DOCU-CLIM: A global documentary climate dataset for climate reconstructions

Angela-Maria Burgdorf[1,2], Stefan Brönnimann[1,2] George Adamson[3], Tatsuya Amano[4], Yasuyuki Aono[5], David Barriopedro[6], Teresa Bullón[7], Chantal Camenisch[1,8], Dario Camuffo[9], Valérie Daux[10], María del Rosario Prieto[11,12,†], Petr Dobrovolný[13,14], David Gallego[15], Ricardo García-Herrera[16,17], Joelle Gergis[18,19], Stefan Grab[20], Matthew J. Hannaford[21], Jari Holopainen[22], Clare Kelso[23], Zoltán Kern[24,25], Andrea Kiss[26], Elaine Lin Kuan-Hui[27,28], Neil J. Loader[29], Martin Možný[30], David Nash[31,20], Sharon E. Nicholson[32], Christian Pfister[1,8], Fernando S. Rodrigo[33], This Rutishauser[2], Sapna Sharma[34], Katalin Takács[35], Ernesto T. Vargas[36] & Inmaculada Vega[15]

1 Institute of Geography, University of Bern, Switzerland

2 Oeschger Centre for Climate Change Research, University of Bern, Switzerland

3 Department of Geography, King's College London, London, UK

4 School of Biological Sciences, The University of Queensland, Brisbane, Australia

5 Graduate School of Agriculture, Osaka Metropolitan University, Osaka, Japan

6 Instituto de Geociencias (IGEO), CSIC-UCM, Madrid, Spain

7 Department of Geography, Autonomous University of Madrid (UAM), Madrid, Spain

8 Institute of History, Department of Economic, Social and Environmental History (WSU), University of Bern, Bern, Switzerland

9 National Research Council-Institute of Atmospheric Sciences and Climate, Corso Stati Uniti 4, Padua, Italy

10 Laboratoire des Sciences du Climat et de l'Environnement, CNRS, CEA, UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France

11 Argentine Institute of Nivology, Glaciology and Environmental Sciences (IANIGLA-CONICET), Mendoza, Argentina

12 Facultad de Filosofía y Letras, Universidad Nacional de Cuyo, Mendoza, Argentina

13 Department of Geography, Faculty of Science, Masaryk University, Brno, Czech Republic

14 Global Change Research Institute, Czech Academy of Sciences, Brno, Czech Republic

15 Departamento de Sistemas Físicos, Químicos y Naturales, Universidad Pablo de Olavide, Seville, Spain

16 Departamento de Física de la Tierra y Astrofísica, Universidad Complutense, Madrid, Spain

17 IGEO, Instituto de Geociencias (CSIC, UCM), Madrid, Spain

18 Fenner School of Environment and Society, Australian National University, Canberra, Australia

19 ARC Centre for Climate Extremes, The Australian National University, Canberra, Australia

20 School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa

21 Department of Geography, College of Science, University of Lincoln, Lincoln, UK

22 Natural Resources Institute Finland, Helsinki, Finland

23 Department of Geography, Environmental Management and Energy Studies, University of Johannesburg, Johannesburg, South Africa

24 Institute for Geological and Geochemical Research, Research Centre for Astronomy and Earth Sciences, Budapest, Hungary

25 CSFK, MTA Centre of Excellence, Konkoly Thege Miklós út 15-17, Budapest, Hungary

26 Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria

27 Research Center for Environmental Changes, Academia Sinica, Taipei, Taiwan

28 Graduate Institute of Environmental Education, National Taiwan Normal University, Taipei, Taiwan

29 Department of Geography, Swansea University, Swansea, UK

30 Department of Biometeorological Applications, Czech Hydrometeorological Institute, Prague, Czech Republic

31 School of Applied Sciences, University of Brighton, Brighton, United Kingdom

32 Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, Florida, USA

33 Department of Chemistry and Physics, University of Almería, Almería, Spain

34 Department of Biology, York University, Toronto, Ontario, Canada

35 Institute for Soil Sciences (TAKI) and Centre for Agricultural Research (ATK), Budapest, Hungary

36 National Museum of Natural Sciences-Spanish Research Council, Madrid, Spain

Corresponding author: Stefan Brönnimann (stefan.broennimann@giub.unibe.ch)

Abstract

Documentary climate data describe evidence of past climate arising from predominantly written historical documents such as diaries, chronicles, newspapers, or logbooks. Over the past decades, historians and climatologists have generated numerous document-based time series of local and regional climates. However, a global dataset of documentary climate time series has never been compiled, and documentary data are rarely used in large-scale climate reconstructions. Here, we present the first global multi-variable collection of documentary climate records. The dataset DOCU-CLIM comprises 621 time series (both published and hitherto unpublished) providing information on historical variations in temperature, precipitation, and wind regime. The series are evaluated by formulating proxy forward models (i.e., predicting the documentary observations from climate fields) in an overlapping period. Results show strong correlations, particularly for the temperature-sensitive series. Correlations are somewhat lower for precipitation-sensitive series. Overall, we ascribe considerable potential to documentary records as climate data, especially in regions and seasons not well represented by early instrumental data and palaeoclimate proxies.

1. Background & Summary

Information on past climates has played an essential role in climate science[1]. While historically the primary research focus has been on reconstructing past annual temperature, the questions raised nowadays with the help of palaeoclimatological data are multifaceted, including changes in the water cycle, the occurrence of weather and climate extreme events, and atmospheric dynamics. This, in turn, is a challenge for producing palaeoclimatic datasets. New approaches, such as off-line palaeodata assimilation[2–5], provide past climate fields at increasing spatial and temporal resolution. However, all reconstructions essentially depend on sufficient high-quality data inputs.

Current climate field reconstructions are largely based on relatively high-resolution proxies measured in natural archives such as tree rings, corals, speleothems, bivalves, sediments, or ice cores. Extensive compilations of such proxies exist [6,7]. In particular, tree rings are widely used, among others, due to their extensive spatial distribution across the globe. Unlike most other natural proxies, tree ring proxies, have an annual resolution. However, their climate signal is mainly limited to the growing season (although there are also winter reconstructions[8]). Documentary proxies, i.e., climate data originating from historical documents, could provide an essential contribution since they potentially cover combinations of seasons and regions (e.g., winter in East Asia) that are otherwise not well represented by natural proxies. Furthermore, documentary data are often calendar dated, and some have a very high temporal resolution. Despite these advantages, they are largely overlooked and only marginally used in large-scale climate reconstructions, since they are not readily available in digital format in the main compilations used by climate scientists, or their quality is not well known (note that classification of events is often based on effects, which requires local context information). The PAGES 2k multiproxy database[9], for instance, only includes 15 documentary proxy series. However, historians have compiled documentary climate information in databases such as EURO-CLIMHIST for many years[10] (note that qualitative weather descriptions are also found in databases of early instrumental meteorological data, e.g., Rodrigo [11]).

In recent years, a major international effort has been done to promote the use of the archives of societies in climate reconstructions. The PAGES CRIAS working group (Climate Reconstruction and Impacts from the Archives of Societies) was founded in 2018 and is working towards that goal. The Palgrave Handbook of Climate History[12] provided a first global overview of documentary climate data organized in regional chapters. Based on this and many other sources, Burgdorf [13] recently inventoried documentary climate series from a literature and databases search. The inventory contains 688 entries; not all are publicly available, and some have not yet been digitized. Here, we publish a subset of the data inventoried in Burgdorf [13], termed DOCU-CLIM. The dataset contributes to a global monthly palaeoclimate reanalysis starting in 1420 and is based on assimilating monthly-to-seasonal proxies, documentary data and instrumental data into an ensemble of atmospheric model simulations using an offline Kalman filter approach similar to Valler et al.[5] For that reason, here we focus on series that provide information in the window 1400-1880 CE at monthly to annual resolution. In this paper we present the dataset (see Supplementary File DOCU-CLIM_Inventory.txt for an overview of all records) and evaluate its usefulness for climate reconstruction using proxy forward models (statistical models that predict the documentary series from climate data rather than vice versa).

2. Methods

2.1. Compilation and Data rescue

Over the past decades, climate historians and historical climatologists have produced numerous datasets in which documentary data have been translated into quantitative climate information. However, as their focus is commonly regional or local, these data are often not submitted to global data repositories such as the NOAA World Data Service for Palaeoclimatology database (https://www.ncei.noaa.gov/access/paleo-search/) but instead, published on project or personal websites, or, unfortunately still very often, not published at all. Even if documentary records are incorporated into databases, they may not always be organized in a manner suitable for climate scientists, particularly when working with time series. In this work, we focus exclusively on quantitative document-based time series data, representing a small, albeit underexploited, subset of the body of documentary climate data.

Figure 1 illustrates the general workflow followed in this project. The compilation of an inventory of documentary climate series was described in a previous paper[13], which lists 688 records. While the latter paper described the metadata, in this paper we compiled the actual data. As detailed in Burgdorf[13], the inventory was based on a search of 14 existing databases (Table 1), contributing about 25% of the entries in the inventory, as well as extensive literature research, contributing the rest. The catalogued data followed a set of criteria, some of which were dictated by our intended use. For instance, we only inventoried material overlapping the period 1400-1880 CE, with a minimum record length of 30 years of which 20 must be before 1880. These criteria were set as we used the data in a data assimilation project starting in 1420 CE and in which instrumental data (which become more frequent after 1880 CE) were also assimilated. The variables of interest were temperature, precipitation, and wind (e.g., onset of seasonal wind regime) and hence only records were compiled that provide information on one of these variables (note that some of the series also depend on further variables). As detailed in Burgdorf[13], the focus was predominantly on English-language literature that was accessible electronically and in which the authors state that the series contains information on one of the three variables (publications about the Mediterranean area and Central Europe in other languages exist and may contain additional series). Except for phenological data, we used only secondary material to ensure the inclusion of expert source interpretation. This includes derived indices, generally accepted to quantify descriptive and qualitative documentary data[14] or even reconstructed time series in physical units.

