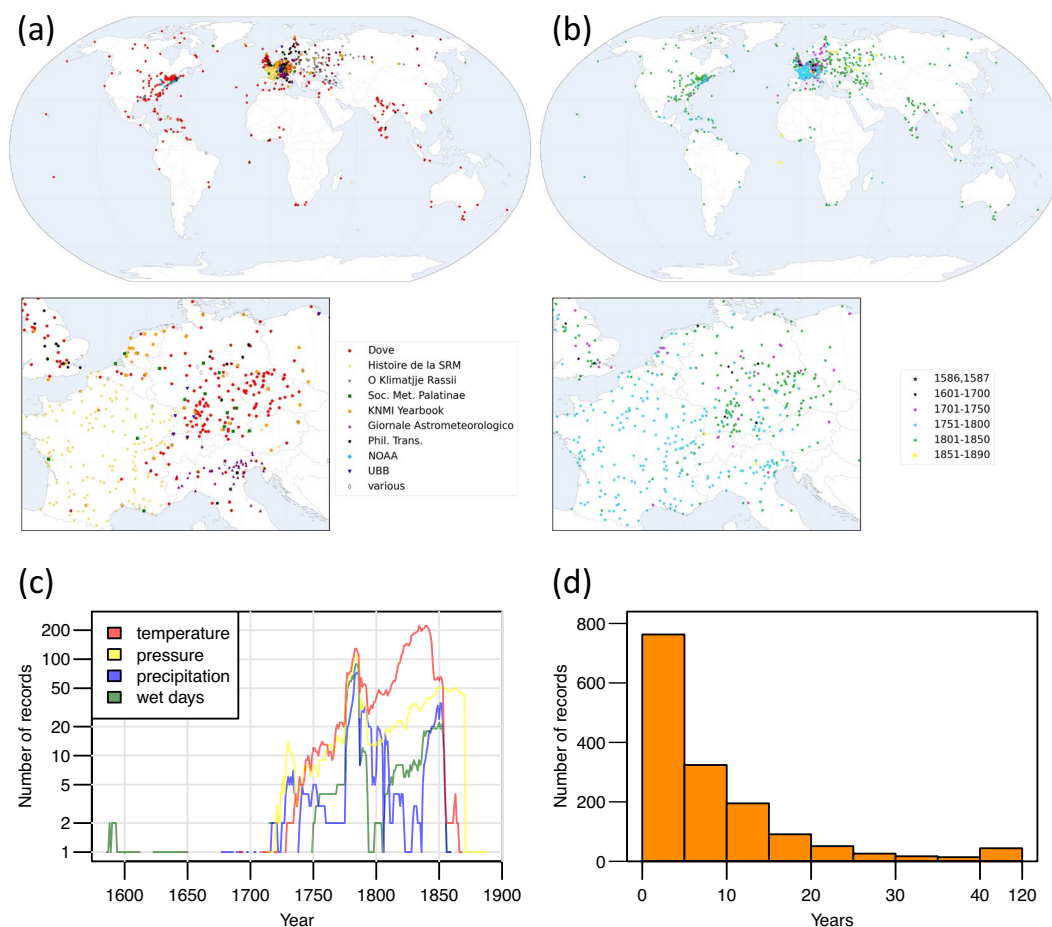


Fig. 2 (a) Spatial distribution of the newly digitized records organized by source and (b) Start year of the rescued records. (c) Time evolution of the number of digitized records. (d) Histogram of the length of the records that have been digitized.

The inventory compiled by Brönnimann *et al.*¹⁴ guided us to the selection of relevant records, which needed to be easily accessible (i.e., hard copy archived in Switzerland) or already available as digital images. We digitized data from 1,235 stations corresponding to 13,822 station years for different variables at various temporal resolutions, with some duplication. The actual typing was carried out by geography students at the University of Bern.

The conversion of outside temperature and pressure to modern units followed the general procedures described in Brugnara *et al.*^{20,21}. A few cases required additional efforts: for example, for the temperature record of Cambridge, Massachusetts (1742–1779), measured with a so-called “Hauksbee thermometer” (made for the Royal Society by Francis Hauksbee, the Younger; see²²), we used the parallel observations with a Fahrenheit

Source	Type	Period	Variables	Resolution	Domain	# of records	# of station years
H. W. Dove ¹⁷	Printed (periodical)	1719–1845	Ta	Monthly	Global	481	4992
Société Royale de Médecine ¹⁸	Printed (periodical)	1776–1786	Ta, P, RR, NR	Monthly	Global	392	1642
KNMI	Printed (yearbook)	1740–1870	P	Monthly	Northern Hemisphere	60	2232
K. Weselowskij	Printed (book)	1750–1853	Ta, RR, NR	Monthly	Russian Empire	198	2028
Societas Meteorologica Palatina ^{*19}	Printed (yearbook)	1781–1792	Ta, P, RR, NR, others	Subdaily, Daily, Monthly	Northern Hemisphere	128	705
NOAA Climate Database Modernization Program	Handwritten	1742–1822	Ta, P, RR, NR, others	Subdaily, Monthly	United States	54	1080
G. Toaldo ⁸⁶	Printed (periodical)	1778–1810	RR	Monthly	Italy, Alps	43	426
Universitätsbibliothek Basel	Handwritten	1739–1863	Ta, P, RR, NR, others	Subdaily, Monthly	Europe	22	289
Others	Various	1586–1888	Ta, P, RR, nr, others	Subdaily, Daily, Monthly	Global	147	944

Table 5. Overview of the newly digitized records. The numbers of records and station years refer to the variables included in HCLIM. A station year is defined as 12 months of non-missing data at a given station for a given variable. *Homogenized⁸⁷.

thermometer provided by the observer to build a conversion function to degrees Celsius. However, in general we discarded a large fraction of temperature records measured before ca. 1770 because of the large uncertainties in temperature scales and the lack of metadata on thermometers (these data can be obtained upon request in their original units).

We reduced pressure observations to normal gravity and, whenever possible, to 0°C. Pressure records that were not corrected for temperature are marked by a specific metadata entry.

Some of the rescued data were already available as monthly means in existing global datasets but have been retranscribed and digitized nonetheless, to ensure a better data quality and traceability, as well as to improve daily and subdaily data availability in future projects. Some of the oldest records are calculated according to the Julian calendar or are averages of monthly extremes. These instances are flagged accordingly in the metadata.

Following best practices in data rescue (e.g.²³), we digitized many additional variables that were observed alongside temperature and pressure. In particular: precipitation amount, precipitation type, monthly number of wet days, wind direction, wet bulb temperature, relative humidity, evaporation, snow depth, cloud cover, as well as qualitative weather descriptions. We digitized records for the number of wet days (i.e., days in which any precipitation was observed) from as early as 1586. Even though these are not strictly instrumental records, we considered them a valuable addition to the database.

The newly digitized raw data²⁴ – including over 2.2 million point observations, over 120,000 daily and over 180,000 monthly statistics – have been submitted to the Global Land and Marine Observations Dataset (GLAMOD^{25,26}) and will be freely available on the Copernicus Climate Change Service data store²⁷. An inventory is provided in the Supplementary Information of this paper.

From subdaily data to monthly averages. The calculation of monthly averages from subdaily observations followed two steps: 1) calculation of daily averages and 2) calculation of monthly averages from the daily averages.

To calculate daily averages, we took into account the time of observations and the effect of the diurnal cycle on averages. This is particularly important when only one observation per day is available, or when observation times are variable throughout the record.

