Climatic Change
DOI 10.1007/s10584-009-9724-x

Monthly, seasonal and annual temperature reconstructions for Central Europe derived from documentary evidence and instrumental records since AD 1500

Petr Dobrovolný · Anders Moberg · Rudolf Brázdil · Christian Pfister · Rüdiger Glaser · Rob Wilson · Aryan van Engelen · Danuta Limanówka · Andrea Kiss · Monika Halíčková · Jarmila Macková · Dirk Riemann · Jürg Luterbacher · Reinhard Böhm

Received: 24 October 2008 / Accepted: 14 July 2009

Abstract Monthly temperature series for Central Europe back to AD 1500 are developed from documentary index series from Germany, Switzerland and the Czech Republic (1500–1854) and 11 instrumental temperature records (1760–2007). Documentary evidence from the Low Countries, the Carpathian Basin and Poland are used for cross-checking for earlier centuries. The instrumental station records are corrected for inhomogeneities, including insufficient radiation protection of early thermometers and the urban heat island effect. For overlapping period (1760–1854), the documentary data series correlate with instrumental temperatures, most strongly in winter (86% explained variance in January) and least in autumn (56% in September). For annual average temperatures, 81% of the variance is explained. Verification statistics indicate high reconstruction skill for most months and seasons. The last 20 years (since 1988) stand out as very likely the warmest 20-year period, accounting for the calibration uncertainty and decreases in proxy data quality before the calibration period. The new reconstruction displays a previously unobserved

P. Dobrovolný (✉) · R. Brázdil · M. Halíčková · J. Macková
Institute of Geography, Masaryk University, 611 37 Brno, Czech Republic
e-mail: dobro@sci.muni.cz
URL: http://www.geogr.muni.cz/en//

A. Moberg
Department of Physical Geography and Quaternary Geology, Stockholm University, Stockholm, Sweden

C. Pfister
Institute of History, Bern University, Bern, Switzerland

R. Glaser · D. Riemann
Institute for Physical Geography, University of Freiburg, Freiburg im Breisgau, Germany

R. Wilson
School of Geography & Geosciences, University of St Andrews, St Andrews, Scotland, UK

Published online: 30 September 2009

long-term decrease in DJF, MAM and JJA temperature variability over last five centuries. Compiled monthly, seasonal and annual series can be used to improve the robustness of gridded large-scale European temperature reconstructions and possible impact studies. Further improvement of the reconstruction would be achieved if documentary data from other European countries are further developed.

1 Introduction

Past climate variations in the pre-instrumental period can be estimated from different kinds of proxy data. Apart from natural proxy archives (e.g. tree rings, ice cores, corals or lake sediments), important information for climate reconstruction can be found in documentary sources. Herein, we present climate information from non-instrumental, man-made sources that are referred to as documentary evidence (see Brázdil et al. 2005 for a review; Pfister et al. 2008). This study is motivated by the fact that documentary evidence can be, in several aspects, complementary, or even more appropriate, compared to natural proxies. We demonstrate that temperature information derived from documentary sources, that are qualitative in nature, can be successfully used to derive quantitative climate reconstructions using the same statistical approaches that are well developed, as in dendroclimatology.

Historical documentary evidence has the potential to reveal details of past climatic changes in Europe for several centuries. It covers mostly the period since the sixteenth century (Pfister and Brázdil 1999), but sources are frequent even for the medieval period in some regions (see, for example, Shabalova and van Engelen (2003) for the Low Countries, or Glaser (2001, 2008) for Germany). Weather descriptions in documentary sources relate to all parts of the year and their temporal resolution frequently goes down to monthly or even daily and sub-daily detail. Hence, documentary data have notable advantages over natural proxy data, which do not resolve different seasons within the same year.

An unavoidable property of documentary evidence is a substantial spatial heterogeneity of the sources. They can be rarely found in the form of consistent and relatively long time series recorded for one particular location. Thus documentary evidence are usually gathered at local, regional and national levels and already have been used to reconstruct climate for several Central European countries: Switzerland

A. van Engelen
Royal Netherlands Meteorological Institute, De Bilt, The Netherlands

D. Limanówka
Institute of Meteorology and Water Management, Kraków, Poland

A. Kiss
Department of Physical Geography and Geoinformatics, University of Szeged, Szeged, Hungary

J. Luterbacher
Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus-Liebig University, Giessen, Germany

R. Böhm
Central Institute for Meteorology and Geodynamics, Vienna, Austria

(Pfister 1984, 1992, 1999), the Czech Lands (Brázdil 1996; Dobrovolný et al. 2009), Germany (Glaser 1997, 2008; Glaser and Stangl 2004; Glaser and Riemann 2009), Poland (Przybylak et al. 2005) and the Carpathian basin (Rácz 1999; Bartholy et al. 2004). Moreover, documentary proxy evidence have been used as an integral part of multi-proxy climate reconstructions (e.g. Mann et al. 2000; Luterbacher et al. 2004, 2007; Guiot et al. 2005; Xoplaki et al. 2005).

The use of documentary evidence for climate reconstruction is limited due to the fact that a lot of information in traditional sources of historical climatology (e.g. institutional sources, diaries or newspapers) fades away with the onset of instrumental measurements. Thus early instrumental measurements frequently substitute descriptive proxies in these sources and this fact hampers the possibility of finding a relatively long overlapping period between documentary evidence and instrumental data. However, the existence of such an overlapping period is crucial for calibration and verification calculations (e.g. Cook et al. 1994). As a consequence, it has often not been possible to quantify the robustness of documentary proxy data. Nevertheless, such data have been used and played an important role in reconstructions of temperature, precipitation and atmospheric pressure fields over Europe back to AD 1500 (Luterbacher et al. 2002, 2004, 2007; Xoplaki et al. 2005; Pauling et al. 2006).

Here we explore the possibilities for using documentary data from six European countries/regions—the Czech Republic (CZ), Germany (DE), Switzerland (CH), the Low Countries (LC), Poland (PL) and the Carpathian Basin (CB)—to reconstruct monthly temperatures by calibrating and verifying the proxy data against early instrumental data. Previous existing index series (e.g. Brázdil 1996; Pfister 1999; Rácz 1999) have been revised and substantially updated for this work as part of the EU Sixth Framework Programme project "European climate of the past millennium" (MILLENNIUM). These documentary data are described in Section 2. Further, we use homogenized instrumental temperature records from 11 Central European (hereafter referred to as CEU) stations where all series have been adjusted for inhomogeneities and errors including instrumental errors, urban warming trends, station relocations, and the effect of poorer protection against radiation (see Böhm et al. 2009). These instrumental records are described in Section 3.

Documentary proxy series and instrumental measurements are compared for the overlapping period 1760–1854, which appears to be sufficient for statistical calibration, verification and for deriving uncertainty estimates. The core methodological aspects of our reconstruction are explained in Section 4. In Section 5 we present the basic features of the new temperature reconstruction at monthly, seasonal and annual resolution. Finally we discuss our results in a broader European context by comparing them with existing reconstructions and we summarize limitations and benefits of our new half millennium long temperature series.

2 Documentary evidence for Central Europe since AD 1500

Compared to natural climate proxies, documentary evidence represents a specific form, namely various weather related man-made historical reports. From the categorization given e.g. by Brázdil et al. (2005), it follows that documentary evidence represents a wide ranging group of direct and indirect data. Long and homogeneous quantitative reports on some temperature-related features can, in some cases, and for some localities, be used for temperature reconstruction directly. For instance

Leijonhufvud et al. (2008, 2009) utilize sources about the beginning of the sailing season in the Stockholm port to derive a winter–early spring temperature reconstruction back to sixteenth century. However, a considerable part of the documentary sources comprises indirect, qualitative, information on weather and climate. Sources including written narrative reports or visual daily weather records are usually transformed into ordinal-scale temperature indices before they can be used for quantitative temperature reconstruction (see Brázdil et al. 2005 for a review).

The transformation from descriptive information to an index scale relies on content analysis of a set of terms that the authors of reports used for description of temperature conditions. Different terms used by individual authors are interpreted as a single index value, where the ordinal scale indices express the extremity of temperature in a given month, usually on a seven-degree scale. Thus, a temperature index value of −3 means extremely cold conditions, −2 very cold, −1 cold, 0 normal, +1 warm, +2 very warm, +3 extremely warm (for more detailed description on construction of temperature indices, see e.g. Brázdil et al. 2005). Seasonal (winter—DJF, spring—MAM, summer—JJA, autumn—SON) and annual (Ann) temperature indices can be calculated as a sum of the corresponding monthly values. This means that seasonal indices may fluctuate between −9 and +9 and annual indices between −36 and +36. This approach to compiling index series was adopted for all countries involved in the present paper except the Low Countries. The LC index series exist only for warm and cold seasons that are defined differently compared to the standard season delimitation. In the LC case, the cold season index value reflects a cumulative index for November through March ranging from 1 (extremely mild) to 9 (extremely severe). Values for the warm season (May to September) range from 1 (extremely cool) to 9 (extremely warm) (van Engelen et al. 2001; Shabalova and van Engelen 2003).

In this study, we utilize temperature index series that were constructed exclusively from descriptive documentary evidence and have therefore not been 'biased' with indices constructed from non-systematic early instrumental measurements or from other proxies. Furthermore we prefer to use monthly indices that were constructed year by year using the seven-degree scale. Temperature index series for the following countries and periods were used: DE (1500–1760), CZ (1500–1854), CH (1501–1816), PL (1501–1700) and CB (1516–1870). The LC index data, because of their different definition, are not considered in this section but are used in Section 6 for cross-comparison.