The next step was to compile the actual data. Not all inventoried series are available in electronic form, and some are subject to a restrictive data policy. We downloaded the series from 14 databases (Table 1) and contacted many authors directly in cases when a dataset was not available in a repository. However, we only compiled data series that are open access and allow us to redistribute the data under a CC-BY license.

Table 1. Overview of available global and national repositories and databases containing documentary evidence. N_{all} is the total number of series or databases available on the platform, and N_{docu} is the number of those series based on documentary evidence and available prior to 1880 CE (from Burgdorf [13]). Note that there may be overlap between the repositories.

Name of repository or database	Abbreviation	Region	N_{all}	N_{docu}	Reference	URL
PAGES2k Global 2,000 Year Multiproxy Database	PAGES2k	Global	692	15	Emile-Geay et al. [6]	https://www.ncei.noaa.gov/access/paleo-search/study/21171 (last access: 30 May 2022)
NOAA/World Data Service for Paleoclimatology archives	NOAA Paleo	Global	>10000	61		https://www.ncei.noaa.gov/access/paleo-search/ (last access: 30 May 2022)
Euro-Climhist	Euro-Climhist	Switzerland/ central Europe	65	27	Pfister et al. [10]	https://www.euroclimhist.unibe.ch/en/ (last access: 30 May 2022)
Tambora.org	Tambora.org	Germany	4	4	Riemann et al. [15]	http://www.tambora.org (last access: 30 May 2022)
National Snow and Ice Database: Global Lake and River Ice Phenology	NSIDC	Northern Hemisphere	865	39	Benson et al.[16], updated 2020	https://nsidc.org/data/g01377/versions/1 (last access: 30 May 2022)
Japan Climate Data Project	JCDP	Japan	14	3		https://jcdp.jp (last access: 30 May 2022)
Climatological Database for the World's Oceans	CLIWOC	Global	1624		García-Herrera et al. [17]	https://www.historicalclimatology.com/cliwoc.html (last access: 30 May 2022)
Institute for Ocean Technology Ice Database	Ice Data	Canada	4	4		http://www.icedata.ca (last access: 30 May 2022)
KNMI Climate Explorer	Climate Explorer	Global	>200	~10		https://climexp.knmi.nl/start.cgi?id=someone@somewhere (last access: 30 May 2022)
Red Española de Reconstrucción Climática a Partir de Fuentes Documentales	RECLIDO	Spain	7	7		http://stream- ucm.es/RECLIDO/es/home-es.htm (last access: 30 May 2022)
Salvá Sinobas	Salvá Sinobas	Iberian Peninsula	18	5		http://salva- sinobas.uvigo.es/index.php (last access: 30 May 2022)
Variabilidad y Reconstrucción del Clima	Vareclim	Global	5	5		https://www.upo.es/vareclim/index.php (last access: 30 May 2022)
Reconstructed East Asian Climate Historical Encoded Series	REACHES	China	1	1	Wang et al. [18]	https://www.ncdc.noaa.gov/paleo- search/study/ 23410 (last access: 30 May 2022)
Tracking Extremes of Meteorological Phenomena Experienced in Space and Time	TEMPEST	United Kingdom	5	5	Veale et al. [19]	https://www.nottingham.ac.uk/research/groups/weather-extremes/research/tempest-database.aspx (last access: 30 May 2022)

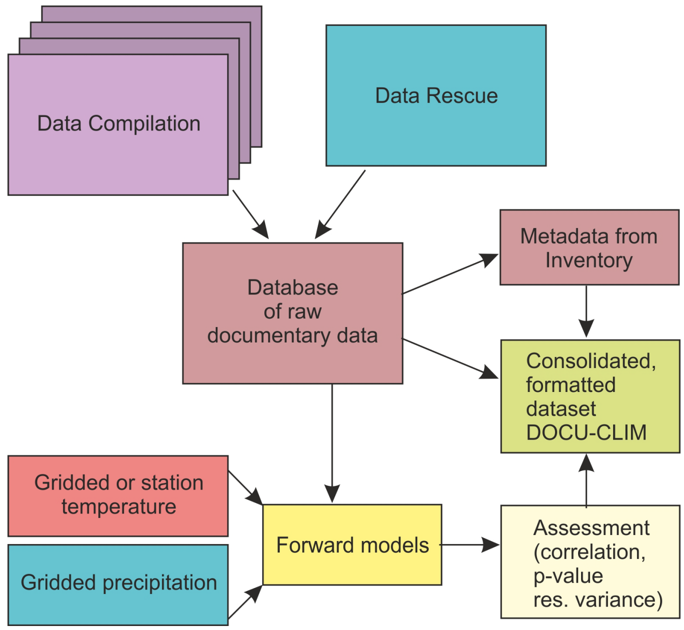


Figure 1. Flow chart depicting the generation of the documentary dataset.

In addition to compiling existing documentary data series, we also rescued a significant amount of data (this includes some series we recently presented in another study[20]). This concerns 137 ice phenology series and 5 precipitation series (Fig. 2). Note that some of the rescued data might be available electronically but we did not find it. The single most important source was a compilation of freezing and thawing dates of Russian rivers by Rykachev[21] (see example in Fig. 3) and a follow-up compilation by Shostakovich[22]. Some (few) series were measured from graphs published in the 1970s where the underlying data were unavailable electronically. Many of the datasets digitized in the 1970s and even 1980s have not made it into the era of electronic publishing and open data policies (a list of rescued series including the sources is given in Supplementary file DOCU-CLIM_Rescued.txt).

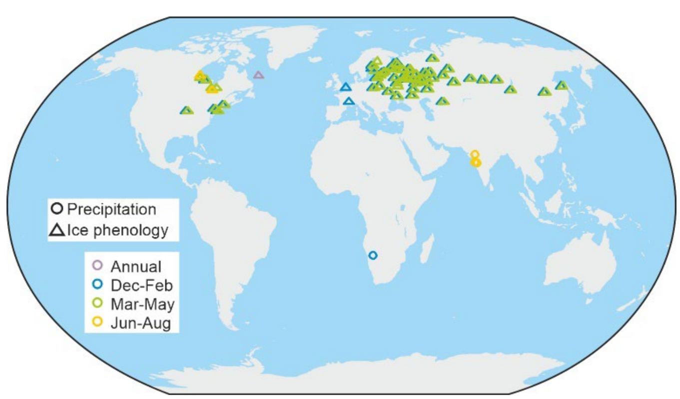
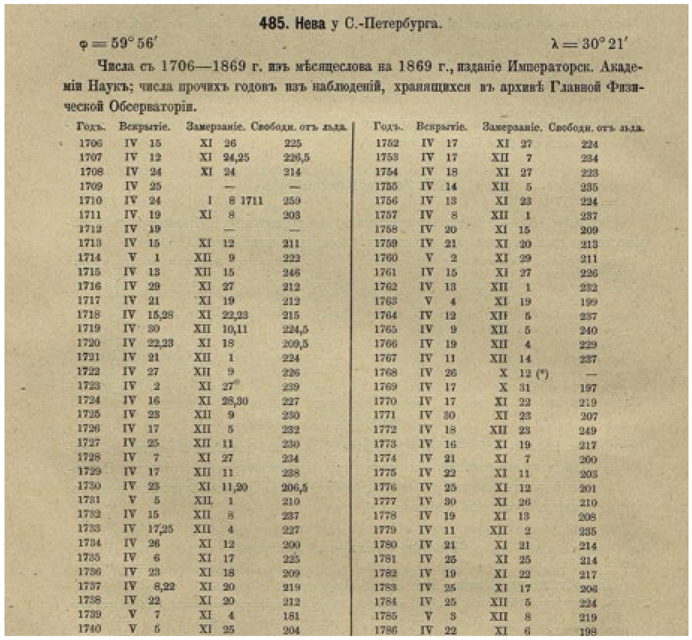


Figure 2. Map of rescued series.



485. Нева у С.-Петербурга.

φ = 59° 56′ λ = 30° 21′

Числа съ 1706—1869 г. изъ мѣсяцеслова на 1869 г., изданіе Императорск. Академіи Наукъ; числа прочихъ годовъ изъ наблюденій, хранящихся въ архивѣ Главной Физической Обсерваторіи.