We obtained the diurnal cycle from the nearest grid point in the ERA5-Land reanalysis, which provides hourly values of temperature and pressure since 1981 with a spatial resolution of ca. 9 km²⁸. We calculated a different diurnal cycle for each calendar month from the reference period 1981–2010. To correct the raw daily means calculated from available observations, we subtracted the average of the corresponding values in the diurnal cycle, after shifting its mean to zero. For example, a daily mean obtained from a single observation in the early morning – near the time of minimum temperature – will be increased by this correction. When the observation times are not known exactly, and for stations on very small islands not resolved in ERA5-Land, the correction is not applied. A metadata entry in the monthly data files informs the user on whether the diurnal cycle correction was applied or not. For precipitation amounts, the calculation of daily values is simply the sum of all observations within a 24-hour period.

Monthly averages and sums (for precipitation) are calculated from daily values following the criteria recommended by the World Meteorological Organization²⁹. The monthly average is set to missing if: (1) daily averages are missing for 11 or more days, or (2) daily averages are missing for a period of 5 or more consecutive days. Monthly precipitation sums are set to missing if any day is missing.

Quality control. The data and metadata (geographical coordinates) in HCLIM have been quality controlled. Quality control or QC is the process to detect and label suspicious or potentially wrong values. This is necessary to avoid possible errors within datasets that could compromise the results of subsequent analysis³⁰.

Parameter	Acronyms	Range
Temperature	ta, Tmax, Tmin (°C)	−60 to +40 °C
Air pressure	P (hPa)	<800 to >1050
Mean Sea Level Pressure	MSLP (hPa)	<980 to >1050
Precipitation	RR (mm, sum)	<0 to >1250
Number of rain days	NR (number)	<0 to > 28

Table 6. Upper and lower values check for the range check for different the parameters.

File	HCLIM	Name	Start	End	Parameter/Resolution
1	X	USFORTS_470273_18530801-18830331_ta_monthly.tsv	1853	1883	ta_monthly
2	X	GHCN_v4_qcf_GHCN_USW00094811_186901-201312_ta_monthly.tsv	1869	2013	ta_monthly
3	X	GHCN_v4_qcf_GHCN_USW00014837_186901-202112_ta_monthly.tsv	1869	2021	ta_monthly
4		GHCN_v4_qcf_GHCN_USC00474966_186901-199112_ta_monthly.tsv	1869	1991	ta_monthly
5		BERKELEY_EARTH-HadleyCentreCRUpartialdatacollection_BE-68168_186901-201105_ta_monthly.tsv	1869	2011	ta_monthly
6		BERKELEY_EARTH-WorldMonthlySurfaceStationClimatology_BE-64124_186901-200412_ta_monthly.tsv	1869	2004	ta_monthly
7		ISTI-WMSSC_726410_Madison_Truxax_Wis_1869_2013_ta_monthly.tsv	1869	2013	ta_monthly
8		ISTI-CRU_726410_Madison_Wisconsin_1869_2013_ta_monthly.tsv	1869	2013	ta_monthly

Table 7. List of records from Madison. X represents inclusion in the HCLIM merging; no selection means they are excluded from merging to this particular Madison record in HCLIM.

All metadata are deposited in a user-friendly inventory for this purpose. Information in this inventory includes station ID, name, latitude, longitude, elevation, start and end years of the time series, source, link, variable, temporal statistics (e.g., average, sum, etc.), unit (e.g., °C, mm) and other information.

The QC of the metadata in the inventory is undertaken by limit tests for latitude, longitude, and elevation, starting and ending dates of the series, variable names and units and cross-checks of the inserted country and the latitude and longitude.

For the QC of the data, we apply the following tests to each variable:

1. Range checks based on constant values. The range is shown in Table 6. This includes a check of physically impossible values such as negative values for precipitation.
2. Climatological outlier checks based on standard deviation, which requires at least 5 years of data. We use a threshold of 5 standard deviations.

The values that fail these tests are then confirmed manually before being flagged in the Station Exchange Format (SEF³¹) (described in the Format section under Data Records). The newly digitized data underwent additional quality checks at subdaily and daily resolution as described in Brunet *et al.*³⁰.

Removing duplicates. The next step was to create an algorithm that recognizes duplicates in the dataset. The same data can appear in several files. These can be copies of identical data compiled from several sources, different datasets (e.g., several observers in the same city), datasets supplemented with data from another city, or all possible combinations thereof (differently merged datasets). We have some examples of data from one meteorological station appearing 19 times within the 28 different databases.

The records were grouped by parameter (temperature, air pressure, precipitation, and number of wet days). Within each group we first calculated a distance (d) matrix. The second step was to calculate correlations for all pairs having d < 50 km. A threshold value of >0.98 for the pair of records was set to define a duplicate. The records fulfilling both the distance and correlation criteria were then included in a merge list of the target record. Proceeding record by record, merge lists were generated for each record, and a merged record was generated according to a priority list described below. Records included in a merge were excluded from the procedure when proceeding to the next record.

Note that we did not use the station name to identify duplicates. This is because the same station might have different names, locations exist in many different languages and spelling, and different locations may have the same station name.

In each merge list, highest priority was given to the records having the earliest start date. These records were extended (or gap filled) with data series starting later. In addition, homogenized sources (of which there are few) were prioritized (e.g., HISTALP³²). In case of identical start years (usually indicating identical data) we proceeded alphabetically. We further show an example from Madison, Wisconsin (USA) of how this method works and how the merging part of the removing duplicates takes place. Table 7 shows the merge list for this station. Note that three other stations with the name “Madison” exist (Fig. 4b) but are not in Wisconsin and represent different stations. As we do not include station names in the criteria to search duplicates, these stations were

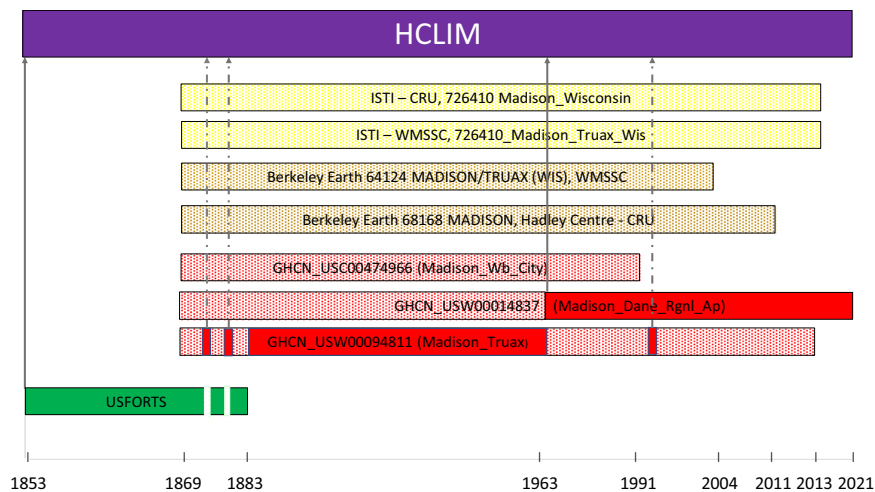


Fig. 3 Records that were checked for duplication, deduplicated and merged from Madison.

treated as separate stations as they are further away than 50 km and have a correlation below 0.98 with Madison Wisconsin.

Eight meteorological stations from Madison had temperature measurements and were tested for duplication and combined or merged, as indicated in Fig. 3.

Three stations are included in this merged Madison record. First the record from US Forts, because the record starts first or is the oldest (this is shown in Fig. 3, highlighted in green), and then two time series from GHCN (highlighted in red). The gray arrows indicate when the time series starts. Periods or observations marked with a dashed arrow show observations that are included when the time series has gaps. The Madison station (GHCN_USW00094811 Madison Truax) had some small gaps in the 1860's and a large gap after 1963. Consequently, another GHCN record (GHCN_USW00014837 - Rgnl_Ap) becomes the dominant record through the merging process.