It is one of the specific features of documentary sources, that they do not usually contain relevant weather information for all months of the period in question. Figure 1 presents an insight into the completeness of the national temperature index series. This figure allows a view of the number of available indices for individual monthly, seasonal and annual series through time. While there is an obvious spatio-temporal heterogeneity of documentary sources in the sixteenth to seventeenth centuries, the national index series in the eighteenth century (and in the nineteenth century whenever available) become more complete. Such temporal distribution of the indices reflects the changing character of the 'man-made' proxies; which is one of the main difficulties encountered when using these data for developing continuous temperature reconstructions.

As can bee seen in Fig. 1, there exist some seasonal index values when corresponding monthly indices are not available and it is important to note that seasonal

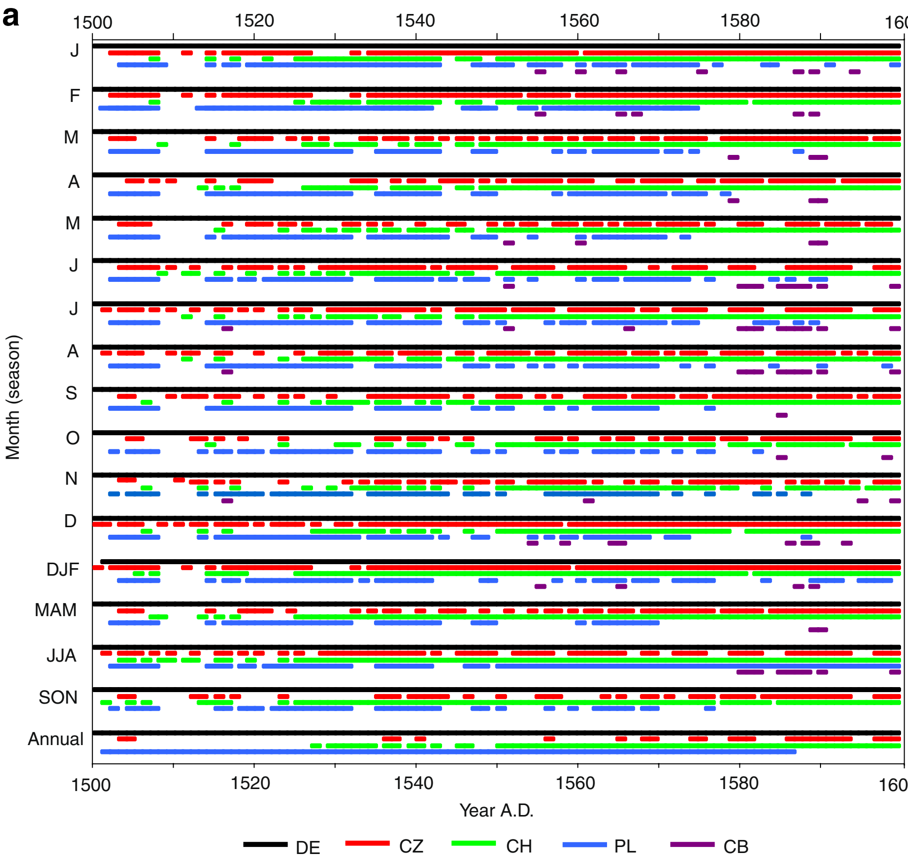


Fig. 1 **a** Completeness of monthly, seasonal and annual temperature index series for Germany (*DE*, 1500–1760), Switzerland (*CH*, 1501–1816), Czech Republic (*CZ*, 1500–1854), Poland (*PL*, 1500–1700) and Carpathian Basin (*CB*, 1516–1870) in the sixteenth century (1500–1599). **b** As **a** but for the seventeenth century (1600–1699). **c** As **a** but for the eighteenth century (1700–1799). **d** As **a** but for the nineteenth century (1800–1899)

indices can sometimes be interpreted directly from documentary sources that provide summary reports or descriptions about the whole season, but do not refer simultaneously to corresponding months. Such summary reports are considered as secondary sources of lower interpretational value in historical climatology. However, besides the conventional means of summing up corresponding monthly indices to provide seasonal values, an interpretation of such summary reports represents another way of deriving indices at seasonal resolution.

There is a long tradition of literary narrative sources (annals, town or family chronicles) in the sixteenth century Central Europe. Narrative reports on weather and climate frequently focus only on individual, mostly extreme events. Personal diaries are another group of sources that date from the second half of the sixteenth century. Compared to subsequent centuries they often cover shorter time spells and are unsystematically derived. However, some outstanding early sources covering

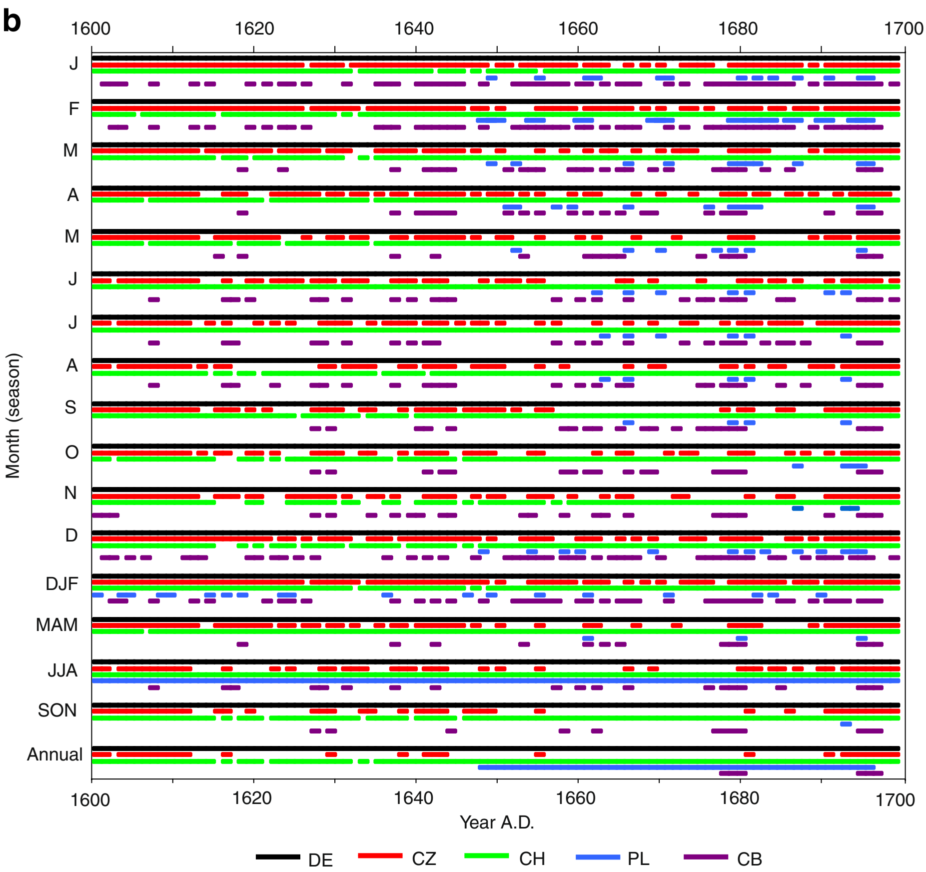


Fig. 1 (continued)

several decades of daily weather observations can be found. For example, Wolfgang Haller (1525–1601), archdeacon at the Cathedral of Zurich (Switzerland), kept a weather diary for more than 30 years (1545–1576) (Flohn 1979; Pfister 1984). Pfister et al. (1999) provide a detailed description of these sixteenth century documentary sources. Over time, new types of documentary sources successively appeared (e.g. newspapers) or some traditional sources became more frequent (weather diaries or sources of economic character). By the end of the seventeenth century weather reports had become a regular topic, e.g. in the "Nordische Mercurius" Newspapers published in Hamburg from 1667 to 1675 (Glaser 2008). Visual daily weather records from the diaries of the Premonstratensian abbey of Hradisko (Olomouc, CZ) for the period 1693–1783 (with some gaps, Brázdil et al. 2008) are of particular value for the eighteenth century.

There is an abundance of plant-phenological evidence in several CEU countries. As an example, Johann Jakob Sprüngli (1717–1803), a parson in the Canton of Bern (Switzerland), systematically documented the seasonal dynamics of 236 plant and 44 animal species as well as the timing of 46 agricultural activities from 1760 to 1803 (Pfister 1984; Burri and Rutishauser 2008). So called institutional sources (Pfister