Годъ.	Вскрытіе.	Замерзаніе.	Свободн. отъ льда.	Годъ.	Вскрытіе.	Замерзаніе.	Свободн. отъ льда.
1706	IV 15	XI 26	225	1752	IV 17	XI 27	224
1707	IV 12	XI 24,25	226,5	1753	IV 17	XII 7	234
1708	IV 24	XI 24	214	1754	IV 18	XI 27	223
1709	IV 25	—	—	1755	IV 14	XII 5	235
1710	IV 24	I 8 1711	259	1756	IV 13	XI 23	224
1711	IV 19	XI 8	203	1757	IV 8	XII 1	237
1712	IV 19	—	—	1758	IV 20	XI 15	209
1713	IV 15	XI 12	211	1759	IV 21	XI 20	213
1714	V 1	XII 9	222	1760	V 2	XI 29	211
1715	IV 13	XII 15	246	1761	IV 15	XI 27	226
1716	IV 29	XI 27	212	1762	IV 13	XII 1	232
1717	IV 21	XI 19	212	1763	V 4	XI 19	199
1718	IV 15,28	XI 22,23	215	1764	IV 12	XII 5	237
1719	IV 30	XII 10,11	224,5	1765	IV 9	XII 5	240
1720	IV 22,23	XI 18	209,5	1766	IV 19	XII 4	229
1721	IV 21	XII 1	224	1767	IV 11	XII 14	237
1722	IV 27	XII 9	226	1768	IV 26	X 12 (*)	—
1723	IV 2	XI 27	239	1769	IV 17	X 31	197
1724	IV 16	XI 28,30	227	1770	IV 17	XI 22	219
1725	IV 23	XII 9	230	1771	IV 30	XI 23	207
1726	IV 17	XII 5	232	1772	IV 18	XII 23	249
1727	IV 25	XII 11	230	1773	IV 16	XI 19	217
1728	IV 7	XI 27	234	1774	IV 21	XI 7	200
1729	IV 17	XII 11	238	1775	IV 22	XI 11	203
1730	IV 23	XI 11,20	206,5	1776	IV 25	XI 12	201
1731	V 5	XII 1	210	1777	IV 30	XI 26	210
1732	IV 15	XII 8	237	1778	IV 19	XI 13	208
1733	IV 17,25	XII 4	227	1779	IV 11	XII 2	235
1734	IV 26	XI 12	200	1780	IV 21	XI 21	214
1735	IV 6	XI 17	225	1781	IV 25	XI 25	214
1736	IV 23	XI 18	209	1782	IV 19	XI 22	217
1737	IV 8,22	XI 20	219	1783	IV 25	XI 17	206
1738	IV 22	XI 20	212	1784	IV 25	XII 5	224
1739	V 7	XI 4	181	1785	V 3	XII 8	219
1740	V 5	XI 25	204	1786	IV 22	XI 6	198

Figure 3. Extract from Rykachev[21] showing the dates of freezing and thawing of the Neva River in St. Petersburg/Leningrad, Russia, from 1706 to 1869 CE.

2.2. Evaluation

To identify any climate signal contained in documentary climate data, we formulated forward models[23] based on instrumental global monthly climate fields. These models not only serve for evaluation in this paper but are directly relevant for climate reconstruction approaches including data assimilation. We used temperature fields from BEST[24] and HadCRUT5[25], and precipitation fields from GPCC[26] to extract the time series from the closest grid point to each documentary site. The number of overlapping years between proxy and climate series had to be superior to 20 years (for all African wetness/dryness indices, as an exception, we accepted 10 overlapping years as otherwise no evaluation would have been possible on the entire dataset). The starting date was usually dictated by the start of the reference dataset, the end date by the end of the documentary dataset (see exceptions below), but never later than 1950 CE in order to avoid calibrating a forward model in a period in which climate or environment are no longer comparable with earlier periods.

The forward model took the form of a multiple regression model (see Fig. 5), in which a documentary series was expressed as a linear combination of monthly series of the corresponding driving variable (either temperature or precipitation). If the season or month was specified in the source (e.g., monthly, seasonal, or annual indices), these months were used. If this information was unavailable, we used annual mean values. In the case of events that were indicated as a specific date (e.g., phenological data), we also included lagged predictor variables (i.e., temperature from one or several previous calendar months). The window to be included was determined in a backward selection approach. In this case, the models initially included 6 months prior to the event in question (defined as the 90th percentile), such that an entire growing season could possibly be covered. Then a selection was carried out, retaining only months that were significant at $p<0.1$. Insignificant months between two significant months were also retained. If no significant months were identified, no model was calculated. For the phenological series covered in Reichen et al.[20], we made use of the more detailed information available. For instance, strongly skewed variables were transformed logarithmically, and we used the reference dataset and reference period given in the paper. The procedure is sketched in Figure 4.

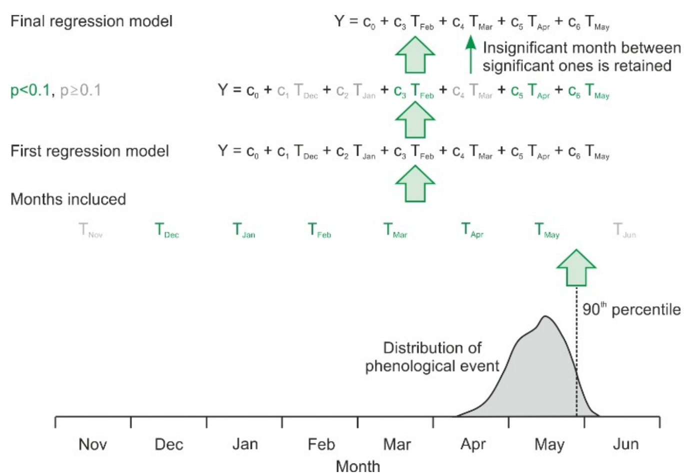


Figure 4. Schematic figure of forward modeling approach.

We then fitted regression models with a least-squares estimator. As a measure of goodness of fit, we used the correlation between the observed and modelled documentary series, along with the p-value. The following information on the evaluation is indicated in the example data file (Fig. 5): the reference period used (1829-1879 CE in this example case), the reference dataset (BEST), the model (monthly mean temperature of March and April; any transformation of variables would be indicated here), correlation, p-value, and error variance of the residuals. For some of the ice phenological records, we have also digitized nearby temperature records as the existing global databases did not have any data in close vicinity. These new data have been published in Reichen et al.[20] and Lundstad et al.[27] (https://doi.org/10.1594/PANGAEA.940724), and in these cases “REFDATA” is denoted by the label “station”.

It should be noted that, first, the residual error is not a measure of the error of the documentary proxy, but of the difference between the actual observation and the forward modelled observation (regression error). Consequently, it also contains the inherent error in the instrumental climate data and in the interpolation (this is the error required for data assimilation approaches). Second, this evaluation measures the error only in recent times when instrumental climate data are available. As a result, the quality of the documentary data in the earlier period may differ[28].

For all proxies that do not have a sufficiently long (see thresholds above) overlap with instrumental climate fields or nearby station records, an independent evaluation was not attempted here. This is indicated with an "NA" throughout the evaluation section of the data file. Other methods of evaluation are possible, but this requires more knowledge and hence we refer to the original publications. It would be possible to compare these cases with reconstructions such as EKF400v2[5], a global, monthly three-dimensional climate reconstruction covering 1600-2003 CE. EKF400v2 is based on an off-line assimilation approach of proxy data (e.g., tree-ring width, maximum late wood density), documentary data, and early instrumental data into an ensemble of atmospheric model simulations. However, in many cases the documentary data were assimilated in EKF400v2 (and hence datasets are not independent), while in cases where no information is locally available, EKF400v2 basically represents a model simulation, so no strong correlation is expected. Accordingly, we use EKF400v2 only in Sect. 4 for a case study.

Some documentary indices continue into the instrumental era as the authors have complemented them with degraded instrumental data or have used instrumental data in addition to documentary data. These data may then not be independent of instrumental data. These values were not removed from DOCU-CLIM. However, in the evaluation conducted in this study, the calibration period in such cases is limited to years before 1900 CE. Where available, information on whether a value was from a documentary source or from degraded observations was added to the "META" column.

3. Data records

The DOCU-CLIM dataset can be downloaded from the BORIS repository (https://boris-portal.unibe.ch/handle/20.500.12422/207).[29] The dataset comprises 621 files (note that a monthly index series is split into 12 files), totaling more than 100,000 values (Fig. 6). Information on all series, including links to the original holding, is given in the readme-file of the dataset.[29] The references of the original series are included in this paper (Ref. 16, 18, 21, 22, 30-132). The files are in ASCII format with 27 columns and a variable number of lines. The files are structured in a way that allows for straightforward inclusion into data assimilation schemes (see Fig. 5). In each file, one line covers one year. As a consequence, monthly data are stored in 12 files, one for each calendar month. The first seven columns contain information about the series, version number, and geographical location. These are identical for each line in the file. Then come the year, and the month (only given if the record resolution is monthly or seasonal). Where documentary information refers to a specific time (e.g., date of freezing), the month is set to NA. The column "STATISTIC" indicates whether the observation is a state (such as a date of freezing), or a mean value (e.g., a seasonal mean index, in which case the indicated month gives the last month of the averaged interval and the column "WINDOW" the number of months averaged. The column "BOREALSEASON" indicates the closest match to a season: winter (Dec-Feb), spring (Mar-May), summer (Jun-Aug), autumn (Sep-Nov), or annual (for ice freezing and thawing series we additionally used “earlywinter” and “latespring”). The next columns indicate variable name, unit, and type (for all series in this paper, the type is “DOCU”), then follows the column “VALUE” that contains the actual time series values. The next seven columns refer to fitting metrics of the forward model, as described in Sect. 2.2. Finally, the last three columns provide metadata such as a reference, the ID of the corresponding series in Burgdorf[13], and a column “META” that contains further information (for several entries, they are separated by “|”). In all cases the META column provides the original value (which is often the same as the value itself). Figure 5 provides an example where the freezing date is given in yr-mon-day. For further metadata on the series and collections, the reader is referred to the inventory by Burgdorf [13].