Breakpoint detection. The removing of duplicates causes some records to be merged, as in the example in Figs. 3,4a. This in turn can introduce large inhomogeneities in the data. We flag them using a Welch's t-test on a 5-year moving window applied to monthly anomalies. The point in time where the inhomogeneity occurs, or breakpoint, is where the maximum of the absolute value of the test statistic occurs. The procedure is similar to the Standard Normal Homogeneity Test³³, but we require the size of the inhomogeneity to be larger than the average of the standard deviations of the two data segments that are separated by the breakpoint. In addition, we consider data gaps of 10 years or longer as breakpoints independently of the results of the statistical test.

Figure 4a shows the merged temperature record from Madison (1853–2021) with 6 breakpoints found through breakpoint detection (1877, 1882, 1952, 1997, 2012 and 2015). Only one of the breakpoints (1882) corresponds to a merging point. It is associated with a large step inhomogeneity. We did not homogenize data in HCLIM, but we provide the breakpoints and the merging information, both of which can be used for homogenization.

Data Records

Overview. After eliminating obvious duplicates and applying the 'removing duplicates' algorithm, we ended up with 12452 merged meteorological time series across 4 parameters. These series constitute the HCLIM dataset.

The HCLIM dataset has been deposited in a public repository and can be easily downloaded from the site¹⁵. Table 8 provides an overview of the numbers of records downloaded for the various parameters in each step of the data processing in the HCLIM dataset.

Table 9 indicates how many years and stations years are included in the HCLIM dataset. In total there are over one million station years, of which 148,843 are before 1891 (Table 8). The largest numbers, both in terms of stations years and number of stations, concern precipitation (Table 8). The variable with the least number of stations and station years is pressure.

Format. All data were reformatted to the Station Exchange Format (SEF). This is a format introduced by the Copernicus Climate Change Service³¹. It provides a simple but standard format for the distribution of historical weather data. SEF files have a .tsv format (tab-separated values) and list basic metadata regarding the station and the data manipulation in a header. The SEF is designed for rescuing observations and present them for wide-spread use in an uncomplicated format and made accessible through publicly available software. The aim of such SEF files is that they can be easily integrated into global repositories³¹. An example is shown in Fig. 5. The data are also available in a single compact flat file (.csv format) where, however, no metadata are provided. Geographical coordinates can be retrieved from an inventory file.

Temporal and spatial coverage. All the earliest meteorological records started in Europe. The first record is for wet days in Resterhufe-Osteel (Germany), which started in 1586. The first instrumental records started in

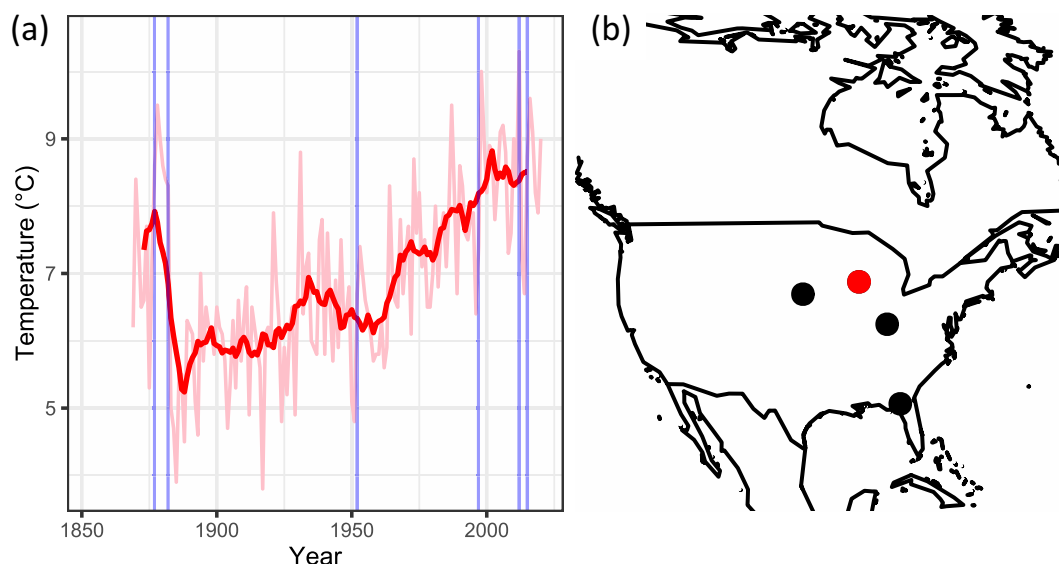


Fig. 4 (a) The merged temperature record from Madison with homogeneity breaks after the breakpoint detection. The thick line shows the rolling mean or moving average, here based on every decade. (b) Meteorological stations named Madison in four states in the USA (Wisconsin, Nebraska, Indiana, and Florida).

Parameter	Total downloaded	Unique records before merging (original data)	After merging => HCLIM dataset
Temperature (Ta)	17228	7573	3632
Air pressure (P)	827	827	806
Mean Sea Level Pressure (MSLP)	317	317	806
Precipitation (RR)	6378	6378	4943
Number of rain days (NR)	3155	3155	3071
Sum	27905	18250	12452

Table 8. Overview of the number of records through the development of HCLIM from downloading and digitization to HCLIM dataset.

Parameter	Station years total	Stations years before 1891	Number of stations
Temperature	328,794	56,384	3632
Pressure	33,264	15,406	806
Precipitation	441,701	57,288	4943
Wet days	215,692	19,931	3071

Table 9. Overview of number of station years and records after the removing duplicates.

1658 for temperature, 1670 for air pressure and 1688 for precipitation; all of which are from Paris. Table 10 lists the oldest records in HCLIM, organized continentally. For example, the first thus far known temperature measurements outside Europe are for Charleston in the USA, beginning in 1738.

Figures 6, 7 show global maps of the available records sorted by start year and length.

Our collection includes 2359 temperature records, 3134 precipitation records, 160 air pressure records and 1551 number of wet days records that contain more than 100 years of data. For temperature 117 records contain more than 200 years and 5 records more than 300 years. The five longest records are listed in Table 11.

Outside Europe, the Boston record is the 11th longest for temperature with 148 observation years before 1891. Westmoreland at Jamaica, situated in the Caribbean Sea in Central America has the longest time series for precipitation record with 131 years of observations before 1891, having started in 1760, and is the 33rd longest in HCLIM. Adelaide in Australia has the longest record outside Europe for number of rain days, parameter with 103 years before 1891. Precipitation had better coverage in Africa and Australia, pressure is even more limited to Europe.

Figure 8 provides an overview of the compilation of all records in HCLIM until 2021. Figure 8a shows the start years and the distribution per parameter, while Fig. 8b shows the length of the records. The maximum length of ~150 years is largely a product of not having considered series starting after 1890.