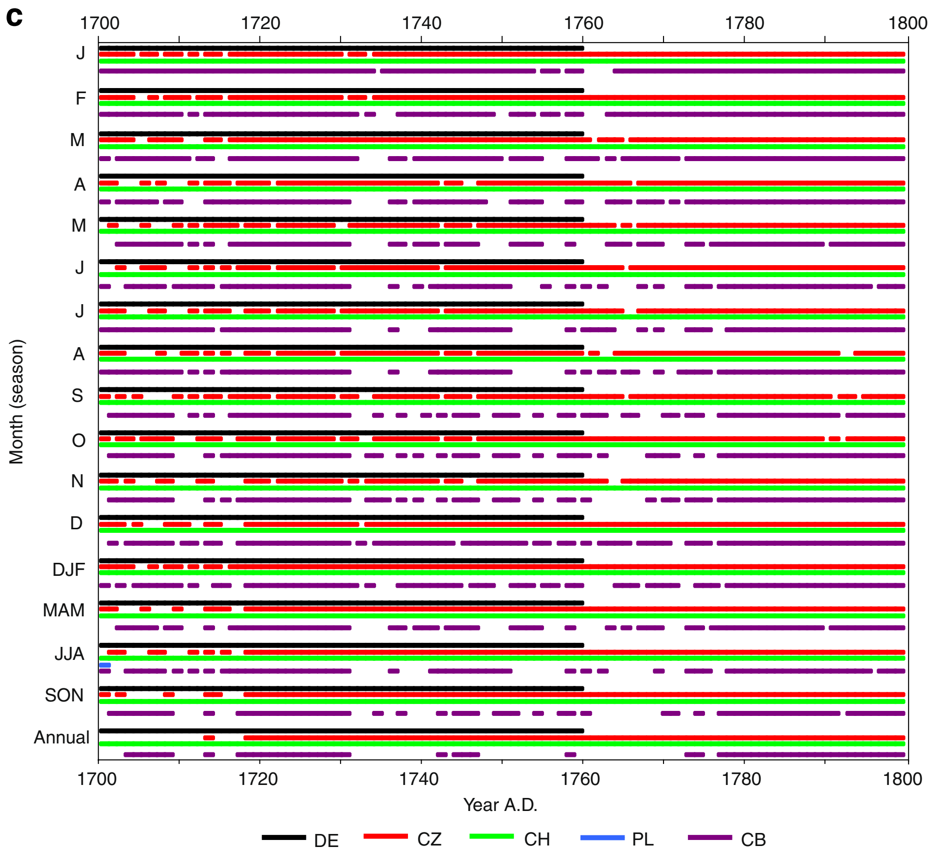


Fig. 1 (continued)

et al. 2008) bring information about the amount of vine production (Pfister 1981) and the beginning of agricultural activities such as hay, crop or vine harvesting. Such sources are already numerous, especially for Switzerland, from the sixteenth century (Pfister 1992; Meier et al. 2007) and for western Hungary since the early seventeenth century.

The description above provides an understanding as to why the national index series in Fig. 1 become gradually more complete towards the nineteenth century. Consequently, their quality improves because the resulting index values can be interpreted from a greater number of different sources. This can be seen as an analogue to the process of data gathering in dendrochronology, where the sample depth usually grows forwards in time. Conversely, the richness of documentary evidence can be influenced by social processes in some periods. While the CH index series show very high data density, except the first three decades of the sixteenth century, the CZ series is almost continuous from the second half of the eighteenth century. While there is a higher density of PL temperature indices in the first part of the sixteenth century, completeness of CB index series greatly improves after the beginning of the eighteenth century.

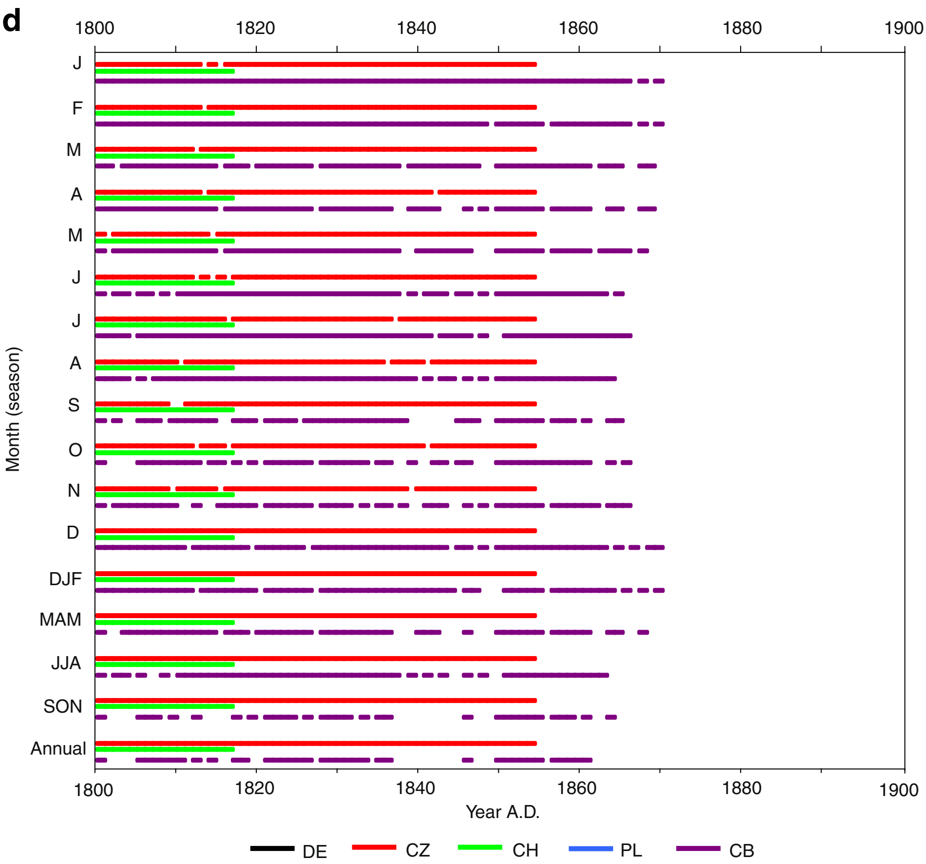


Fig. 1 (continued)

Figure 2 shows the overall completeness of national index series for the AD 1500–1850 period. The common period was chosen for a better comparison among all national series and Figs. 1 and 2 summarize to what extent each national series can

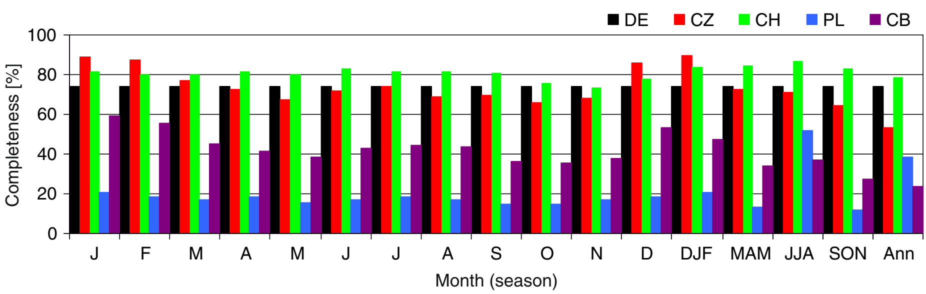


Fig. 2 Overall percentage of collected monthly, seasonal and annual temperature indices for DE, CH, CZ, PL and CB in the period 1500–1850

contribute to the reconstruction of temperatures in the pre-instrumental period, but also to what extent it can contribute to calibration and verification (see Section 4). While the indices for Germany are complete from 1500 to 1760, no DE index series from documentary evidence is available after this year. The PL index series ends in 1700, while the CZ, CH and CB index series have data up to various time points in the nineteenth century. The CB series is the most extensive in the nineteenth century and extends to AD 1870.

Data coverage for DE, CH and CZ over the common period 1500–1850 is comparable and the overall completeness varies from 70% to 90% for all months and seasons. Completeness of the other series is substantially lower reaching about 40% for CB series and 20% for PL series, respectively. Typically there is better coverage for the winter months compared to the other seasons of the year.

Substantial time heterogeneity of the index series when compared with natural proxies results from the fact that the weather-related notes were made by a varying number of individuals or institutions for a limited number of years. The number of missing values is smaller in time periods covered by multiple documentary sources when several sources are utilized to derive the final index value. However, in periods poorly replicated with documentary evidence, the way to overcome the problem of missing values is to average several regional or national series. Such averaging can usefully reduce some site-specific local influences, but it should be preceded by a comprehensive analysis of mutual relationships of the index series used. Only those index series that come from climatologically homogeneous regions should be combined (see Dobrovolný et al. 2009 for the CZ). Figure 3 shows results of correlation analysis for the five Central European index series. Each individual national series is compared to a simple average from all other national series excluding the series in question (for example DE is correlated with the average of CH, CZ, PL, and CB, etc.). An average value was computed only when at least two indices existed in the four 'other' national series.

The ordinal scale of documentary indices prompts the use of a rank correlation method. However, for the annual data, the index scale has a considerable number of possible values and with averages taken over four national series, the average series are similar to continuous normally distributed data. Because of this mix of data, we calculate correlation coefficients with both the Spearman (S) rank method and the Pearson (P) method. Figure 3 summarizes the results. Corresponding P and S values are very similar. Furthermore, we also found that the index series do not differ significantly from the normal distribution (not shown). Thus, we argue that an application of parametric statistics in our methodological approach (in Section 4) to temperature reconstruction using ordinary scale indices do not significantly influence the results.

High and statistically significant correlations (0.025 level, one-tailed test) are consistently found between the average index series and national DE, CH and CZ series respectively. The highest correlations are found for the winter months and for the DJF season. This can be explained by reference to the higher spatial coherence of winter temperatures in Central Europe, which are more frequently conditioned by large-scale circulation compared to summer conditions that are more determined by smaller scale processes. Several CB monthly index series also show a strong and significant correlation with the other series. The PL indices show overall weaker, and insignificant, correlations with the remaining national index series.