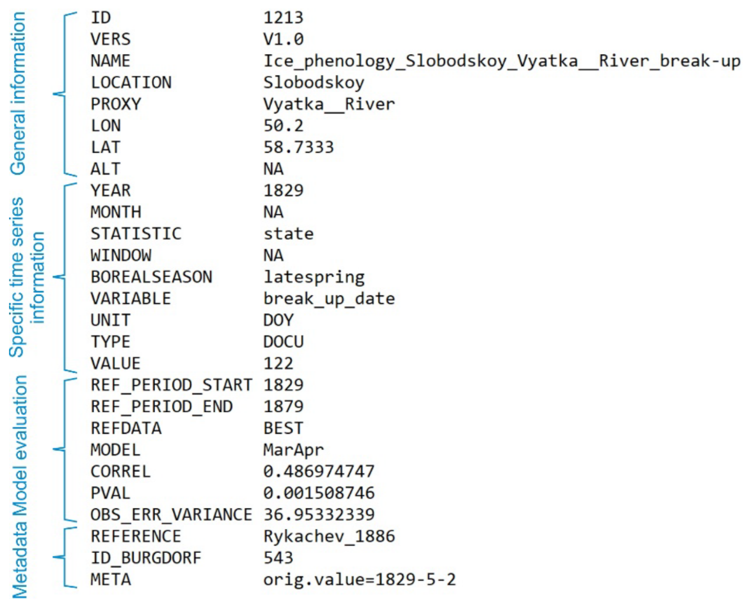


Figure 5. Data format with 27 columns using the example of the first line of series 1213 (transposed for clearer visualization).

The file structure is further explained in the supplementary file “DOCU_CLIM_File_Description.txt”. All file names start with “DOCU_CLIM” followed by the ID as six-digit integer, followed by the version number “V1.0” and the name of the variable; all of these elements are separated by an underscore character “_”. As an example, the file shown in Fig. 5 is named: “DOCU_CLIM_001213_V1.0_Ice_phenology_Slobodskoy_Vyatka__River_break-up.txt”. R-code to read the files is given in the Supplementary material.

Most of the document-based climate records are from Europe, which is partly due to our selection criteria. However, there are also records from Asia and North America. Data for Africa mostly concern precipitation[30]. We only have a few documentary records from South America [31,32] and only two from Australia[33,34].

One of the advantages of documentary proxy data is that they encompass all seasons (left bottom panel in Fig. 6). For example, plant phenology reflects temperature in spring and summer (sometimes autumn). Ice phenology reflects conditions from late autumn to spring as ice-on dates are associated with autumn and winter air temperatures and ice-off dates are highly correlated to winter and spring air temperatures. Many of the indices such as temperature and precipitation indices for various regions of Europe are seasonal or even monthly. Some documentary data such as wetness/dryness in Africa indicate annual conditions (interpreted as yearly means).[30]

The earliest records that extend back to the 15th century or further are mainly from Europe, China and Japan (top right panel in Fig. 6). Some records from South America begin as early as the 16th century, and those from North America date back to the 18th century, while the earliest African records start around 1800. Many of the ice phenological data date back to the 1800s, but there are also longer records such as that of Lake Suwa, Japan, beginning in 1443. The oldest records are typically the longest (right panels in Fig. 6), as many continue to the start of instrumental observations or even beyond.

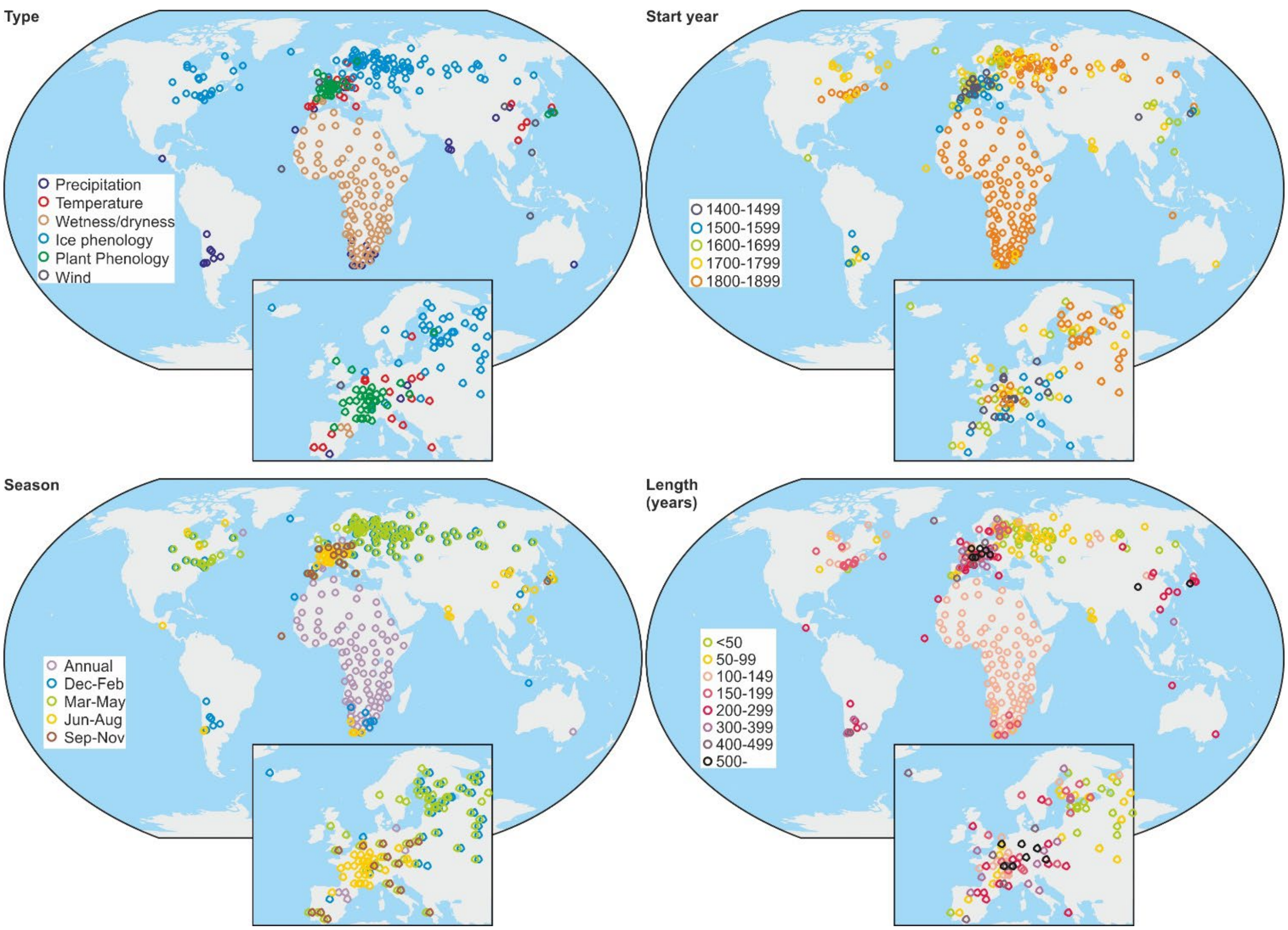


Figure 6. Map of all documentary data categorized according to (top left) the variable, (top right) the start year, (bottom left) the season covered, and (bottom right) the length.

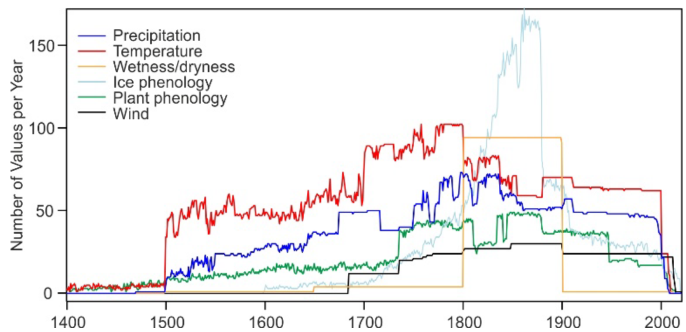


Figure 7. Number of values in the DOCU-CLIM dataset as a function of year and type of proxy.

Many records start after 1500 CE (Fig. 7) and the maximum coverage is in the late 19th century (note that from 1880 onward, no new records were added, although many phenological series start later). Temperature and precipitation indices are the most frequent record type, and wind indices[34–37] the least. However, the numbers for different types vary in time (Fig. 7). During the 19th century, when many weather stations were already measuring temperature in Europe, the number of documentary temperature series decreases and ice phenological records dominate.

As an example of series in DOCU-CLIM, Figure 8 shows three time series of freeze-up dates of Russian rivers. Ice phenology is the dominating type of documentary data in DOCU-CLIM during most of the 19th century. While the Neva series is continuous, the one from the Ob in Barnaul has a long gap. The series from the Volga in Saratov are shorter and contain an outlier (freeze-up date: 14 Feb). Note that outliers were not filtered out for the following evaluation.

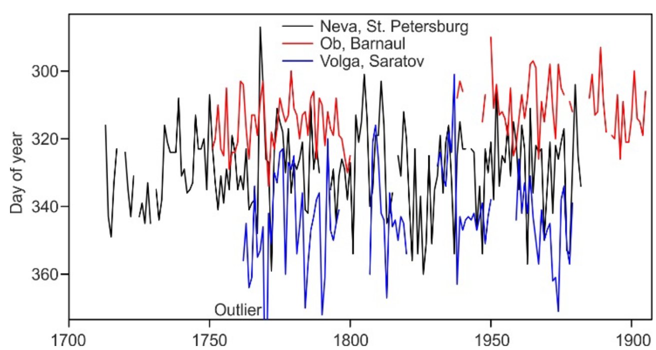


Figure 8: Time series of the freeze-up date of three Russian rivers. Note the reverse scale on the y-axis.

4. Technical validation

For many of the time series, a technical validation was performed by the original authors. These evaluations include arguably the most specific, expert judgement, combining local knowledge on both the historical sources and the local-to-regional climate characteristics. Readers are encouraged to consult the original publication for specific details (references are indicated in the inventory file as well as in the data files; a link is also given to the inventory by Burgdorf[13] where further information is available).

Here we report the results of our independent validation, as described in Section 2. In Figure 9 we show correlations for forward models that we calibrated in gridded instrument-based datasets. All records that have no overlap with observations (and thus evaluation was not possible) are denoted by grey dotted circles.