	Temperature	Air pressure	Precipitation	Number of rain days
Europe	Paris (1658)	Paris (1670)	Paris (1688)	Resterhufe-Osteel (1586)
Africa	Funchal (1747)	Funchal (1747)	Funchal (1747)	Freetown (1793)
Asia	Chandannagar (1741)	Beijing (1757)	Beijing (1757)	Beijing (1757)
North America	Charleston (1738)	Toronto (1734)	Cambridge (1742)	Charleston (1738)
Central- and South America	Westmoreland, Jamaica (1764)	Rio de Janeiro (1795)	Westmoreland, Jamaica (1760)	Mexico City (1775)
Australia	Sydney (1788)	Sydney (1788)	Paramatta (1822)	Sydney (1788)

Table 10. First and oldest meteorological observations/measurements organized according to continents and parameters in HCLIM. The start year is given in parenthesis.

```

SEF      1.0.0
ID      Global-Datasets_Dove-5
Name    Berlin
Lat     52.5
Lon     13.4
Alt     NA
Source  PALAEO-RA
Link    NA
Vbl     ta
Stat    mean
Units   C
Meta    Observer=Brand Gronau. Kirch.Mädler | Obs.times=7 1/2, 12 1/2, 9 1/2. dann 8, 1,11 und tägl. Extr.
Year    Month Day Hour Minute Period Value Meta
1719   1      1      1      1      month 2.8 orig=2.25R
1719   2      1      1      1      month 1.1 orig=0.90R
1719   3      1      1      1      month 5.2 orig=4.15R
1719   4      1      1      1      month 9 orig=7.20R
1719   5      1      1      1      month 15.1 orig=12.08R
1719   6      1      1      1      month 19 orig=15.20R
1719   7      1      1      1      month 21.4 orig=17.15R
1719   8      1      1      1      month 18.8 orig=15.00R
1719   9      1      1      1      month 13.9 orig=11.14R
1719  10      1      1      1      month 9 orig=7.19R
1719  11      1      1      1      month 6.6 orig=5.24R
1719  12      1      1      1      month 0.3 orig=0.22R

```

Fig. 5 Example of a SEF file: Temperature record from Berlin 1719, from the Dove Collection.

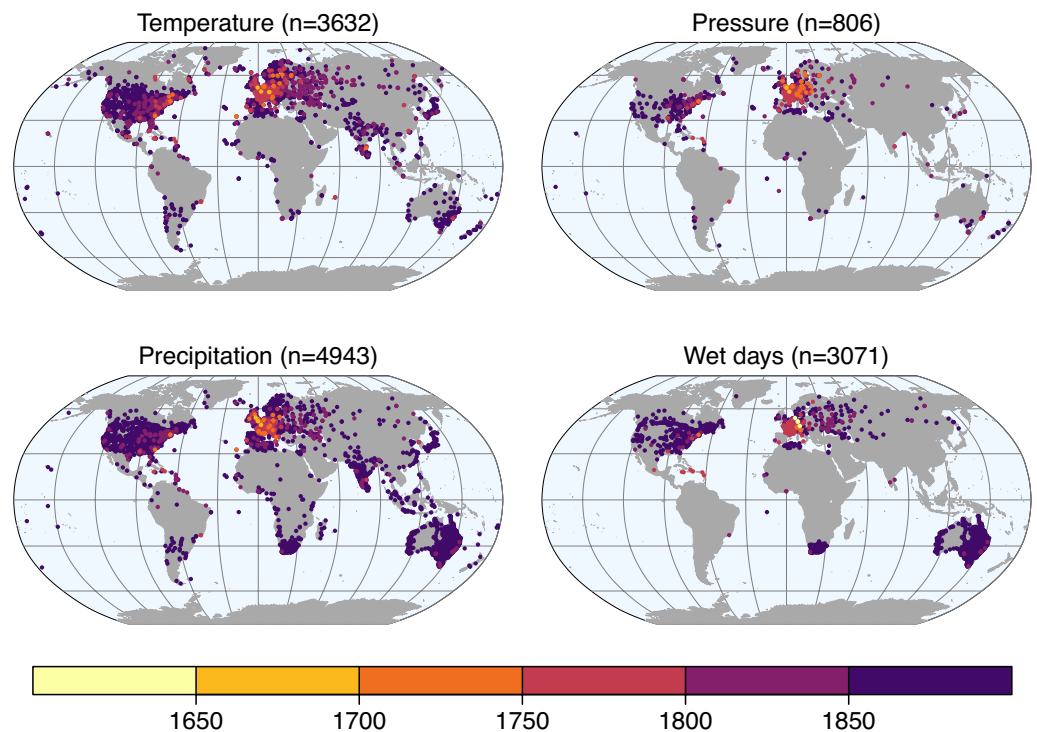


Fig. 6 Temperature, pressure, precipitation, and wet day records in HCLIM by start year.

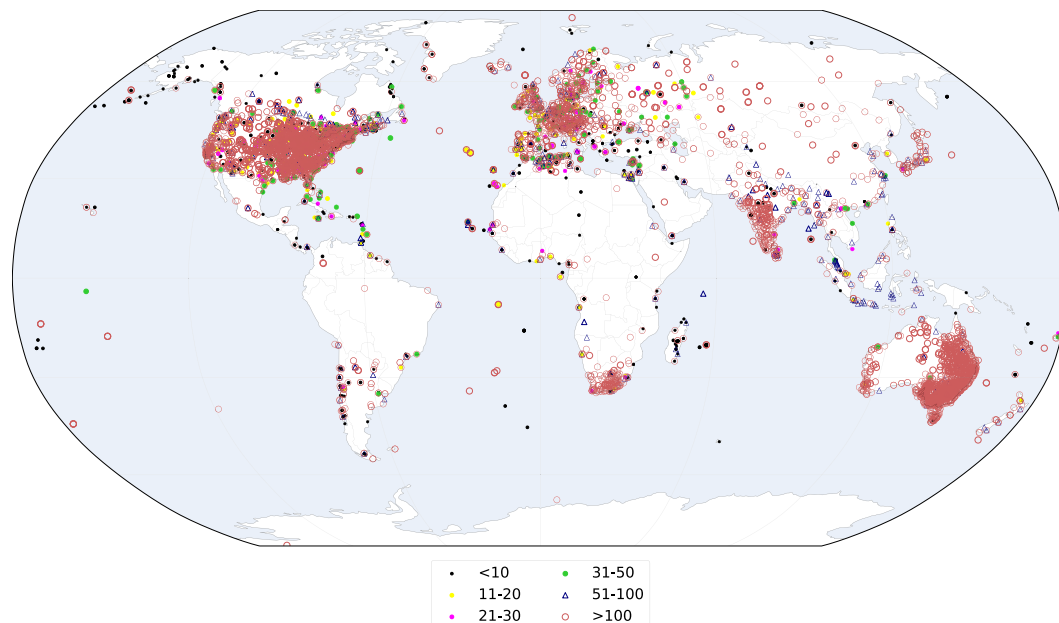


Fig. 7 Length of the time series in years before 1891.

Temperature (Ta)	Precipitation (RR)	Air Pressure (P/MSLP)	Wet Days (NR)
Paris (233)	Paris (203)	Paris (221)	Padova (291)
CET, London (232)	Kew, London (194)	London (199)	Hoheissenberg (238)
Berlin (190)	Zürich (183)	Zürich (183)	Palermo (211)
DeBilt (185)	Padova (166)	Vienna (182)	Praha-Klementinum (201)
Padova (165)	PodeHole-3, Spalding (165)	Prešov (174)	Bologna (194)

Table 11. Overview of the five longest weather records per parameter before 1891. The length in years is given in parenthesis, the operating years is calculated into years affected (missing months may occur).