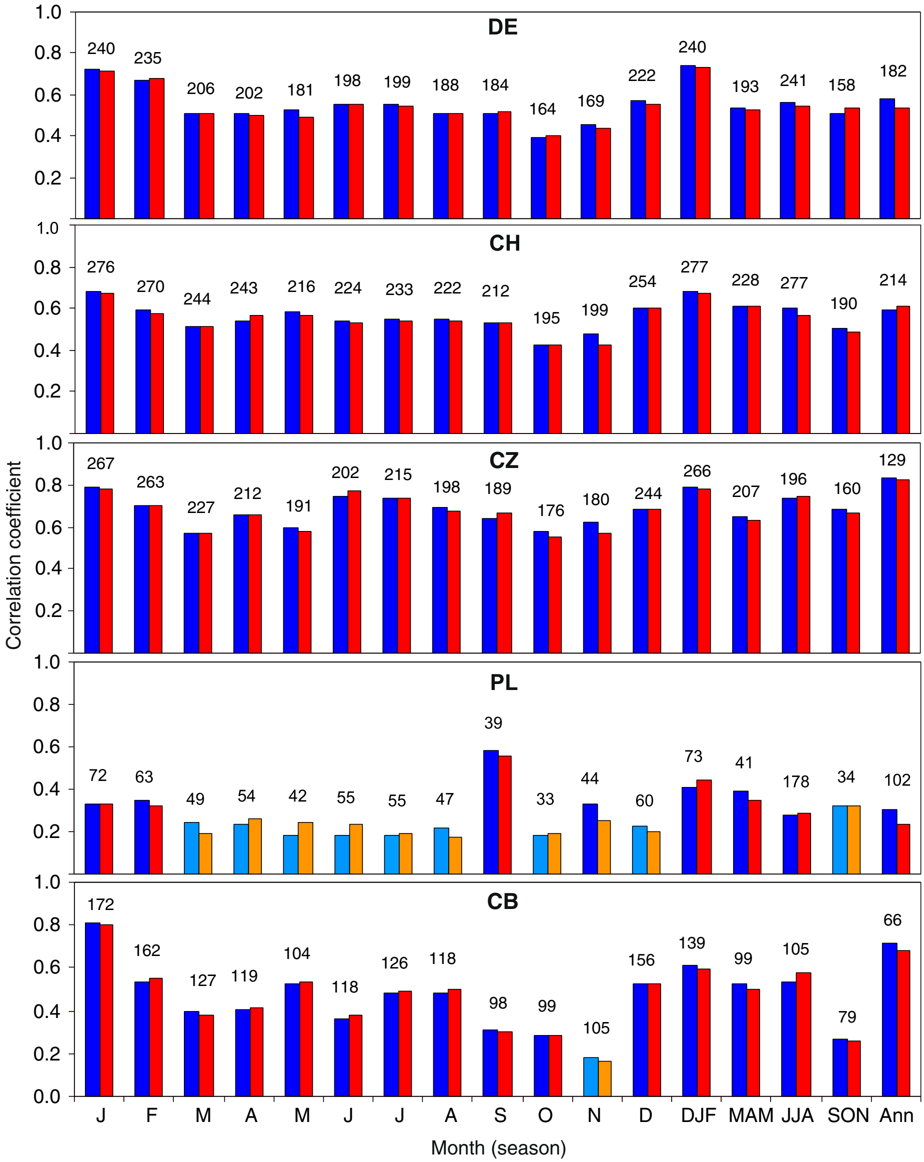


Fig. 3 Pearson correlation (*blue*) and Spearman rank correlation (*red*) coefficients between individual national index series and the average computed from the remaining national index series. Number of years for which correlations were calculated are introduced *above the individual bars*. *Pale colors* mark correlations not significant at 0.025 level for a one-sided significance test

We point to the fact that the correlation coefficients presented in Fig. 3 reflect both data quality and the spatial coherence of temperature variability within CEU. For example, temperature patterns of Poland or the Carpathian Basin can be quite different compared to Switzerland and Germany, especially in the summer half-year. A direct comparison of results for different countries in Fig. 3, is complicated by different lengths of the series. In particular, the PL data are available only before 1700, whereas the CB data are mostly complete after this period.

The analysis of data completeness of the five national index series and also the results of the correlation analysis lead us to the conclusion that, for the purpose of developing a long temperature reconstruction, the strongest potential is currently provided by the index series from Germany, the Czech Republic and Switzerland. Hence, in the following sections, we only use these three national index series for compilation of an averaged Central European temperature index series (as described in Section 4).

3 Instrumental temperature data

Instrumental temperature data are used for the following reasons: (1) for statistical calibration and verification of the documentary index data, and (2) for providing the necessary information for comparing temperature variations in the more recent period when documentary data are not available. In this section we provide a brief background related to long instrumental records and describe how we select station records and combine them to national averages and Central European average temperature series.

Continuous temperature measurements in Europe started at several places for different years during the eighteenth century and in some cases even in the late seventeenth century (Balling et al. 1998; Camuffo and Jones 2002; Jones 2001; Jones and Briffa 2006). The long temperature records, however, are in general affected by artificial factors such as station relocations, instrument changes, changes in observation hours, and growing urban heat island. The process of identifying and adjusting for such problems is called homogenization (Peterson et al. 1998). Careful homogenization has been undertaken for several long instrumental records from the Greater Alpine Region—the HISTALP dataset (Auer et al. 2007; Begert et al. 2005)—which partly covers the Central European region of interest here. The most recent of the HISTALP homogenization efforts were made by Böhm et al. (2009), to correct for the bias due to insufficient protection of early thermometers against direct solar radiation (Frank et al. 2007a). This feature has been labeled the Early Instrumental (EI) warm-bias problem.

We used ten station records from the new EI-adjusted version of the homogenized HISTALP dataset; four series from Germany and three each from Switzerland and Austria. The Austrian series are included, despite the lack of Austrian documentary index data, because they geographically connect the Czech Republic with Switzerland (Fig. 4). Moreover, due to the increased number of stations, they help to decrease the effect of possible uncorrected errors in individual station records.

We included the long instrumental Prague (Klementinum observatory) series from the Czech Republic. The homogenization of this series was based on results from the Standard Normal Homogeneity Test (SNHT) (Alexandersson 1986), with the averaged ten selected HISTALP station records as reference series. Corrections to the Prague data were applied separately to each monthly series, whenever significant breakpoints were determined by SNHT. Hence, for this study, we use 11 homogenized station records (Table 1).

All selected German stations are from the southern part of the country. Although some long records also exist from more northerly located German sites, we did not

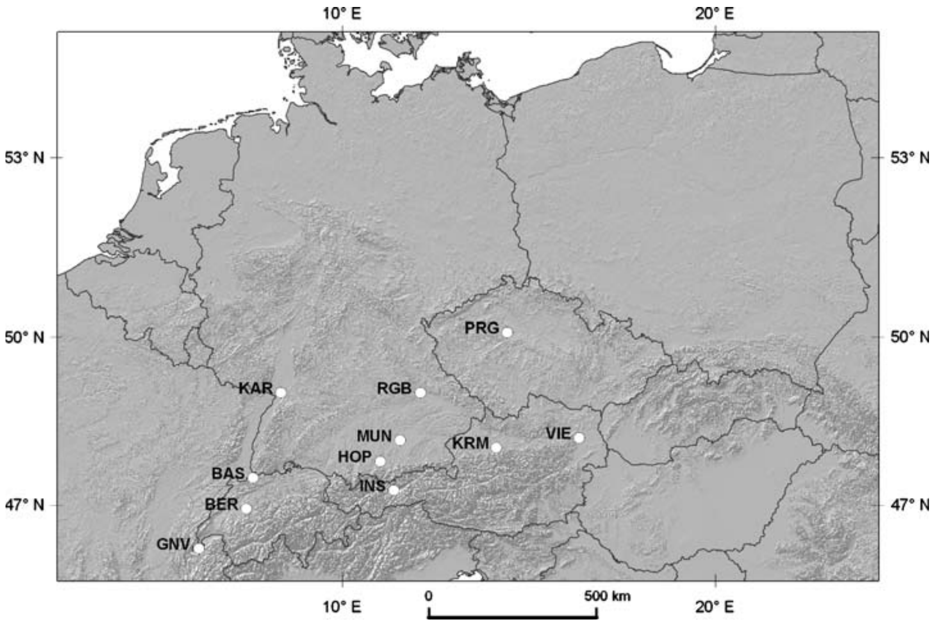


Fig. 4 Stations used for compilation of Central European temperature series. Original data © ESRI 2008. For station abbreviations see Table 1

include any of those as they have not been analyzed and corrected for the EI warm-bias. Temperature variability in the region covered by the selected stations is highly coherent. For seasonal mean temperatures, the station correlation matrix (Table 2) shows highest values for DJF and MAM, while the correlation is weaker for JJA and SON, but still mostly above 0.8. The strong spatial coherence of temperature variability within the region can be further highlighted using principal component (PC) analysis (e.g. von Storch and Zwiers 1999) of the station records. The first PC explains 91% of variability in DJF, MAM and annual temperatures. Explained variability for JJA and SON temperatures is slightly lower (88%), but all 11 station records have component loadings on the first PC exceeding 0.90.