We found robust and highly significant correlations across Europe, North America and Asia. Many of them are related to derived indices, plant phenology, or ice phenology. Somewhat lower but still highly significant correlations are found over South America. Spatially-varying correlations are found over Africa, where most of the series were evaluated based on only 10 years of overlap. Moreover, the precipitation dataset[26], which the documentary data were compared with, may have substantial errors in these pre-1900 years. However, several significant correlations can be found in locations including southern Africa and Australia.

The distribution of the correlations (Fig. 9) shows that an overwhelming number of series exhibits correlations above 0.5 and the peak of the distribution occurring at correlations between 0.7 and 0.9, which is higher than the correlations found for forward tree rings[7] modelling. The highest correlations are observed for temperature indices, which are however often not fully independent from instrumental data in the overlapping period. Very high correlations are however, also found for ice phenological data and grape harvest dates.

The evaluation of the 421 records with models demonstrates that many documentary series have significant potential for quantitative applications. However, the series that were not evaluated (due to the absence of instrumental series in the vicinity, or total lack of overlap) will require further investigation and consultation in the literature before incorporation into a climate reconstruction.

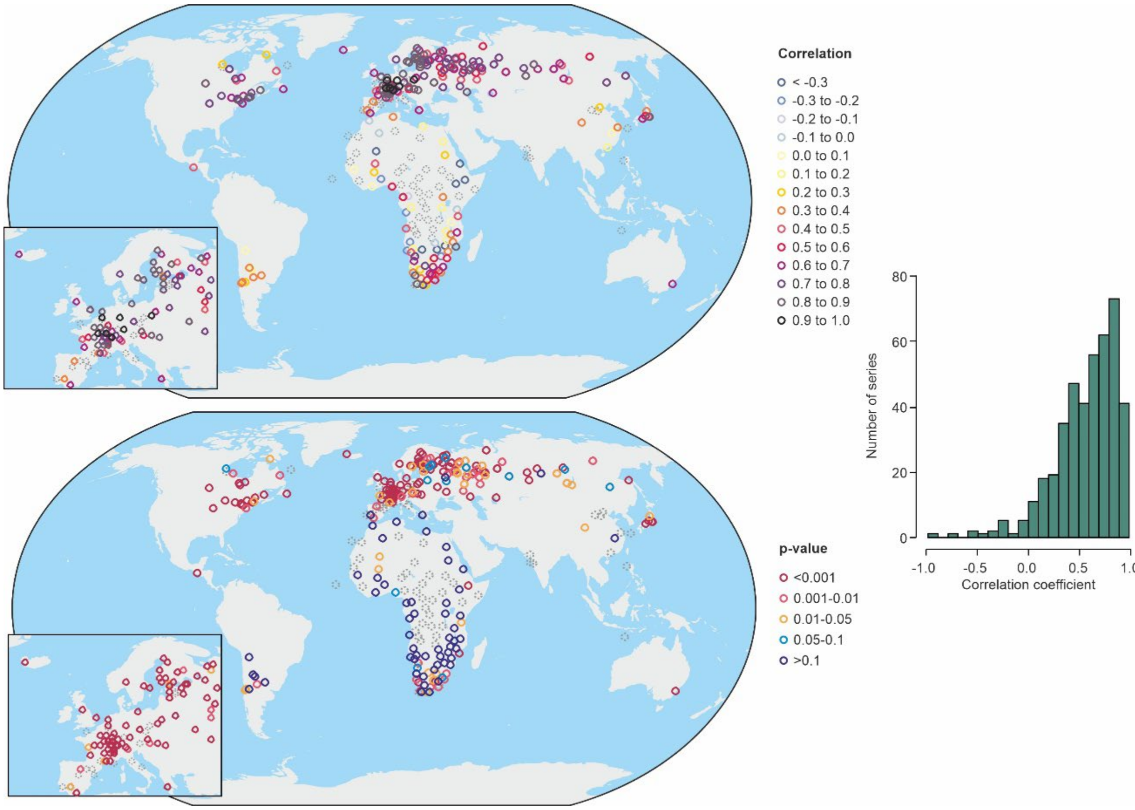


Figure 9. Map of Pearson correlation coefficients between documentary data and forward modelled data (top), *p*-values (bottom) and histogram showing the distribution of the correlation coefficients. Grey dotted circles indicate series where no evaluation was possible.

Usage notes

The DOCU-CLIM dataset[29] provides climate information from documentary data with the main aim of facilitating climate reconstructions. The dataset contains information on the correlation with corresponding forward models. This information should be carefully considered before using the data. Although climate reconstruction is the primary aim, the data can also be used in the form of individual time series.

To demonstrate the potential of this new documentary dataset for quantitative climate analysis, we present a case study for the year 1835. In January 1835, the volcano Cosigüina in Nicaragua erupted and released massive amounts of sulfuric aerosols into the atmosphere. It is considered one of the largest historical volcanic eruptions in the Americas and led to widespread environmental impacts[133]. We investigated temperature-related series, and precipitation or wetness/dryness-related time series, for the year 1835 CE. For temperature, we differentiated two extended seasons: boreal spring to summer (March to July), and autumn to early winter (August to December), as this division fits best with the material contained in the documentary data (thawing dates and spring/summer phenology, freezing dates and autumn phenology). Monthly series were averaged to these seasons. For precipitation and wetness/dryness indices, we considered annual indices or means of all seasons. As a reference we chose the period 1841-1870 CE (retaining only series with more than 20 years of data in this period) and standardized the series with respect to this reference. Finally, the sign of series was adjusted such that, for temperature, positive indicates warmer conditions (e.g., spring flowering or thawing dates were multiplied by -1 as earlier dates indicate warming; the sign of freezing dates was kept as early freezing indicates low temperatures). Series that were already assimilated in EKF400v2 were excluded from this analysis. We then compared these anomalies to the EKF400v2 reanalysis where we performed the same procedure with global monthly fields. Seasonal and annual averages were calculated, and the fields for 1835 were presented as standardized anomalies from the 1841-1870 base period. Figure 10 shows the standardized anomalies of the documentary proxies (top row) and EKF400v2 (bottom row) for the year 1835 CE. The two sets of data are entirely independent, and they are plotted on the same scale.

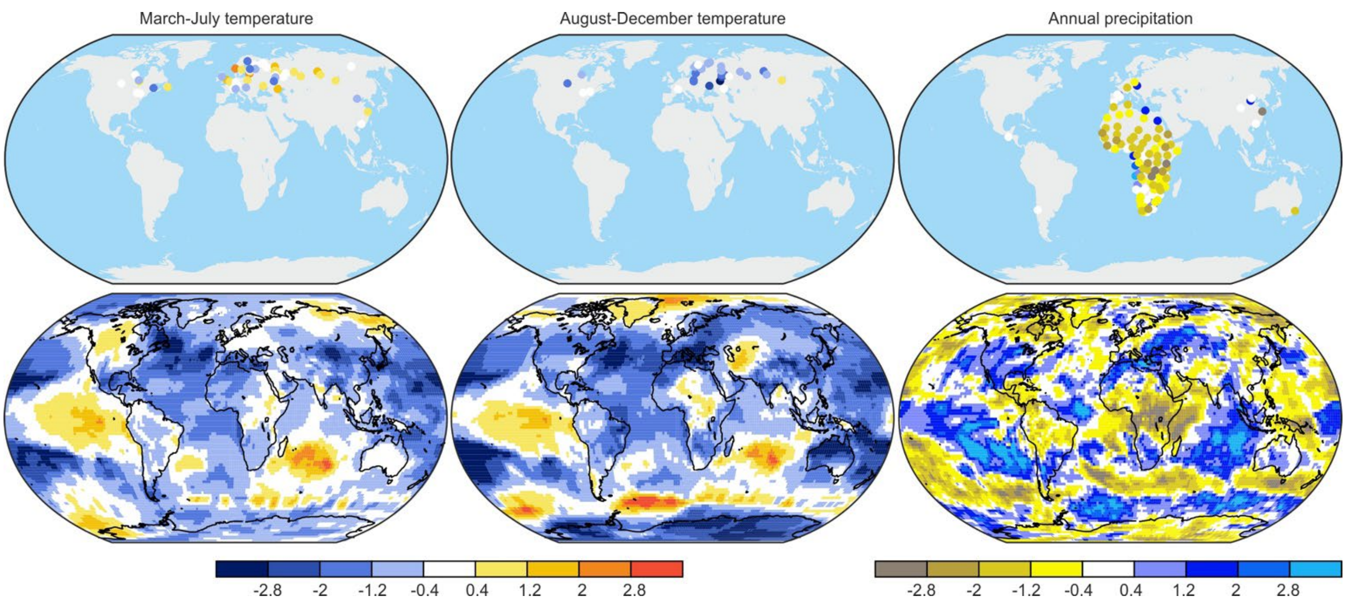


Figure 10. Standardized climatic anomalies in 1835 (with respect to the period 1841-1870) for (top) documentary data and (bottom) EKF400v2. Variables include temperature in (left) spring to summer and (middle) autumn to early winter and (right) annual precipitation.

EKF400v2 shows a general cooling that is arguably related to the volcanic eruption. However, not all regions cooled in all seasons. In boreal spring and summer, Europe and northern Eurasia have standardized anomalies around zero, and in some regions, temperature anomalies are positive. Although EKF400v2 assimilates no ice phenology from Siberian rivers except for the Angara in Irkutsk, these regions also showed neutral or slightly warm conditions. During autumn and early winter, the documentary data (particularly the early freezing of rivers) suggest a general cooling across the northern mid-latitudes. This coldness corresponds well with the EKF400v2 anomaly fields. Finally, precipitation in EKF400v2 indicates drying in most parts of Africa and wetting around the Mediterranean. This pattern is also observed in most of the documentary data in Africa (none of which were assimilated into EKF400v2). The large-scale cooling and the drying of areas influenced by the African monsoon agree with the expected effects of a tropical volcanic eruption[134]. Overall, our analysis shows that our documentary dataset (DOCU-CLIM)[29] can capture spatial climate variability associated with the prominent volcanic eruption of 1835.