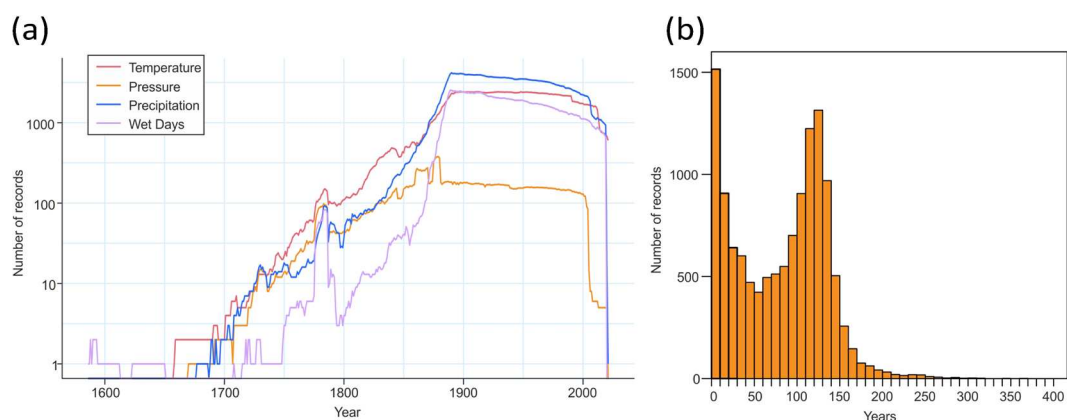


Fig. 8 (a) Time evolution of the number of records per parameter (note the logarithmic scale). (b) Histogram of the length of all the records until 2021.

As seen in Fig. 8a, the typical trend is that most records began in the late 19th century round year 1880–90 (pressure is little earlier) and it increases and ends up with an explosive development towards the end of the seventeenth century for all parameters.

However, the largest increase in the number of meteorological records occurred in the mid and late 19th century. This development does not apply to air pressure data, mostly because not as many data rescue projects have targeted this variable.

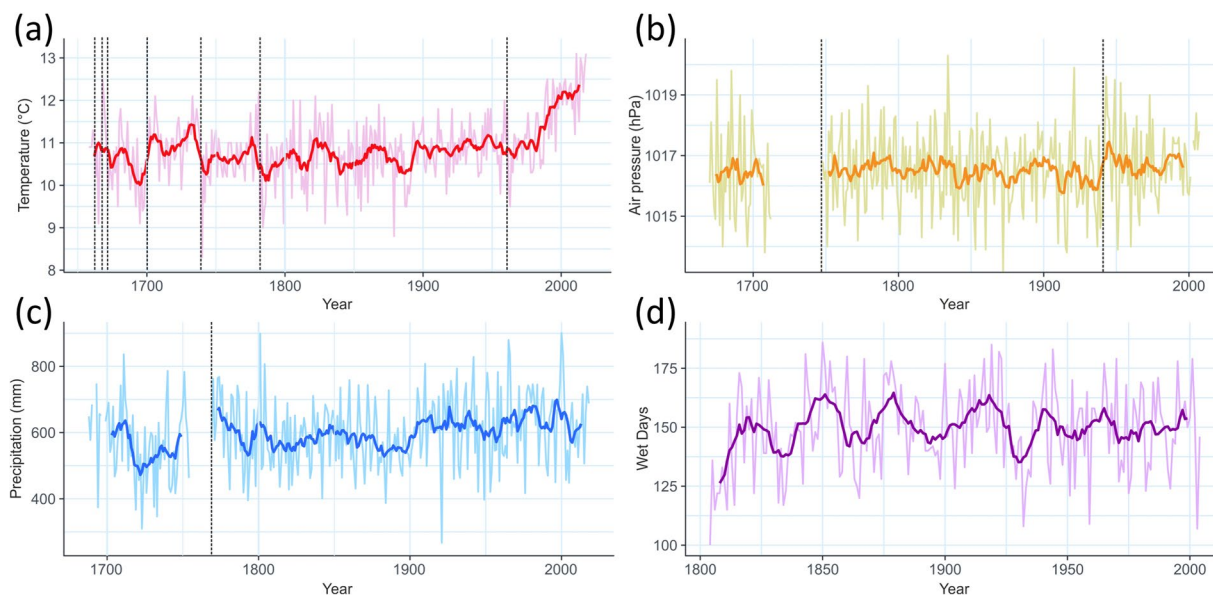


Fig. 9 (a) The oldest and longest temperature time series in the world from Paris that shows mean annual temperature for the years 1658–2018. The black dashed lines indicate homogeneity breaks (HCLIM)¹⁵. (b) The oldest and longest air pressure time series in the world from Paris, showing mean annual air pressure in hPa for the years 1670–2007 (the time series has a gap between 1726–1747). (c) The oldest and longest precipitation annual record from Paris (1688–2018). (d) The oldest and longest annual record for number of rain days from Prague^{88–103}.

Examples. The era of the first recorded meteorological observations begins during the 17th century in Europe. Here we show a few examples of the earliest and longest time series for each parameter.

Paris has the earliest and longest meteorological records lasting over 200 years for three meteorological parameters (temperature, pressure, and precipitation). Rousseau^{34,35} developed a monthly temperature record available back to 1658. This is the longest continuous instrumental meteorological record (duration: 360 years). Figure 9a shows the annual temperature time series from Paris, with breakpoints marked. Annual averages are calculated as the average or the sum, it depends on the parameter (WMO³⁶), and if there are missing values, the additional uncertainty introduced in the estimation of an average monthly value described in WMO²⁹, for example, is also taken into account.

After Evangelista Torricelli's invention of the barometer in 1643, systematic measurements of pressure began in 1670, also in Paris. From this, the Paris MSLP (Mean Sea-Level Pressure) record was compiled and published by Cornes *et al.*³⁷. Unfortunately, a gap in the series still exists for the period 1726–1747, for which it seems that no barometer observations have survived, however, there is no clear inhomogeneity in the longer dataset. The annual average MSLP for Paris (1670–2007) is presented in Fig. 9b.

The Paris precipitation record (Fig. 9c) begins in 1688 and is the longest known continuous precipitation record³⁸. There is a long gap in the record (18 years after 1755), to which a breakpoint is assigned by definition.

Prague has the longest time series (Fig. 9d) for rain days without any gap.

Technical Validation

As described in the Methods section, we provide raw data. Result from the QC control provides an indication of the general data quality (Table 12). Most QC flags are for precipitation. The well-known global databases have few QC flags (Table 12). This became clear when every single database that was downloaded was quality controlled. These are probably previously well controlled. It should also be mentioned that until the end of the 19th century there was no standard regulation for meteorological observations^{16,39}. However, to be precise the first international standards were set in 1873, although it was mostly about observation times and reporting standards⁴⁰.

For these reasons, testing the raw data and all the available metadata is the best option with which to optimize the use of the database. Every individual user will be able to apply the type of post-processing that is best suited to their needs.

After applying our breakpoint detection algorithm, we find large inhomogeneities in ca. 76% of the temperature records, 41% of the pressure records, 48% of the precipitation records, and 77% of the wet days' records (after de-duplication). This corresponds to an average homogeneous period (i.e., number of station years divided by the number of breakpoints) of 33, 56, 129, and 37 years, respectively.

We stress that smaller inhomogeneities remain undetected and that the detection is less effective for precipitation series, where the signal-to-noise ratio is very low. This can be relevant for applications such as trend analysis and would require more advanced detection methods that make use of reference series.

Source	Period	Variables	Fraction of flags after quality control	
			1658–1890 (232 years)	1891–2021 (130 years)
HCLIM-Dataset	1658–2019	Ta	0.0066%	0.022%
		P	0.27%	0.015%
		RR	0.4%	0.41%
		NR	0.098%	0.12%

Table 12. Overview of the percentage of flagged data for the quality control in HCLIM.