Table 1 Stations used for compilation of Central European temperature series

Station	Abr.	Country	Lat [deg. N]	Lon [deg. E]	H [m a.s.l.]	Starting year
Kremsmünster	KRM	AT	48.05	14.13	389	1767
Vienna	VIE	AT	48.22	16.35	209	1775
Innsbruck	INS	AT	47.27	11.38	609	1777
Regensburg	RGB	DE	49.03	12.10	366	1773
Munich	MUN	DE	48.18	11.55	525	1781
Hohenpeissenberg	HOP	DE	47.80	11.02	986	1781
Karlsruhe	KAR	DE	49.03	08.35	112	1779
Basel	BAS	CH	47.60	07.60	316	1760
Geneva	GNV	CH	46.19	06.15	380	1760
Bern	BER	CH	46.93	07.42	565	1777
Prague	PRG	CZ	50.08	14.42	191	1771

Table 2 Pearson's correlation coefficients between seasonal temperatures for 11 CEU stations

Station	KRM	WIE	INS	BAS	GNV	BER	RGB	HOP	MUN	KAR	PRG
DJF											
KRM	1.00	0.97	0.84	0.90	0.82	0.86	0.97	0.87	0.96	0.93	0.95
WIE	0.96	1.00	0.82	0.88	0.79	0.83	0.94	0.86	0.92	0.92	0.96
INS	0.91	0.88	1.00	0.91	0.89	0.92	0.87	0.89	0.87	0.88	0.80
BAS	0.87	0.81	0.89	1.00	0.95	0.97	0.94	0.92	0.95	0.97	0.87
GNV	0.82	0.76	0.86	0.95	1.00	0.97	0.86	0.88	0.89	0.90	0.78
BER	0.86	0.81	0.88	0.96	0.95	1.00	0.90	0.90	0.91	0.92	0.81
RGB	0.94	0.92	0.92	0.91	0.86	0.90	1.00	0.89	0.97	0.96	0.94
HOP	0.93	0.88	0.94	0.94	0.90	0.94	0.94	1.00	0.91	0.91	0.84
MUN	0.95	0.91	0.93	0.93	0.89	0.92	0.97	0.96	1.00	0.97	0.92
KAR	0.89	0.84	0.89	0.96	0.92	0.94	0.92	0.94	0.93	1.00	0.92
PRG	0.95	0.95	0.88	0.85	0.81	0.84	0.94	0.90	0.93	0.88	1.00
MAM											
JJA											
KRM	1.00	0.94	0.86	0.83	0.78	0.82	0.91	0.91	0.91	0.83	0.89
WIE	0.93	1.00	0.85	0.81	0.76	0.81	0.90	0.90	0.88	0.82	0.90
INS	0.87	0.82	1.00	0.86	0.84	0.86	0.85	0.89	0.87	0.84	0.81
BAS	0.85	0.81	0.85	1.00	0.95	0.95	0.88	0.92	0.90	0.94	0.80
GNV	0.77	0.71	0.81	0.94	1.00	0.93	0.83	0.88	0.85	0.88	0.75
BER	0.82	0.78	0.84	0.96	0.92	1.00	0.86	0.92	0.88	0.90	0.79
RGB	0.94	0.89	0.86	0.88	0.78	0.85	1.00	0.92	0.95	0.88	0.89
HOP	0.88	0.85	0.90	0.91	0.85	0.89	0.89	1.00	0.94	0.89	0.86
MUN	0.92	0.89	0.84	0.85	0.75	0.82	0.95	0.89	1.00	0.89	0.86
KAR	0.87	0.82	0.83	0.95	0.87	0.91	0.89	0.88	0.87	1.00	0.86
PRG	0.91	0.92	0.81	0.86	0.78	0.82	0.90	0.85	0.85	0.86	1.00
SON											

Values above the main diagonal belong to DJF or JJA series (upper and lower table, respectively), while those below belong to MAM or SON series. For station abbreviations and length of corresponding common period see Table 1

From the selected station records, we created national average temperature series for Germany, Switzerland and Austria, respectively. All data were first converted to anomalies with respect to the 1961–1990 reference period. Because the starting year differs somewhat among the stations, the number of available records is less in the earliest years than in the main part of the series. To ensure that this does not cause any artificial change of variance in the national average series, we applied a variance adjustment to the averaged data. We essentially followed the procedure outlined by Jones et al. (2001). However, their purpose was to adjust the entire time series to have the theoretical variance of infinitely sampled grid-boxes. We rather adjusted the variance only in the years when some stations are missing, to correspond to the variance level for the finite number of complete series in the main part of the time period. For the Czech Republic, however, the national series is simply identical to the Prague series.

Finally, the four national series were arithmetically averaged to form a CEU average temperature series covering the period 1760–2007. Variance adjustment was applied to the average in the early years as not all four national series have data. As can be seen from Table 1, only two stations (Basel and Geneva) have data before 1767 and not until 1781 do all 11 stations contribute to the CEU average. Our choice of arithmetically averaging the nation-averages rather than equally weighting all 11 stations makes however, due to the very strong coherence of the data, little significant difference to the final result.

The degree of spatial integrity of the CEU average instrumental series across the whole European region can be examined using the HadCRUT3 5° × 5° gridded temperature dataset (Brohan et al. 2006). Figure 5 presents spatial correlation fields between seasonal CEU average series and HadCRUT3 data for the period 1850–2007. The significant correlations cover a broad region, which can be geographically defined as Central Europe. From that core region correlations decrease with increasing distance. The core region of highest coherence is spatially larger for winter than for summer. This is due to the fact that winter climate is more related to large-scale circulation whereas during summer local scale effects of convective cloudiness and associated differences in radiation balance are more relevant. The high spatial correlations shown in Fig. 5, along with correlation analysis of both index and instrumental series in Fig. 3 and Table 2, argue for labeling our analyzed series as Central European.

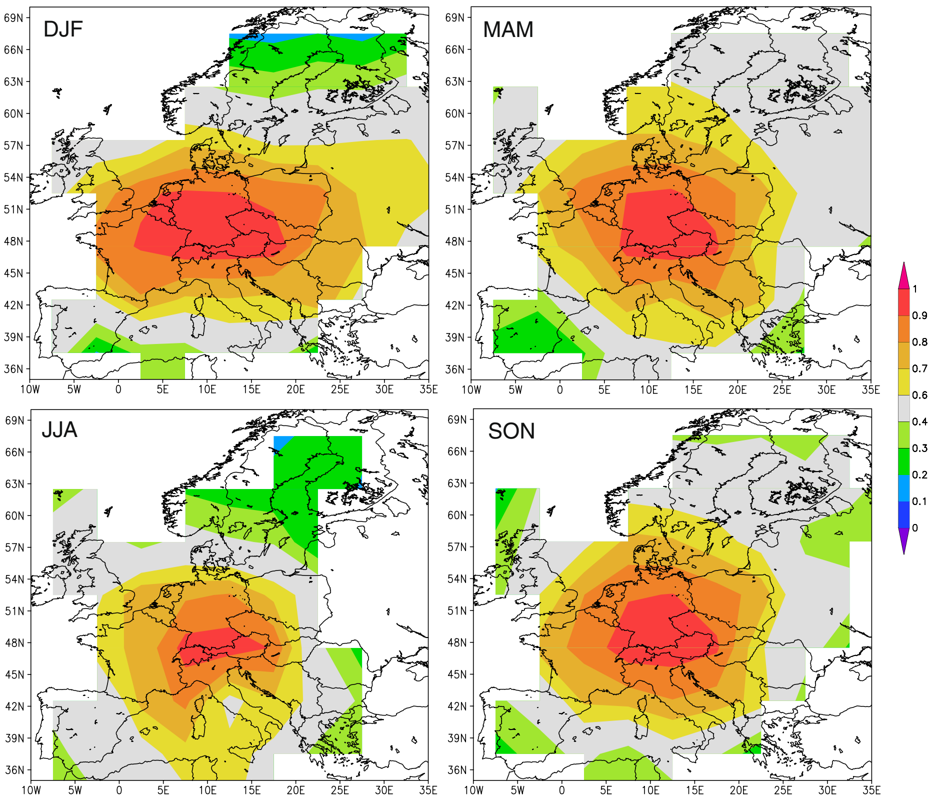


Fig. 5 Spatial correlations between seasonal average CEU instrumental series and HadCRUT3 5° × 5° gridded temperatures (Brohan et al. 2006) for period 1850–2007

4 Temperature reconstruction methods

In this section, we explain the different steps undertaken to derive calibrated monthly, seasonal and annual CEU temperatures and associated uncertainty estimates, based on a composite of the DE, CH and CZ documentary temperature index series. Inspired by working procedures that are well established in dendroclimatology (Cook and Kairiukstis 1990; Cook et al. 1994) and multi-proxy climate reconstructions from annually resolved paleoclimatic data (e.g. Rutherford et al. 2005; Luterbacher et al. 2004, 2007; Xoplaki et al. 2005; Guiot et al. 2005), but which are rarely used for documentary evidence, we explore how to use the information from temporally overlapping instrumental and documentary data and also parallel information in the different national documentary index series in the pre-instrumental period.

First, we provide a data quality overview. Figure 6 shows running correlations between all pairs of national index series, as well as the corresponding values for instrumental temperature data. The average of the nation-pair correlations (RBAR) is also shown. A running window of length 45 years is chosen (similar to the length of our calibration period—see Section 4.2). However, in order not to loose too many correlation values due to incomplete index series, we allow up to 22 missing values in any particular 45-year windows.

For instrumental data, the running correlations are highest and most stable in the winter months, with RBAR being near 0.9 in the DJF season. Somewhat lower and less stable correlations are seen for the other seasons. Compared to instrumental series, running correlations between documentary index series are more variable and mostly weaker. In some periods, the correlations between the index series are nearly as strong as for the instrumental data (e.g. before around AD 1620 in JJA), but there are also cases when the correlations break down (e.g. near the end of the seventeenth century for April). The strongest correlations among the index series are found (in agreement with instrumental data) for the winter months, with RBAR values between 0.6 and 0.8. Although correlations are weaker for other seasons, the RBAR values rarely fall below the level of a one-tailed significance test at the 0.025 level ($r = 0.29$ for $n = 45$), clearly indicating a common signal in the index data in all months and seasons. The most long-lasting case is the first two centuries in October, when RBAR values are constantly near or below the significance level. The weaker correlations for April and October, as typical transition months, can be partly related to the higher number of missing indices compared to winter or summer months (Fig. 2).