The DOCU-CLIM dataset[29] can be used for climate reconstruction, particularly for data assimilation, which can make full use of the data and the metadata provided on the forward modeling. Some documentary time series could not be validated and should be further analyzed. DOCU-CLIM is a global dataset[29] and can now be combined with other multi-proxy compilations such as the PAGES 2k[6] datasets, or instrumental datasets such as H-CLIM[27], to generate new climate reconstructions.

Care should be taken when evaluating the series for trends. We have not analyzed the suitability of the records for trend analyses and advise testing this further before using the dataset for this purpose. The answer may well depend on the proxy type considered (phenological data, thermal index, etc.). Possible future updates of the DOCU-CLIM dataset may offer the data in a range of other existing data formats.[135]

Code availability

R code for generating the plots in this paper, for reading in all files and extracting desired information, and for the forward modeling is available from https://github.com/sbroennimann/DOCU-CLIM.

References

1. IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* (Cambridge University Press, 2021). doi:10.1017/9781009157896.

2. Bhend, J., Franke, J., Folini, D., Wild, M. & Brönnimann, S. An ensemble-based approach to climate reconstructions. *Climate of the Past* **8**, 963–976 (2012).

3. Hakim, G. J. *et al.* The last millennium climate reanalysis project: Framework and first results. *Journal of Geophysical Research: Atmospheres* **121**, 6745–6764 (2016).

4. Tardif, R. *et al.* Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling. *Climate of the Past* **15**, 1251–1273 (2019).

5. Valler, V., Franke, J., Brugnara, Y. & Brönnimann, S. An updated global atmospheric paleo-reanalysis covering the last 400 years. *Geosci Data J* **00**, 1–19 (2021).

6. Emile-Geay, J. *et al.* A global multiproxy database for temperature reconstructions of the Common Era. *Sci Data* **4**, 170088 (2017).

7. Breitenmoser, P., Brönnimann, S. & Frank, D. Forward modelling of tree-ring width and comparison with a global network of tree-ring chronologies. *Climate of the Past* **10**, 437–449 (2014).

8. Shah, S. K., Pandey, U., Mehrotra, N., Wiles, G. C. & Chandra, R. A winter temperature reconstruction for the Lidder Valley, Kashmir, Northwest Himalaya based on tree-rings of Pinus wallichiana. *Clim Dyn* **53**, 4059–4075 (2019).

9. PAGES2k. A global multiproxy database for temperature reconstructions of the Common Era. *Sci Data* **4**, 170088 (2017).

10. Pfister, C., Rohr, C. & Jover, A. C. C. Euro-Climhist: eine Datenplattform der Universität Bern zur Witterungs-, Klima- und Katastrophengeschichte. *Wasser Energie Luft* **109**, 45–48 (2017).

11. Rodrigo, F. S. The climate of Granada (southern Spain) during the first third of the 18th century (1706-1730) according to documentary sources. *Climate of the Past* **15**, 647–659 (2019).

12. *The Palgrave Handbook of Climate History*. (Palgrave Macmillan UK, 2018). doi:10.1057/978-1-137-43020-5.

13. Burgdorf, A.-M. A global inventory of quantitative documentary evidence related to climate since the 15th century. *Climate of the Past* **18**, 1407–1428 (2022).

14. Nash, D. J. *et al.* Climate indices in historical climate reconstructions: a global state of the art. *Clim. Past* **17**, 1273–1314 (2021).

15. Riemann, D., Glaser, R., Kahle, M. & Vogt, S. The CRE tambora.org – new data and tools for collaborative research in climate and environmental history. *Geosci Data J* **2**, 63–77 (2015).

16. Benson, B., Magnuson, J. & Sharma, S. Global Lake and River Ice Phenology Database, Version 1. *NSIDC: National Snow and Ice Data Center.* (2000) doi:https://doi.org/10.7265/N5W66HP8.

17. García-Herrera, R. *et al.* CLIWOC: A Climatological Database for the World's Oceans 1750–1854. *Clim Change* **73**, 1–12 (2005).

18. Wang, P. K. *et al.* Construction of the REACHES climate database based on historical documents of China. *Sci Data* **5**, 180288 (2018).

19. Veale, L. *et al.* Dealing with the deluge of historical weather data: the example of the TEMPEST database. *Geo* **4**, e00039 (2017).

20. Reichen, L. *et al.* A decade of cold Eurasian winters reconstructed for the early 19th century. *Nat Commun* **13**, 1–9 (2022).

21. Rykachev, M. *Openings and freezings of rivers in the Russian Empire*. (тип. Имп. Акад. наук, 1886).

22. Shostakovich, V. B. *Opening and freezing of waters of Asian Russia*. (паровая типо-литография П. И. Макушина и В. М. Посохина, 1909).

23. Dee, S. G., Steiger, N. J., Emile-Geay, J. & Hakim, G. J. On the utility of proxy system models for estimating climate states over the common era. *J Adv Model Earth Syst* **8**, 1164–1179 (2016).

24. Rohde, R. *et al.* A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview* **01**, 1–7 (2013).

25. Morice, C. P. *et al.* An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. *Journal of Geophysical Research: Atmospheres* **126**, e2019JD032361 (2021).

26. Schneider, U. *et al.* Evaluating the Hydrological Cycle over Land Using the Newly-Corrected Precipitation Climatology from the Global Precipitation Climatology Centre (GPCC). *Atmosphere 2017, Vol. 8, Page 52* **8**, 52 (2017).

27. Lundstad, E. et al. The global historical climate database HCLIM. *Scientific Data* **10,** 44 (2022). https://doi.org/10.1038/s41597-022-01919-w

28. Dobrovolný, P. *et al.* Monthly, seasonal and annual temperature reconstructions for Central Europe derived from documentary evidence and instrumental records since AD 1500. *Clim Change* **101**, 69–107 (2010).

29. Burgdorf, A.-M. *et al.* DOCU-CLIM: A global documentary climate dataset for climate reconstructions. *BORIS*, https://doi.org/10.48620/167 (2022)

30. Nicholson, S. E., Klotter, D. & Dezfuli, A. K. Spatial reconstruction of semi-quantitative precipitation fields over Africa during the nineteenth century from documentary evidence and gauge data. *Quaternary Research* **78**, 13–23 (2012b).

31. Prieto, M. R. & García Herrera, R. Documentary sources from South America: Potential for climate reconstruction. *Palaeogeogr Palaeoclimatol Palaeoecol* **281**, 196–209 (2009).

32. Neukom, R. *et al.* An extended network of documentary data from South America and its potential for quantitative precipitation reconstructions back to the 16th century. *Geophys Res Lett* **36**, L12703 (2009).

33. Fenby, C. & Gergis, J. L. Rainfall variations in south-eastern Australia part 1: Consolidating evidence from pre-instrumental documentary sources, 1788-1860. *International Journal of Climatology* **33**, 2956–2972 (2013).

34. Gallego, D., García-Herrera, R., Peña-Ortiz, C. & Ribera, P. The steady enhancement of the Australian Summer Monsoon in the last 200 years. *Sci Rep* **7**, 1–7 (2017).

35. Gallego, D., Ordóñez, P., Ribera, P., Peña-Ortiz, C. & García-Herrera, R. An instrumental index of the West African Monsoon back to the nineteenth century. *Quarterly Journal of the Royal Meteorological Society* **141**, 3166–3176 (2015).

36. Barriopedro, D. *et al.* Witnessing North Atlantic westerlies variability from ships' logbooks (1685-2008). *Clim Dyn* **43**, 939–955 (2014).

37. Vega, I. *et al.* Reconstructing the western North Pacific summer monsoon since the late nineteenth century. *J Clim* **31**, 355–368 (2018).

38. Adamson, George C.D. & David J. Nash. Documentary Reconstruction of Monsoon Rainfall Variability over Western India, 1781–1860. *Clim. Dyn.* **42,** 749–69 (2014).

39. Alcoforado, M.-J. et al. Temperature and precipitation reconstruction in southern Portugal during the late Maunder Minimum (AD 1675-1715). *The Holocene* **10,** 333-340 (2000).

40. Amano, T., Smithers, R. J., Sparks, T. H. & Sutherland, W. J. A 250-year index of first flowering dates and its response to temperature changes. *Proc. R. Soc. B Biol. Sci.* **277,** 2451–2457 (2010).

41. Aono, Y. & Kazui, K. Phenological Data Series of Cherry Tree flowering in Kyoto Japan and Its Application to Reconstruction of Springtime Temperatures since the 9th Century. *Int. J. Climatol.* **28,** 905–914 (2008).

42. Aono, Y. & Omoto, Y. Estimation of Temperature at Kyoto since the 11th Century. Using Flowering Data of Cherry Trees in Old Documents. *J. Agricult. Meteorol.* **49,** 263–272 (1994).

43. Aono, Y. & Saito, S. Clarifying Springtime Temperature Reconstructions of the Medieval Period by Gap-Filling the Cherry Blossom Phenological Data Series at Kyoto, Japan. *Int. J. Biometeorol.* **54,** 211–219 (2010).

44. Aono, Y. & Tani, A. Autumn temperature deduced from historical records of autumn tints phenology of maple tree in Kyoto, Japan. *Climate in Biosphere* **14,** 18-28 (2014).