Usage Notes

The data products can be widely used in climate change research, such as reconstructions and data assimilations. Our database is based on an equivalent methodology that was previously developed by many others (GHCN, ISTI, Berkeley Earth etc). But this product represents the most comprehensive pre-industrial global dataset at a monthly temporal resolution. The data have been quality controlled and duplicate dataset removed. Although a breakpoint detection has been performed, complete homogenization is still required. Hence, we do not recommend using the dataset for trend analyses at the current stage, but the utility of the database is equally valuable, for analysis of singular extreme events, or the impact of volcanic eruptions (Laki in 1783, Tambora in 1815 and the Year Without a Summer, etc).

Code availability

R code used for formatting, quality control, removing duplicates, and breakpoint detection are publicly available under <https://github.com/elinlun/Hclim>.

The data are available at PANGAEA¹⁵.

Received: 14 June 2022; Accepted: 19 December 2022;

Published online: 19 January 2023

References

- Brázdil, R. *et al.* Extreme droughts and human responses to them: the Czech Lands in the pre-instrumental period. *Clim. Past* **15**, 1–24, <https://doi.org/10.5194/cp-15-1-2019> (2019).
- Cornes, R. C., Jones, P. D., Briffa, K. R. & Osborn, T. J. Estimates of the North Atlantic Oscillation back to 1692 using a Paris–London westerly index. *International Journal of Climatology* **33**(1), 228–248, <https://doi.org/10.1002/joc.3416> (2013).
- Timmreck, C. *et al.* The unidentified volcanic eruption of 1809: why it remains a climatic cold case. *Clim. Past* **17**(4), 1455–1482, <https://doi.org/10.5194/cp-17-1455-2021> (2021).
- Küttel, M., Luterbacher, J. & Wanner, H. Multidecadal changes in winter circulation–climate relationship in Europe: frequency variations, within-type modifications, and long-term trends. *Clim. Dyn.* **36**, 957–972, <https://doi.org/10.1007/s00382-009-0737-y> (2011).
- Valler, V., Franke, J., Brugnara, Y. & Brönnimann, S. An updated global atmospheric paleo-reanalysis covering the last 400 years. *Geoscience Data journal* 00:1–19, <https://doi.org/10.1002/gdj3.121> (2021).
- Slivinski, L. C. *et al.* Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. Royal Meteorological Society, Volume **145**, Issue 724 <https://doi.org/10.1002/qj.3598> (2019).
- Brunet, M. & Jones, P. D. Data rescue initiatives: Bringing historical climate data into the 21st century. *Climate Research* **47**(1), 29–40, <https://doi.org/10.3354/cr00960> (2011).
- Hawkins, E. *et al.* Estimating Changes in Global Temperature since the Preindustrial Period. *American Meteorological Society* **98**(9), 1841–1856, <https://doi.org/10.1175/BAMS-D-16-0007.1> (2017).
- Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., Lawrimore, J. H. The global historical climatology network monthly Temperature dataset, Version 4. *J. Climate* **31** (24): 9835–9854 <https://www.jstor.org/stable/26661466> (2018).
- Rohde, R. *et al.* New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011, *Geoinfor Geostat: An Overview 1:1* <https://doi.org/10.4172/2327-4581.1000101> (2013).
- Nicholson, S. E., Dezfuli, A. K. & Klotter, D. A two-century precipitation dataset for the continent of Africa. *Bull. Amer. Meteorol. Soc* **93**, 1219–1231, <https://doi.org/10.1175/BAMS-D-11-00212.1> (2012).
- Przybylak, R., Wyszynski, P., Nordli, Ø. & Strzyzewski, T. Air temperature changes in Svalbard and the surrounding seas from 1865 to 1920. *Int. J. Climatol.* **36**, 2899–2916, <https://doi.org/10.1002/joc.4527> (2016).
- Ashcroft, L., Gergis, J. & Karoly, D. J. A historical climate dataset for southeastern Australia, 1788–1859. *Geosci. Data J.* **1**(2), 158–178, <https://doi.org/10.1002/gdj3.19> (2014).
- Brönnimann, S. *et al.* Unlocking pre-1850 Instrumental Meteorological records – A global Inventory. *Bulletin of the American Meteorological Society* **100**, 389–413, <https://doi.org/10.1175/BAMS-D-19-0040.1> (2019).
- Lundstad, E., Brugnara, Y. & Brönnimann, S. Global Early Instrumental Monthly Meteorological Multivariable Database (HCLIM). PANGAEA <https://doi.pangaea.de/10.1594/PANGAEA.940724> (2022).
- Atmospheric Circulation Reconstructions over the Earth (ACRE): <https://www.met-acre.net/>
- Dove, H. W. *Temperaturtafeln nebst Bemerkungen, über die Verbreitung der Temperatur auf der Oberfläche der Erde und ihre jährlichen periodischen Schwankungen*. Erste, Zweite, Dritte und Vierte Abhandlung. Berlin (1838, 1839, 1842).
- La Société royale de médecine (Royale de Médecine or Royal Society of Medicine) in Paris (1776–1789).
- Societas Meteorologica Palatina, Ephemerides Societatis Meteorologicae Palatinae: Observations anni 1781–92, Vol. **12**, edited by: Hemmer, J. J., Mannheim (1783–95).
- Brugnara, Y. Swiss Early Meteorological Observations (CHIMES). PANGAEA <https://doi.pangaea.de/10.1594/PANGAEA.909141> (2020).
- Brugnara, Y. *et al.* A collection of subdaily pressure and temperature 3 observations for the early instrumental period with a focus on the “year without a summer” 4 1816. *Clim. Past* **11**, 1027–1047, <https://doi.org/10.5194/cp-11-1027-2015> (2015).
- Middleton, W. E. K. *Invention of the Meteorological Instruments*. The Johns Hopkins Press: Baltimore (1969).
- World Meteorological Organization (WMO). Guidelines on Best Practices for Climate Data Rescue. WMO-No. **1182** https://library.wmo.int/doc_num.php?explnum_id=3318 (2016).