4.1 Derivation of variance-adjusted, unitless, composite CEU series from index data

We derive a unitless average of the three national index series, separately for each month and season, for the period 1500–1854. Prior to averaging, each individual index series is standardized by subtracting its mean and dividing by its standard deviation, which are calculated for the years when all three series of DE, CH and CZ have data in a given month or season. Due to the incompleteness of all three index series, the number of available index series varies with time (see Fig. 1). It is therefore

Fig. 6 Running correlations (45 year window) between the German, Swiss and Czech Republic temperature indices and between the corresponding instrumental records. The average of the nation-pair correlations (*RBAR*) is also shown. The 0.025 level ($r \sim 0.29$) for a one-sided significance test is shown with a dashed horizontal line

not appropriate to use a simple arithmetic average to represent the temperature variations. This is because the variance of an average of three series is less than the variance of an average of two series, which in turn is less than the variance of one single series. The situation is similar to that in dendroclimatology where the number of individual trees in a chronology varies with time. Dendroclimatologists generally apply a variance correction to their averaged tree-ring chronologies defined by the following equation (Osborn et al. 1997):

$$Y(t) = X(t)\sqrt{\frac{n(t)}{1 + (n(t) - 1)\bar{r}}} \tag{1}$$

where $Y(t)$ is the adjusted mean value at time t. $X(t)$ is the mean value at time t, $n(t)$ is the number of series at time t, and $\bar{r}$ is the average inter-series correlation between all pairs of time series (RBAR). The theoretical basis for this adjustment is the same as for the case of instrumental temperature data (Section 3). As a slight refinement to Osborn et al. (1997), following suggestions outlined in Frank et al. (2007b), the RBAR values are allowed to change with time (see also Leijonhufvud et al. 2009). To achieve this we use the running 45-year RBAR-values (Fig. 6). As we need RBAR values at the beginning and end of the time series, we 'padded' these parts of the series with the average of the 45 first and last available RBAR values, respectively. Furthermore, to avoid the irregular year-to-year variations due to the finiteness of the running window, we applied a Gaussian low-pass filter (Alexandersson and Eriksson 1989; van Vliet et al. 1998) with its standard deviation (σ) set to 9. This filter is frequently used in climatological applications to obtain a smoothing that approximately corresponds to 30-year running averages (e.g. Tuomenvirta et al. 2000). After having obtained the smoothed RBAR series, a variance adjusted CEU series was calculated for each month and each season using the equation above. These time series can be regarded as unitless temperature reconstructions, which in the following subsection are calibrated against instrumental data.

4.2 Calibration and verification

To perform calibration of the unitless reconstruction to temperature anomalies from the 1961–1990 average, simple linear regression was applied using the instrumental CEU temperature series as the predictand and the unitless reconstruction as predictor over a calibration period of data overlap. Verification of the calibrated data was then made over two other periods of overlapping data.

Unfortunately, it is impossible to find calibration and verification periods in which all three national instrumental and index series have data. In particular, the DE index series ends in 1760, which is the starting year of our CEU instrumental record. Moreover, only the CZ index series is available after 1816. As a compromise, we chose the 46-year period 1771–1816 for calibration. Three of the four national instrumental series are complete after 1771 (DE data are available from 1773 onwards), while 1816 marks the end of the CH index series. Therefore, only CZ and CH data are used for the calibration.

The verification was performed over one longer period (1817–1854), when only CZ index data are available, and a shorter period (1760–1770) when both CZ and CH index data exist. Unfortunately, it is not possible to validate the CEU temperature reconstruction directly against the instrumental CEU temperatures in any period where DE index data are included. Validation that includes DE data is only made implicitly (in Section 4.3) by means of correlations between the national index series.

Table 3 summarizes the calibration and verification statistics, which are commonly used in dendroclimatology but also applied to other annually resolved proxy data (e.g. Cook et al. 1994; Wilson et al. 2006). The squared correlation (r^2), the standard error of estimate (SE) and the Durbin–Watson (DW) test diagnose the calibration, whereas r^2, the Reduction of Error (RE), the Coefficient of Efficiency (CE) and the Root Mean Square Error (RMSE) diagnose the verification.

The r^2 quantifies the amount of temperature variance explained by the reconstruction, while SE measures the uncertainty in °C. The DW tests the first-order

Table 3 Summary of calibration (1771–1816) and verification (verification 1—1817–1854; verification 2—1760–1770) statistics

Month (season)	Calibration			Verification 1				Verification 2			
	r^2	SE	DW	r^2	RE	CE	RMSE	r^2	RE	CE	RMSE
J	0.86	1.09	2.26	0.81	0.79	0.79	1.33	0.91	0.87	0.85	1.21
F	0.74	1.11	1.50	0.62	0.54	0.52	1.68	0.57	0.54	0.54	1.28
M	0.79	0.99	1.86	0.50	0.49	0.48	1.48	0.44	0.41	0.21	1.34
A	0.74	1.10	1.79	0.27	−0.05	−0.08	1.88	0.78	0.62	0.61	1.02
M	0.65	0.97	1.74	0.65	0.66	0.60	1.07	0.31	0.11	−0.10	1.13
J	0.58	0.78	2.04	0.44	0.44	0.44	1.05	0.34	0.26	0.25	0.67
J	0.77	0.69	1.82	0.51	0.40	0.38	1.02	0.62	0.59	0.56	0.67
A	0.75	0.69	1.23	0.56	0.55	0.44	1.10	0.08	0.05	−0.14	1.03
S	0.56	0.85	1.23	0.56	0.29	0.16	1.09	0.62	0.50	0.40	0.71
O	0.62	0.97	1.51	0.19	−0.25	−0.26	1.51	0.76	0.75	0.63	0.86
N	0.68	0.88	1.75	0.44	0.45	0.44	1.41	0.77	0.74	0.74	0.63
D	0.72	1.41	2.00	0.61	0.62	0.60	1.75	0.87	0.85	0.84	0.85
DJF	0.83	0.69	2.23	0.86	0.81	0.81	0.82	0.92	0.86	0.84	0.67
MAM	0.80	0.58	1.78	0.65	0.62	0.57	0.86	0.74	0.74	0.61	0.50
JJA	0.77	0.49	1.02	0.64	0.58	0.53	0.64	0.30	−0.33	−0.34	0.44
SON	0.73	0.55	1.51	0.43	0.25	0.25	0.84	0.74	0.67	0.59	0.57
Ann	0.81	0.37	1.22	0.62	0.53	0.48	0.54	0.53	0.55	−0.52	0.34

See text for explanation of the measures

autocorrelation within the regression residuals (von Storch and Zwiers 1999). Critical values of DW depend on the number of independent variables and also on the time series length, but values between 1.5 and 2.5 (with an ideal target of 2.0) are generally acceptable. DW values outside this range indicate problems with reconstructing multi-decadal variations.

The RE statistic compares the Mean Square Error (MSE) of the reconstruction to the MSE of a ‘reconstruction’ that is constant in time with a value equal to the mean value for the measured temperatures in the calibration period. The CE instead compares the MSE of the reconstruction to a ‘reconstruction’ that is constant and equal to the mean value of the measured temperatures in the validation period. Both RE and CE can take values between 1 and negative infinity. CE is always less than, or equal to, RE. For both measures, positive values indicate that the linear regression model has some potential for reconstruction skill. CE provides a more rigorous test for the data than RE. Definitions and further discussion on how to interpret RE and CE can be found elsewhere (e.g. Cook et al. 1994; Rutherford et al. 2005; Wilson et al. 2006; Wahl and Ammann 2007).

As can be seen from the r^2 values (Table 3), the documentary evidence can explain a large fraction of temperature variability. In the calibration period, the seasonal r^2 values vary from 73% for SON to 83% for DJF, whereas it is 81% for the annual mean. In individual months, r^2 vary between 56% (September) and 86% (January). The DW statistics are within the acceptable range (1.5–2.5) in most cases, but not for August, September, JJA and annual series. This suggests problems with the capability of the documentary data to reliably portray longer-term temperature trends in the summer season and this appears to affect also the annual means.

The RE and CE scores for the longer verification period (1817–1854) are positive for all four seasons and the annual mean, with values up to 0.81 for both RE and

CE in DJF. Also MAM and JJA have rather high RE and CE scores (approximately 0.5 to 0.6), whereas SON has the weakest score (0.25). Values for individual months are generally weaker than those for the seasonal averages, with even negative RE and CE in April and October, indicating that the monthly reconstructions are less reliable than the seasonal ones. Nevertheless, the overall strong verification statistics in the longer verification period demonstrate that CZ national index series alone is able to capture a significant fraction of the CEU average temperature variability in all seasons. Verification results for the shorter period (1760–1770; with index data for both CZ and CH) are also generally quite strong, although negative RE and CE values are found for JJA. Moreover, CE for the annual mean is negative despite the rather high RE (0.55) and r^2 (0.53). This is yet another indication of potential problems with reliability of long-term trends in the warm season and that this can affect the annual mean trends. Regarding individual months, May and August have the poorest verification statistics in the shorter period, whereas all other months have acceptable or even strong values.

A simple way to further illustrate the performance of the documentary data is to compare the reconstruction and the instrumental target data in the calibration and verification periods (Fig. 7). These time series confirm an overall strong capacity of the documentary data to portray the temperature variations, but also shed more light on the cases with rather poor calibration/verification statistics. For example, the negative CE in the annual mean data in the short verification period is due to the instrumental annual target data being slightly colder than the reconstruction in this short 11-year period.

4.3 Error estimation

From the calibration and verification exercise, we conclude that the documentary data are useful for reconstructing seasonal and monthly Central European temperatures, albeit with problems noted mainly for parts of the summer season. Here, we derive error estimates for the temperature reconstructions, which are illustrated as time series in Section 5.

The basis for our error estimation for the reconstruction is to simply use the Standard Error (SE) of the estimate, defined by the regression relationship between proxy and instrumental data in the calibration period. This is a standard approach. A drawback, however, is that this uncertainty estimate alone only represents the data properties in the calibration period and hence it is a simplification of the total error estimation problem. Notably, our calibration involves proxy data from only Switzerland and the Czech Republic, but not from Germany. Moreover, the calibration uncertainty does not account for any temporal changes in data quality. Therefore, we seek to account also for these additional factors.