45. Aono, Y. Cherry Blossom Phenological Data since the Seventeenth Century for Edo (Tokyo), Japan, and Their Application to Estimation of March Temperatures. *Int. J. Biometeorol.* **59,** 427–434 (2015).

46. Ball, T. F. Historical and instrumental evidence of climate: western Hudson Bay, Canada, 1714-1850. In: Climate since the A.D. 1500 (eds. Bradley, R. S. & Jones, P. D.), p. 40–73, Routledge (1995).

47. Barrat, J. M. D. First Appearance of Ice, the Closing and the time of Opening of the Connecticut River at Middletown. *The American Journal of Science and Arts* **39,** 88–90 (1840).

48. Bartholy, J. et al. Classification and analysis of past climate information based on historical documentary sources for the Carpathian Basin. *Int. J. Climatol.* **24,** 1759-1776 (2004).

49. Bullón, T. Winter Temperatures in the Second Half of the Sixteenth Century in the Central Area of the Iberian Peninsula. *Clim. Past* **4,** 357–367 (2008).

50. Camenisch, C. Endless Cold: A Seasonal Reconstruction of Temperature and Precipitation in the Burgundian Low Countries during the 15th Century Based on Documentary Evidence. *Clim. Past* **11,** 1049–1066, (2015).

51. Camuffo, D. et al. 500-Year Temperature Reconstruction in the Mediterranean Basin by Means of Documentary Data and Instrumental Observations. *Clim. Change* **101,** 169–99 (2010).

52. Camuffo, D. et al. Climate change in the Mediterranean over the last five hundred years. In: Planet earth 2011 – global warming challenges and opportunities for policy and practice (ed. Carayannis, E. G.). InTech, Rijeka (2011).

53. Catchpole, A. J. W. & Faurer, M.-A. Summer sea ice severity in Hudson Strait, 1751–1870. *Clim. Change* **5,** 115–139 (1983).

54. Catchpole, A. J. W. Hudson's Bay Company Ships' Log-Books as Sources of Sea Ice Data, 1751–1870. In: Climate since the A.D. 1500 (eds. Bradley, R. S. & Jones, P. D.), p. 17–39, Routledge (1995).

55. Catchpole, A. J. W., Moodie, D. W. & Milton, D. Freeze-Up and Break-Up of Estuaries on Hudson Bay in the Eighteenth and Nineteenth Centuries. *Can. Geogr. Géographe Can.* **20,** 279–297 (1976).

56. Daux, V. et al. An Open-Access Database of Grape Harvest Dates for Climate Research: Data Description and Quality Assessment. *Clim. Past* **8,** 1403–1418 (2012).

57. de Vries, J. Histoire du climat et économie : des faits nouveaux, une interprétation différente. *Ann. Hist. Sci. Soc.* **32,** 198–226 (1977).

58. Defila, C. & Clot, B. Phytophenological trends in Switzerland. *Int. J. Biometeorol.* **45,** 203–207 (2001).

59. Dobrovolný, P. et al. Precipitation Reconstruction for the Czech Lands, AD 1501–2010. *Int. J. Climatol.* **35,** 1–14 (2015).

60. Eklund, A. Isläggning och islossning i svenska sjöar. (Long observation series of ice freeze and break up dates in swedish lakes), SMHI Hydrologi No. 81. Sveriges Meteorologiska och Hydrologiska Institut, Nörrköping, pp. 1-24 (1999).

61. Fernández-Fernández, M. I. et al. The climate in Zafra from 1750 to 1840: precipitation. *Clim. Change* **129,** 267–280 (2015).

62. García, R. et al. Reconstruction of the precipitation in the Canary Islands for the period 1595-1836. *B. Am. Meteorol. Soc.* **84,** 37-40 (2003).

63. Ge, Q. S. et al. Reconstruction of Historical Climate in China, High-Resolution Precipitation Data from Qing Dynasty Archives. *B. Am. Meteorol. Soc.* **86,** 671–769 (2005).

64. Ge, Q. S. et al. The Rainy Season in the Northwestern Part of the East Asian Summer Monsoon in the 18th and 19th Centuries. *Quat. Sci. Rev.* **229,** 16–23 (2011).

65. Gergis, J. & Ashcroft, L. Rainfall Variations in South-Eastern Australia Part 2: A Comparison of Documentary, Early Instrumental and Palaeoclimate Records, 1788–2008. *Int. J. Climatol.* **33,** 2973–2987 (2013).

66. Gimmi, U. et al. A Method to Reconstruct Long Precipitation Series Using Systematic Descriptive Observations in Weather Diaries: The Example of the Precipitation Series for Bern, Switzerland (1760–2003). *Theor. App. Clim.* **87,** 185–197 (2007).

67. Gioda, A. Para una Historia Climática de La Paz en los Últimos Cinco Siglos. *Revista de La Coordinadora de Historia* **3,** 13–33 (1999).

68. Gioda, A. & Prieto, M. R. Histoire des sécheresses andines. Potosí, El Niño et le petit age glaciaire. *La Météorologie* **8,** 33–42 (1999).

69. Glaser, R. & Riemann, D. A thousand-year record of temperature variations for Germany and Central Europe based on documentary data. *J. Quaternary Sci.* **24,** 437–449 (2009).

70. Grab, S. & Zumthurm, T. The land and its climate knows no transition, no middle ground, everywhere too much or too little: a documentary-based climate chronology for central Namibia, 1845–1900. *Int. J. Climatol.* **38,** e643-e659 (2018).

71. Guevara-Murua, A. et al. 300 years of hydrological records and societal responses to droughts and floods on the Pacific coast of Central America. *Clim. Past* **14,** 175-191 (2018).

72. Hannaford, M. J. et al. Early-nineteenth-century southern African precipitation reconstructions from ships' logbooks. *The Holocene* **25,** 379-390 (2015).

73. Hao, Z. X. et al. Winter Temperature Variations over the Middle and Lower Reaches of the Yangtze River since 1736 AD. *Clim. Past* **8,** 1023–1030 (2012).

74. García-Herrera, R. & Prieto, M. R. Floods in the semiarid Argentinean Chaco during the 17th to 19th centuries. In: Proceedings of Palaeofloods, Historical Data & Climatic Variability: Applications in Flood Risk Assessment (ed. Thorndycraft, V. R. et al.), pp. 107-112, CSIC-Centro de Ciencias Medioambientales, Madrid (2003).

75. Hildebrandsson, H. H. Sur le prétendu changement du climat européen en temps historiques. *Nova Acta R. Soc. Sci. Upps.* **4(4),** 5 (1915).

76. Hill, B. T. & Jones, S. J. The Newfoundland ice extent and the solar cycle from 1860 to 1988. *J. Geophys. Res.* **95,** 5385 (1990).

77. Hirano, J. et al. Reconstruction of July temperature variations since the 1830s in Kawanishi based on historical weather documents. *Geographical Review of Japan (Series A)* **86,** 451-464 (2013). (in Japanese).

78. Holopainen, J. et al. Plant phenological data and tree-rings as palaeoclimate indicators since AD 1750 in SW Finland. *Int. J. Biometeorol.* **51,** 61-72 (2006).

79. Kajander, Ja. Methodological Aspects on River Cryophenology Exemplified by a Tricentennial Break-up Time Series from Tornio. *Geophysica* **29,** 73-95 (1993).

80. Kelso, C. & Vogel, C. The Climate of Namaqualand in the Nineteenth Century. *Clim. Change* **83,** 357–380 (2007).

81. Kiss, A. et al. An experimental 392-year documentary-based multi proxy (vine and grain) reconstruction of May-July temperatures for Koszeg, West-Hungary. *Int. J. Biometeorol.* **55,** 595-611 (2011).

82. Koslowski, G. & Glaser, R. Reconstruction of the ice winter severity since 1701 in the Western Baltic. *Clim. Change* **31,** 79-98 (1995).

83. Kouraev, A. et al. The ice regime of Lake Baikal from historical and satellite data: Relationship to air temperature, dynamical, and other factors. *Limnology and Oceanography* **52,** 1268-1286 (2007).

84. Koslowski, G. & Glaser, R. Variations in reconstructed ice winter severity in the Western Baltic from 1501 to 1995, and their implications for the North Atlantic oscillation. *Clim. Change* **41,** 175-191 (1999).

85. Leijonhufvud, L. et al. Five Centuries of Stockholm Winter/Spring Temperatures Reconstructed from Documentary Evidence and Instrumental Observations. *Clim. Change* **101,** 109–141 (2010).

86. Livingstone, D. M. Break-up dates of Alpine lakes as proxy data for local and regional mean surface air temperatures. *Clim. Change* **37,** 407–439 (1997).

87. Loader, N. J. et al. Spring temperature variability in northern Fennoscandia AD 1693-2011. *J. Quat. Sci.* **26,** 566-570 (2011).

88. Magne, M. A. Two Centuries of River Ice Dates in Hudson Bay Region from Historical Sources. (University of Manitoba, 1981).

89. Magnuson, J. J. et al. Historical trends in lake and river ice cover in the Northern Hemisphere. *Science* **289,** 1743–1746 (2000).

90. Matuszko, D. Klimat Krakowa w XX wieku (Climate of Cracow in 20th Century). IGiGP UJ, 251 pp. (2007)

91. Maurer, C., Koch, E., Hammerl, C., Hammerl, T. & Pokorny, E. BACCHUS temperature reconstruction for the period 16th to 18th centuries from Viennese and Klosterneuburg grape harvest dates. *J. Geophys. Res.* **114,** D22106, doi:10.1029/2009JD011730 (2009).