24. Brugnara, Y. Global early instrumental climate data digitized at the University of Bern. Bern Open Repository and Information System. 10.48620/7 (2021).
25. Global Land and Marine Observations Dataset (GLAMOD): <https://climate.copernicus.eu/global-land-and-marine-observations-database>
26. Noone, S. *et al.* Progress towards a holistic land and marine surface meteorological database and a call for additional contributions. *Geosci. Data J.* **8**, 103–120, <https://doi.org/10.1002/gdj3.109> (2021).
27. Copernicus Climate Change Service data store: <https://cds.climate.copernicus.eu#!/home>
28. Muñoz Sabater, J. *et al.* ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **13**, 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021> (2021).
29. World Meteorological Organization (WMO). Guidelines on the Calculation of Climate Normals. WMO-No. 1203 https://library.wmo.int/doc_num.php?explnum_id=4166 (2017).
30. Brunet, M. *et al.* Best Practice Guidelines for Climate Data and Metadata Formatting, Quality Control and Submission. Copernicus Climate Change Service <https://doi.org/10.24381/kctk-8j22> (2020).
31. Station Exchange Format (SEF) <https://datarescue.climate.copernicus.eu/>
32. Historical Instrumental Climatological Surface Time Series of the Greater Alpine Region: <http://www.zamg.ac.at/histalp/index.php>
33. Alexandersson, H. & Moberg, A. Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *International Journal of Climatology* **17**(1), 25–34, [https://doi.org/10.1002/\(SICI\)1097-0088\(199701\)17:1%3C25::AID-JOC103%3E3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0088(199701)17:1%3C25::AID-JOC103%3E3.0.CO;2-J) (1997).
34. Rousseau, D. Les moyennes mensuelles de températures à Paris de 1658 à 1675, d'Ismaël Boulliau (1658-1660) à Louis Morin (1665–1675). *La Météorologie, 8e série* **81**, 11–22 (2013).
35. Rousseau, D. Les températures mensuelles en région parisienne de 1676 à 2008. *La Météorologie, 8e série* **67**, 43–55 (2009).
36. World Meteorological Organization (WMO). Guide to Climatological Practices WMO-No. 100 https://library.wmo.int/doc_num.php?explnum_id=5541 (2018).
37. Cornes, R. C., Jones, P. D., Briffa, K. R. & Osborn, T. J. A daily series of mean sea-level pressure for Paris, 1670–2007. *International Journal of Climatology* **32**(8), 1135–1150, <https://doi.org/10.1002/joc.2349> (2012).
38. Slonosky, V. C. Wet winters, dry summers? Three centuries of precipitation data from Paris. *Geophysical Research letters* **29**(No.19), 1895, <https://doi.org/10.1029/2001GL014302> (2002).
39. Kingston, G. T. *Meteorological Service, Dominion of Canada: Instruction to Observers*. Toronto (1878).
40. The Meteorological congress at Vienna (1873). *Nature publishing group* <https://www.nature.com/articles/010017a0> (1874).
41. Stockholm Historical Weather Observations (Moberg, A.) <https://bolin.su.se/data/stockholm>
42. Allan, R. *et al.* The International Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative. *Bulletin of the American Meteorological Society* **92**(11), 1421–1425, <http://www.jstor.org/stable/26218599> (2011).
43. Atmospheric Circulation Reconstructions over the Earth (Allen, R.) <http://www.met-acre.org/Home>
44. Zaiki, M. *et al.* Recovery of 19th century Tokyo/Osaka meteorological data in Japan. *Int. J. Climatol.* **26**, 399–423, <https://doi.org/10.1002/joc.1253> (2006).
45. Deutscher Wetterdienst: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/
46. Berkeley Earth data: <https://climatedataguide.ucar.edu/climate-data/global-surface-temperatures-best-berkeley-earth-surface-temperatures>
47. Jones, P. D., Parker, D.E., Osborn, T.J., Briffa, K.R. *Global and Hemispheric Temperature Anomalies – Land and Marine Instrumental Records*. CRU & Hadley Centre <https://www.osti.gov/dataexplorer/biblio/dataset/1389299> (2006).
48. Climatic Research Unit: <https://crudata.uea.ac.uk/cru/data/temperature/>
49. Global Historical Climatology Network (GHCN): <https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-monthly-version-4>
50. Franke, J., Brönnimann, S., Bhend, J. & Brugnara, Y. A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations. *Sci Data* **4**, 170076, <https://doi.org/10.1038/sdata.2017.76> (2017).
51. Compo, G. P. *et al.* *The International Surface Pressure Databank version 4*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory <https://doi.org/10.5065/9EYR-TY90> (2019).
52. The International Surface Pressure Databank (ISPD): <https://rda.ucar.edu/datasets/ds132.2/>
53. Rennie, J. J. *et al.* The International surface temperature initiative global land surface databank: monthly temperature data release description and methods. *Geosci. Data J.* **1**, 75–102, <https://doi.org/10.1002/gdj3.8> (2014).
54. ISTI data: <http://www.surfacetemperatures.org/>
55. Slonosky, V. C. Hazardous weather events in the St. Lawrence Valley from the French regime to Confederation: descriptive weather in historical records from Quebec City and Montreal, 1742–1869 and 1953—present. *Nat Hazards* **98**, 51–77, <https://doi.org/10.1007/s11069-019-03612-5> (2019).
56. Slonosky, V. C. Historical climate observations in Canada: 18th and 19th century daily temperature from the St. Lawrence Valley, Quebec. *Geosci. Data J.* **1**, 103–120, <https://doi.org/10.1002/gdj3.11> (2014).
57. Open Data Rescue (Canada): <https://opendatarescue.org>
58. Brugnara, Y. *et al.* Early instrumental meteorological observations in Switzerland: 1708–1873. *Earth Syst. Sci. Data* **12**, 1179–1190, <https://doi.org/10.5194/essd-12-1179-2020> (2020).
59. Kaspar, F., Tinz, B., Mächel, H. & Gates, L. Data rescue of national and international meteorological observations at Deutscher Wetterdienst. *Adv. Sci. Res.* **12**, 57–61, <https://doi.org/10.5194/asr-12-57-2015> (2015).
60. DWD – overseas data: https://www.dwd.de/EN/ourservices/overseas_stations/overseas_stations.html
61. Klein Tank, A. M. G. *et al.* Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. Climatol.* **22**, 1441–1453, <https://doi.org/10.1002/joc.773> (2002).
62. European Climate Assessment & Dataset: <https://www.ecad.eu/>
63. Domínguez-Castro, F. *et al.* Early meteorological records from Latin America and the Caribbean during the 18th and 19th centuries. *Sci. Data* **4**, 170169, <https://doi.org/10.1038/sdata.2017.169> (2017).
64. Camuffo, D. & Jones, P. D. Improved Understanding of Past Climatic Variability from Early Daily European Instrumental Sources. *Climatic Change* **53**(1), 1–4, <https://doi.org/10.1023/A:1014902904197> (2002).
65. Camuffo, D. & Jones, P. D. *Improved Understanding of Past Climatic Variability from Early Daily European Instrumental Sources*. Springer <https://link.springer.com/book/10.1007/978-94-010-0371-1#editorsandaffiliations> (2002).
66. Auer, I. *et al.* HISTALP – historical instrumental climatological surface time 13 series of the Greater Alpine Region. *Int. J. Climatol.* **27**, 17–46, <https://doi.org/10.1002/joc.1377> (2007).
67. Japan-Asia Climate Data Program: <https://jcdp.jp/instrumental-meteorological-data/>
68. Brunet, M., Gilbert, A., Jones, P. D. & Efthymiadis, D. A historical surface climate dataset from station observations in Mediterranean North Africa and Middle East areas. *Geoscience Data Journal* **1**, 121–128, <https://doi.org/10.1002/gdj3.12> (2014).
69. C3-EURO4M-MEDARE Mediterranean historical climate data (MEDARE): <https://doi.org/10.5281/zenodo.7531>
70. Przybylak, R., Vizi, Z. & Wyszynski, P. Air temperature changes in the Arctic from 1801 to 1920. *Int. J. Climatol.* **30**, 791–812, <https://doi.org/10.1002/joc.1918> (2010).
71. HARD 2.0 – Historical Arctic Database: <http://www.hardv2.prac.umk.pl/>