Following ideas and methods discussed in Leijonhufvud et al. (2009), we argue that information about changes in data quality can be derived from the mean of the running inter-series correlations, i.e. the RBAR values used in Section 4.1. In periods when RBAR is weak, the reconstruction uncertainty is arguably larger than in a period with a strong RBAR. We design a method that accounts for these changes in uncertainty, where we multiply the calibration SE with time varying inflation factors. Our approach to find these factors is to add noise to the proxy data and then perform the calibration again, repetitively for several levels of noise. With more

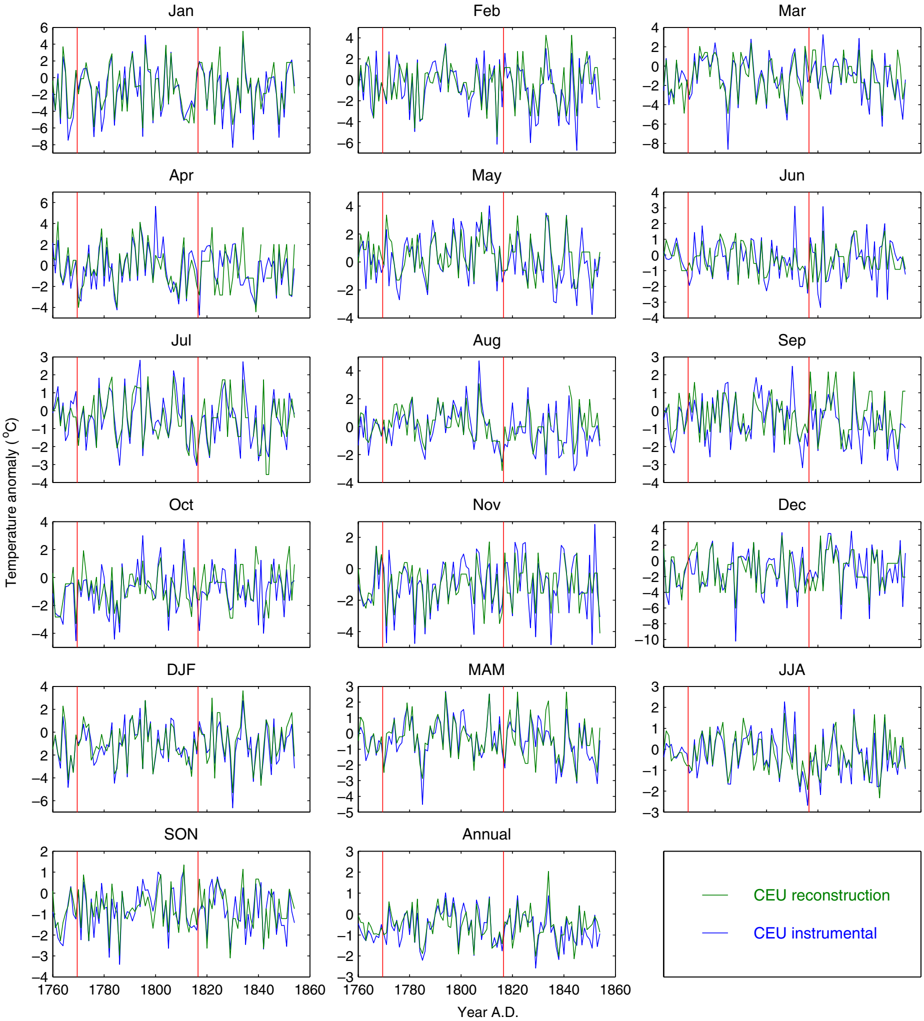


Fig. 7 Comparison of measured (*blue*) and reconstructed (*green*) monthly, seasonal and annual Central European temperatures during 1760–1854. The calibration period (1771–1816) and the two verification (1760–1770 and 1817–1854) periods are separated with vertical red lines

noise, the SE becomes larger up to a maximum level where the proxy data have zero correlation with the instrumental data. Hence, we can derive a numeric relationship between data quality and error inflation. Below we describe the procedure in detail, which also accounts for the circumstance that no German data were included in the calibration.

Leijonhufvud et al. (2009) hypothesize that the RBAR values provide an estimate of the common signal among the series and that the strength, or weakness, of this signal can be used to derive an approximation of the error in the composite. A basis for these ideas is the concept of the Expressed Population Signal (EPS), which was first described by Wigley et al. (1984). These authors were interested in finding an estimate of how strongly an average of a sample of correlated time series correlates with the theoretical population of associated time series. They demonstrated that this theoretical correlation is dependent on the average correlation between the time series pairs (i.e. RBAR) and the number of individual series (n) contributing to the average, according to the following equation:

$$\mathrm{EPS}\,(t) = \frac{n\,(t)\,\bar{r}}{n\,(t)\,\bar{r} + (1 - \bar{r})} \approx \frac{\text{signal}}{\text{total variance}} \tag{2}$$

The approach of Leijonhufvud et al. (2009) is to calculate the EPS-value for the proxy data in the calibration period, and then add noise to the calibration proxy data and repetitively calculating the EPS-value again for several levels of added noise. Thus, they estimate how SE changes as a function of EPS. We follow this idea, but as there are no DE index data in the calibration period, we first had to construct German surrogate proxy data. This was done by adding white noise to the DE instrumental data, such that the resulting surrogate proxy data have the same signal-to-noise ratio (SNR) relative to the original DE instrumental data as the average SNR for the CH and CZ proxy data. Then we obtained an estimate of the calibration EPS and SE for the hypothetical case where the reconstruction is based on index data from all three countries. To obtain stable results, we added 2000 noise realizations to the DE instrumental data. The final estimate of the hypothetical calibration EPS and SE was taken as the average of all cases.

Next, we repetitively estimated hypothetical calibration EPS and SE values again after adding different levels of noise to all three national proxy series. For each noise level, we estimated EPS and SE by taking the average of 2000 cases. Figure 8 shows (blue circles) how the ratio between the SE for the noise-added data and the original calibration SE (obtained from CZ+CH index data and represented by the red dot) increases with decreasing EPS. Results are only shown for the seasons, but calculations were also performed for monthly and annual data. The blue circle at EPS = 0 is obtained in a different way; here we calculate SE for a 'reconstruction' that consists of a constant equal to the mean of the instrumental calibration data. This theoretically corresponds to the case when the correlation between the predictor and predictand is zero, and provides the maximum possible error as far as only the regression relationship is concerned.

The ratio between the hypothetical estimated SE and the real calibration SE is henceforth called the SE inflation factor (SE_inflation). We can obtain an estimate of the reconstruction error at any time point by calculating EPS (derived from the smoothed RBAR series described in Section 4.1), and multiplying the calibration SE with the corresponding SE_inflation found by interpolation between the points in Fig. 8.

Figure 9 shows how the monthly and seasonal EPS values vary with time. The slowly varying component of the EPS series is due to variations in RBAR, whereas the year-to-year variations reflect the changing number of contributing index series. DJF has the strongest and most stable signal, mostly above 0.8. Despite the calibration/verification problems reported above, JJA has the second strongest EPS-values, with pre-1650 values being at about the same level as for DJF. EPS values for MAM and SON are often weaker and show rather strong low-frequency variability. The previously noted breakdown of RBAR in April in the late 1600s is clearly seen also in the EPS series. This is a good example of a period where the error bars should

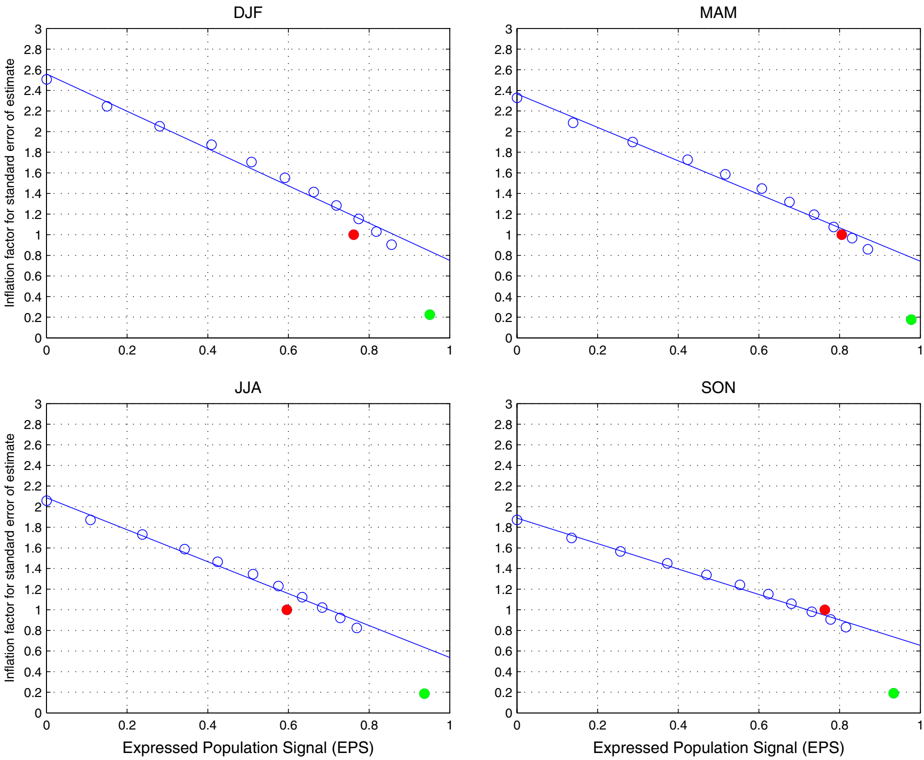


Fig. 8 Estimated relationship (*blue circles* and *fitted straight line*) between the standard error (SE) of the reconstruction and the expressed population signal (EPS), expressed as the ratio (SE_inflation) between the estimated SE and the original calibration SE. Red circles correspond to the original calibration period statistics. Green circles are obtained using instrumental temperature data as predictor

be wider than in the calibration period, where the EPS is quite strong. To interpolate between the pairs of EPS values and SE_inflation, we follow Leijonhufvud et al. (2009) and fit a straight line as shown in Fig. 8; thus we derive a time series of SE_inflation factors for each year in the reconstruction. Before the SE time series was multiplied with the SE_inflation series, however, we smoothed the SE_inflation series with a Gaussian low-pass filter ($\sigma = 9$) to get rid of the year-to-year variations inherited from the EPS calculations.