92. Meier, N. et al. Grape harvest dates as a proxy for Swiss April to August temperature reconstructions back to AD 1480. *Geophys. Res. Lett.* **34,** L20705 (2007).

93. Mikami, T. Long term variations of summer temperatures in Tokyo since 1721. *Geographical Reports of Tokyo Metropolitan University* **31,** 157-165 (1996).

94. Mikami, T. Climatic Variations in Japan Reconstructed from Historical Documents. *Weather* **63,** 190–193 (2008).

95. Mougin, M. Etudes glaciologiques en Savoie. *Études Glaciologiques* **3,** 1–113, (1912).

96. Možný M. et al. Cereal harvest dates as a proxy for Czech Republic March to June temperature. *Clim. Change* **110,** 801-812 (2012).

97. Možný M. et al. Drought reconstruction based on grape harvest dates for the Czech Lands, 1499−2012 *Clim. Res.* **70,** 119-132 (2016).

98. Nash, D. J. & Endfield, G. H. 'Splendid Rains Have Fallen': Links Between El Niño and Rainfall Variability in the Kalahari, 1840–1900. *Clim. Change* **86,** 257–290 (2008).

99. Nash, D. J. & Endfield, G. H. A 19th century climate chronology for the Kalahari region of central southern Africa derived from missionary correspondence. *Int. J. Climatol.* **22,** 821–841 (2002).

100. Nash, D. J. & Grab, S. W. 'A Sky of Brass and Burning Winds': Documentary Evidence of Rainfall Variability in the Kingdom of Lesotho, Southern Africa, 1824–1900. *Clim. Change* **101,** 617–653 (2010).

101. Nash, D. J. et al. Seasonal rainfall variability in southeast Africa during the nineteenth century reconstructed from documentary sources. *Clim. Change* **134,** 605-619 (2016).

102. Neukom, R. et al. Multi-Proxy Summer and Winter Precipitation Reconstruction for Southern Africa Over the Last 200 Years. *Clim. Dyn.* **42,** 2713–2726 (2014).

103. Nicholson, S. E. A Semi-Quantitative, Regional Precipitation Data Set for Studying African Climates of the Nineteenth Century, Part I. Overview of the Data Set. *Clim. Change* **50,** 317-353 (2001).

104. Nicholson, S. E. et al. A Two-Century Precipitation Dataset for the Continent of Africa. *B. Am. Meteorol. Soc.* **93,** 1219–1231 (2012).

105. Nordli, Ø. et al. A Late-Winter to Early-Spring Temperature Reconstruction for Southeastern Norway from 1758 to 2006. *Annals of Glaciology* **46,** 404–408 (2007).

106. Ogilvie, A. E. J. & Jónsdóttir, I. Sea Ice, Climate, and Icelandic Fisheries in the Eighteenth and Nineteenth Centuries. *Arctic* **53,** 383–394 (2000).

107. Ogilvie, A. E. J. The past climate and sea-ice record from Iceland, Part 1: Data to A.D. 1780. *Clim. Change* **6,** 131–152 (1984).

108. Ouellet-Bernier, M.-M. & de Vernal, A. Winter freeze-up and summer break-up in Nunatsiavut, Canada, from 1770 to 1910. *Past Glob. Chang. Mag.* **28,** 52–53 (2020).

109. Pfister, C. Monthly temperature and precipitation in central Europe 1525-1979: quantifying documentary evidence on weather and its effects. In: Climate since the A.D. 1500 (eds. Bradley, R. S. & Jones, P. D.), p. 92-117, Routledge (1995).

110. Pfister, C. The creation of high resolution spatio- temporal reconstructions of past climate from direct meteorological observations and proxy data. Methodological considerations and results. Climate in Europe 1675-1715. In: Climatic Trends and Anomalies in Europe 1675–1715 (eds. Frenzel, B., Pfister C. & Glaeser, B.), p. 329–376, Stuttgart (1994).

111. Prieto, M. R. & Herrera, R. De sequías, hambrunas, plagas y otras varias y continuas calamidades acaecidas en la Jurisdicción de Córdoba durante el siglo XVIII, Serie Economía y Sociedad. *Cuadernos de Historia* **4,** 131-158 (2001).

112. Prieto, M. R. ENSO Signals in South America: Rains and Floods in the Paraná River Region during Colonial Times. *Clim. Change* **83,** 39–54 (2007).

113. Prieto, M. R. et al. Historical Evidences of the Mendoza River Streamflow Fluctuations and Their Relationship with ENSO. *The Holocene* **9,** 472–481 (1999).

114. Prieto, M. R., Herrera, R. & Dussel, P. Archival evidence for some aspects of historical climate variability in Argentina and Bolivia during the 17th and 18th centuries. In: Southern Hemisphere Paleo and Neo-Climates: Methods and Concepts (eds. Smolka, P. P. & Volkheimer, W.), pp. 127-142, Springer, Berlin (2000).

115. Rannie, W. F. Breakup and freezeup of the Red River at Winnipeg, Manitoba Canada in the 19th century and some climatic implications. *Clim. Change* **5,** 283-296 (1983).

116. Rodrigo, F. S. et al. A 500-year precipitation record in Southern Spain. *Int. J. Climatol.* **19,** 1233-1253 (1999).

117. Rodrigo, F. S. et al. A reconstruction of the winter North Atlantic Oscillation Index back to AD 1501 using documentary data in Southern Spain. *J. Geophys. Res.* **106,** 14805-14818 (2001).

118. Rutishauser, T. Cherry Tree Phenology. Interdisciplinary Analyses of Phenological Observations of the Cherry Tree in the Extended Swiss Plateau Region and Their Relation to Climate Change. Diploma thesis, University of Bern 122 pp. (2003).

119. Seinä A. & Palosuo E. The classification of the maximum annual extent of ice cover in the Baltic Sea 1720–1995. *MERI-Report Series of the Finnish Inst. of Marine Res.* **27,** 79–91 (1996).

120. Seinä, A. et al. Ice seasons 1996–2000 in Finnish sea areas. *MERI-Report Series of the Finnish Inst. of Marine Res* **43,** 132 (2001).

121. Shabalova, M. V. & van Engelen A. F. V. Evaluation of a Reconstruction of Winter and Summer Temperatures in the Low Countries, AD 764–1998. *Clim. Change* **58,** 219–242 (2003).

122. Shipman, T. G. Ice-conditions on the Mississippi River at Davenport, Iowa. *Northwest Sci.* **12,** 590–594 (1938).

123. Sparks, T. H. & Carey, P. D. The Responses of Species to Climate Over Two Centuries: An Analysis of the Marsham Phenological Record, 1736-1947. *J. Ecol.* **83,** 321 (1995).

124. Takács, K., Kern, Z. & Pásztor, L. Long-term ice phenology records from eastern–central Europe. *Earth Syst. Sci. Data* **10,** 391–404 (2018).

125. Tan, L. et al. Precipitation variations of Longxi, northeast margin of Tibetan Plateau since AD 960 and their relationship with solar activity. *Clim. Past* **4,** 19-28 (2008).

126. Tarand, A. & Nordli, Ø. The Tallinn temperature series reconstructed back half a millennium by use of proxy data. *Clim. Change* **48,** 189-199 (2001).

127. Taulis, E. De la distribution des pluies au Chili. *Materiaux pour l'étude des calamités* **33,** 3-20 (1934).

128. Teillet, J. V. A Reconstruction of Summer Sea lce Conditions in the Labrador Sea Using Hudson's Bay Company Ships' Log-Books, 1751 to 1870. University of Manitoba, Winnipeg (1988).

129. Tejedor, E. et al. Rogation ceremonies: a key to understanding past drought variability in northeastern Spain since 1650. *Clim. Past* **15,** 1647–1664 (2019).

130. Van Engelen, A. F. V. et al. A Millennium of Weather, Winds and Water in the Low Countries. In History and Climate: Memories of the Future? (ed. Jones P. D. et al.), pp. 101–123. New York, Springer (2001).

131. Vogel, C. H. A Documentary-Derived Climatic Chronology for South Africa, 1820–1900. *Clim. Change* **14,** 291–307 (1989).

132. Wetter, O. & Pfister, C. Spring-Summer Temperatures Reconstructed for Northern Switzerland and Southwestern Germany from Winter Rye Harvest Dates, 1454–1970. *Clim. Past* **7,** 1307–1026 (2011).

133. Longpré, M. A., Stix, J., Burkert, C., Hansteen, T. & Kutterolf, S. Sulfur budget and global climate impact of the A.D. 1835 eruption of Cosigüina volcano, Nicaragua. *Geophys Res Lett* **41**, 6667–6675 (2014).

133. Raible, C. C. *et al.* Tambora 1815 as a test case for high impact volcanic eruptions: Earth system effects. *WIREs Climate Change* **7**, 569–589 (2016).

134. McKay, N. P. and Emile-Geay, J. Technical note: The Linked Paleo Data framework – a common tongue for paleoclimatology. *Clim. Past* **12,** 1093–1100, https://doi.org/10.5194/cp-12-1093-2016 (2016).

Acknowledgements

This work was supported by the European Commission (ERC Grant PALAEO-RA, 787574) and by Swiss National Science Foundation project WeaR (188701). Simulations underlying EKF400v2 were performed at the Swiss National Supercomputing Centre CSCS. ET is supported by a Marie Skłodowska-Curie Action ("ITHACA-101024389", and the Government of Aragón through the "Program of research groups" (group H09_20R, "Climate, Water, Global Change, and Natural Systems").

Competing interests statement

The authors declare no competing interests.

Author contributions

AMB compiled all datasets in this paper, SB formatted the data and performed the forward modeling, all authors contributed datasets, all authors commented on the manuscript.