72. Ashcroft, L. *et al.* A rescued dataset of subdaily meteorological observations for Europe and the southern Mediterranean region, 1877–2012. *Earth Syst. Sci. Data* **10**, 1613–1635, <https://doi.org/10.5194/essd-10-1613-2018> (2018).
73. SEA – Historical Climate Data (Australia): <https://lindenashcroft.com/research/>
74. Klein Tank, A. M. G. *et al.* Changes in daily temperature & precipitation extremes in central and south Asia. *J. Geophys. Res.* **111**, D16105, <https://doi.org/10.1029/2005JD006316> (2006).
75. Southeast Asian Climate Assessment & Dataset (SACA&D): <https://www.climateurope.eu/southeast-asian-climate-assessment-dataset-sacad/>
76. Precipitation data from Africa: <https://doi.org/10.1175/BAMS-D-11-00212.1>
77. Westcott, N. E. *et al.* *Quality Control of 19th Century Weather Data*. Midwestern Regional Climate Center, Illinois State Water Survey, Illinois (2011).
78. Westcott, N. E., Cooper, J., Andsager, K., Stoecker, L. A. & Shein, K. Status and Climate Applications of the 19th Century Forts and Volunteer Observer Database. *J. Applied and Service Climatology*, Vol. **2021**(2) <https://doi.org/10.46275/JOASC.2021.09.001> (2021).
79. Forts and volunteer Observer Database: https://mrcc.purdue.edu/data_serv/cdmp/cdmp.jsp
80. Koninklijk Nederlands Meteorologisch Instituut: http://projects.knmi.nl/klimatologie/daggegevens/antieke_wrn/index.html
81. MétéoFrance: <https://meteofrance.com/comprendre-climat/etudier-climat-passe>
82. MétéoFrance: <http://archives-climat.fr/>
83. Norsk Meteorologisk Institutt: <https://seklima.met.no/>
84. Roshydromet Всероссийский научно-исследовательский институт: http://meteo.ru/english/climate/cl_data.php
85. Sveriges meteorologiska och hydrologiska institut: <https://www.smhi.se/data/meteorologi/temperatur>
86. Toaldo, G. *La Meteorologia applicata all'Agricoltura*, Storti, Venice (1775).
87. Pappert, D. Unlocking weather observations from the Societas Meteorologica Palatina (1781–1792). *Clim. Past* **17**, 2361–2379, <https://doi.org/10.5194/cp-2021-57> (2021).
88. Alcoforado, M. J., Vaquero, J. M., Trigo, R. M. & Taborada, J. P. Early Portuguese meteorological measurements (18th century). *Clim. Past* **8**, 353–371, <https://doi.org/10.5194/cp-8-353-2012> (2012).
89. Brázdil, R., Kiss, A., Luterbacher, J. & Valášek, H. Weather patterns in eastern Slovakia 1717–1730, based on records from the Breslau meteorological network. *International Journal of Climatology* **28**(12), 1639–1651, <https://doi.org/10.1002/joc.1667> (2008).
90. Kunz, M., Kottmeier, C., Lähne, W., Bertram, I. & Ehmann, C. The Karlsruhe climate time series since 1779. *Meteorologische Zeitschrift* **31**(No. 3), 175–202, <https://doi.org/10.1127/metz/2022/1106> (2022).
91. Camuffo, D., Valle, A. D., Becherini, F. & Zanini, V. Three centuries of daily precipitation in Padua, Italy, 1713–2018: history, relocations, gaps, homogeneity, and raw data. *Climatic Change* **162-4**, 923–942, <https://doi.org/10.1007/s10584-017-1931-2> (2020).
92. Camuffo, D., Becherini, F. & Valle, A. D. Temperature observations in Florence, Italy, after the end of the Medici Network (1654–1670): the Grifoni record (1751–1766). *Climatic Change* **162**, 943–963, <https://doi.org/10.1007/s10584-020-02760-z> (2020).
93. Camuffo, D., Valle, A. D., Bertolin, C. & Santorelli, E. Temperature observations in Bologna, Italy, from 1715 to 1815: a comparison with other contemporary series and an overview of three centuries of changing climate. *Climatic Change* **142**, 7–22, <https://doi.org/10.1007/s10584-017-1931-2> (2017).
94. Cornes, R. C. Robert Boyle's weather journal for the year 1685. *Weather* **75**, 272–277, <https://doi.org/10.1002/wea.3531> (2020).
95. Demarée, G. R. & Ogilvie, A. E. J. Climate-related Information in Labrador/Nunatsiavut: Evidence from Moravian Missionary Journals. *Meded. Zitt. K. Acad. Overzeese Wet.* **57**(2-4), 391–408 (2011).
96. Domínguez-Castro, F. *et al.* Early Spanish meteorological records (1780–1850). *Int. J. Climatol.* **34**(3), 593–603, <https://doi.org/10.1002/joc.3709> (2014).
97. Filipiak, J. & Miętus, M. History of the Gdańsk Pre-Instrumental and Instrumental Record of Meteorological Observations and Analysis of Selected Air Pressure Observations. In: Przybylak R., Majorowicz, J., Brázdil, R., Kejna M. (eds) *The Polish Climate in the European Context: An Historical Overview*, Springer, Berlin Heidelberg New York https://doi.org/10.1007/978-90-481-3167-9_12 (2010).
98. Rodrigo, F. S. Early meteorological data in southern Spain during the Dalton Minimum. *Int. J. Climatol.* **39**, 3593–3607, <https://doi.org/10.1002/joc.6041> (2019).
99. EMOWA data <http://repositorio.ual.es/handle/10835/10636> (Rodrigo, 2021).
100. Slonosky, V. C. Daily Minimum and Maximum Temperature in the St-Lawrence Valley, Quebec: Two Centuries of Climatic Observations from Canada. *Int. J. Climatol.* **35**(7), 1662–81, <https://doi.org/10.1002/joc.4085> (2015).
101. Slonosky, V. C. The Meteorological Observations of Jean-François Gaultier, Quebec, Canada: 1742–56. *J. Clim.*, **16**, 2232–2247 [10.1175/1520-0442\(2003\)16<2232:TMOOJG>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)16<2232:TMOOJG>2.0.CO;2) (2003).
102. Thomas Jefferson's weather journal: <https://jefferson-weather-records.org/>
103. Ives, G. A history of the monsoon in southern India between 1730 and 1920 and its impact on society https://etheses.whiterose.ac.uk/27995/1/Gemma_Ives_Corrected_Thesis.pdf (2020).

Acknowledgements

The work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No 787574 (PALAEO-RA), by the Swiss National Science Foundation (project WeaR 188701), by Copernicus Climate Change Service (C3S) 311a Lot 1, by the Federal Office of Meteorology and Climatology MeteoSwiss in the framework of GCOS Switzerland (project “Long Swiss Meteorological Series”). The work of R.P. was supported by the NCN projects No. DEC 2020/37/B/ST10/00710 and No. 2020/39/B/ST10/00653, and by funds from the Nicolaus Copernicus University–Emerging field: Global Environmental Changes. We thank all digitizers for their help in rescuing the meteorological data. Thank you for the collaboration and the data sharing from Anders Moberg⁴¹, Rob Allen^{42,43}, Masumi Zaiki⁴⁴, and Lisa Hannak⁴⁵. EL would like to thank colleagues at GiUB (Geographisches Institut, Universität Bern) and colleagues at the Norwegian Meteorological Institute (Anita Verpe Dyrørdal, Oscar Landgren, Line Båserud and Reidun Gangstø Skaland) for help with formatting and quality control advice. Thanks also to the Norwegian Meteorological Institute for loaning an office during the COVID-19 pandemic, when E.L. mostly spent time in Norway given the reduced outbreak and fewer restrictions in that country. The reconstructions and all analyses presented in this data descriptor have been performed based on free and open-source software (R & Python).

Author contributions

E.L. processed the data and undertook quality control of the data, as also produced the figures; Y.B. was responsible for the data rescue initiative and provided figures for this section, a review of the QC and the breakpoint detection section of the manuscript; S.B. conceptualized the project and its intellectual configuration; Y.B. and S.B. provided methodological advice and provided code, especially for the removing duplicates script.

E.L. developed the first draft of the manuscript. E.L., Y.B., S.B. contributed to editing and approved the final version of the manuscript. D.P., J.K., E.S., A.H. have helped with data editing. A.A., B.C., R.C., G.D., J.F., L.G., G.L.L., J.M.J., S.J., A.K., S.E.N., R.P., P.J., D.R., B.T., F.S.R., S.G., F.D-C., M.B., V.S., J.C. and M.B. contributed data. All the authors contributed to reviewing and editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01919-w>.

Correspondence and requests for materials should be addressed to E.L.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2023