As a last step in the error estimation procedure, one final restriction was imposed. We argue that one should not have stronger confidence in any earlier proxy data than those in the calibration period. It is obvious that we cannot strictly verified if any pre-instrumental proxy data are better than those in the calibration period. Hence, we never deflate any SE values—we only adjust them in periods when they should be inflated.

From a practical point of view, it seems that only two pairs of EPS/SE_inflation values are needed to establish the linear relationship; namely the case representing the original calibration (the red dots in Fig. 8) and the case where EPS $= 0$. This would simplify the procedure designed above, as no noise modelling would be needed, and could perhaps be considered by potential users of our error inflation

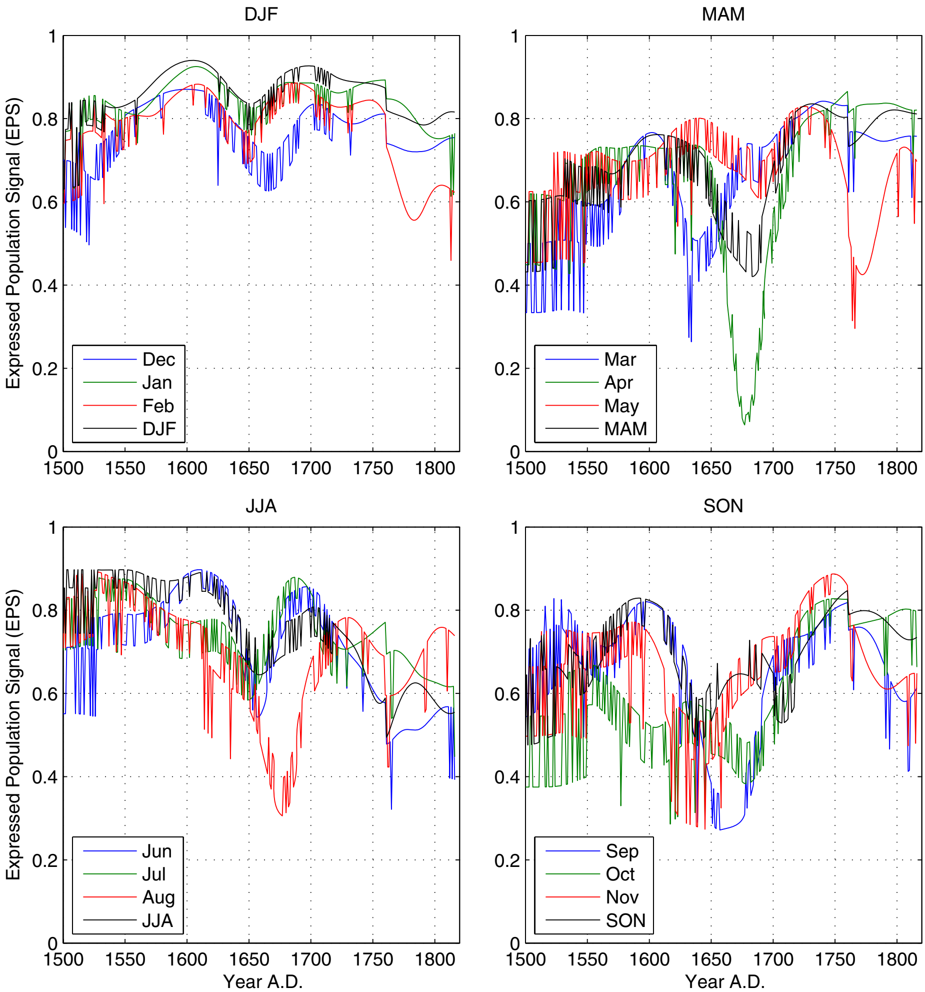


Fig. 9 Estimated Expressed Population Signal (EPS) for each month and season over the 1500–1816 period

approach. The fit of such a straight line between EPS and SE_inflation pairs is a reasonable approximation as long as we are not interested in extrapolation to stronger EPS-values than in the calibration period. For very strong EPS-values, however, the linear approximation may not be appropriate. In theory, one may argue that if the inter-series correlation among the index series is equal to 1, then they could perfectly portray the temperature variations and hence there would be no error in the temperature estimation. This unrealistic case would correspond to EPS = 1 and SE_inflation = 0. In Fig. 8, the green circle is inserted to represent a case where the average of the instrumental DE, CH and CZ series is used as the predictor. It can be seen that this point lies approximately on a curve that connects the blue circles with the theoretical point (EPS = 1, SE_inflation = 0). Hence, a curve of some kind fitted to the data seems theoretically more correct than a straight line, but for our purpose a straight line is a reasonable approximation.

5 The Central European temperature reconstructions

The new Central European temperature reconstructions are presented here as time series with error bars, providing a view of the recent climate in the context of the last five centuries.

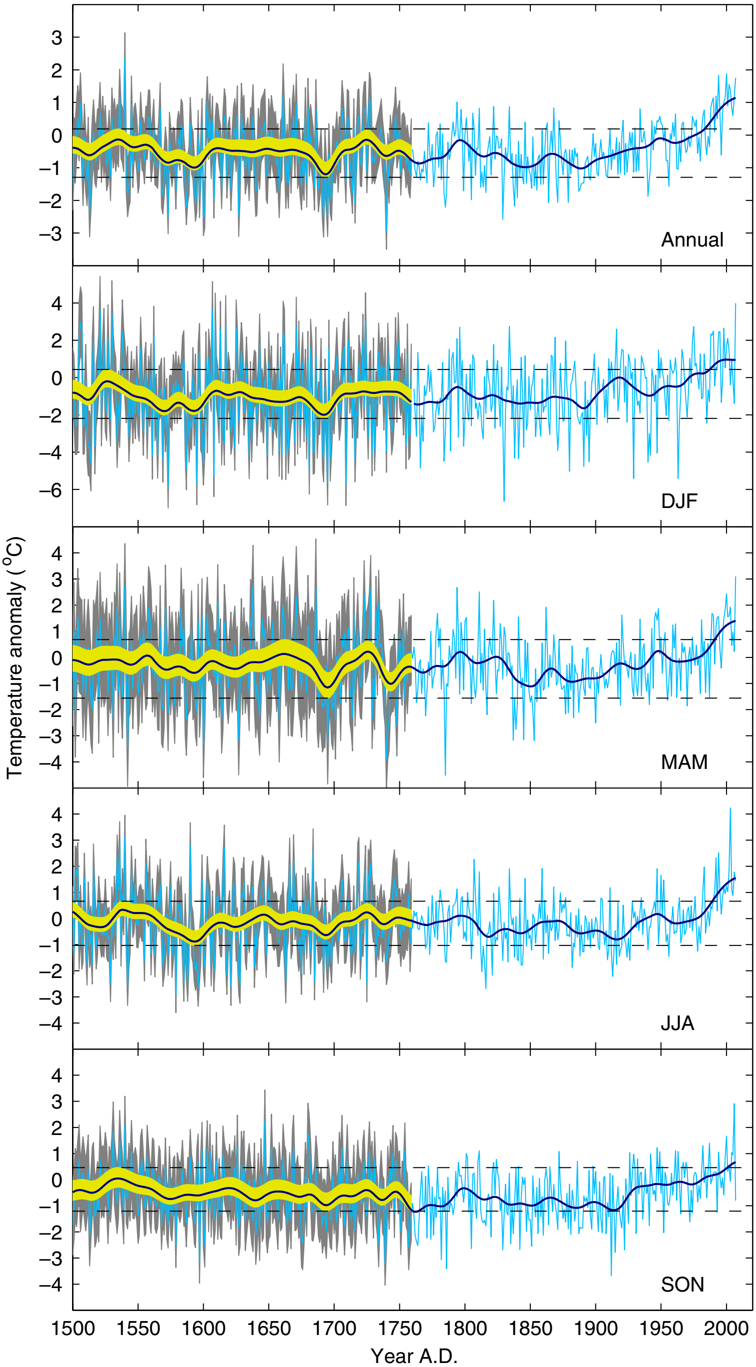
Figures 10 and 11 show annual, seasonal and monthly series, where the reconstructions (1500–1759) are spliced onto the instrumental data (1760–2007). Here, the variance and mean level of the proxy data have been adjusted to agree with the instrumental data (CEU.instr) in the full period of data overlap (1760–1854). This is done to avoid the artificial reduction of variance due to linear regression calibration (Esper et al. 2005). The disadvantage of using the variance-and-mean adjusted (CEU.scal) data instead of those obtained from regression (CEU.regr) is that the errors are not minimized in a least squares sense. There is thus a trade-off between minimizing prediction errors and conservation of variance. However, as the correlation between the proxy and instrumental data is very strong, the differences between CEU.regr and CEU.scal are small (see also Section 6.1).

The reconstruction uncertainty derived in Section 4.3 is illustrated in Figs. 10 and 11 as grey error bands defined as CEU.regr ± 2 SE (adjusted with the SE_inflation factor), which provide approximate 95% confidence intervals for past temperatures, accounting for calibration uncertainty and temporally changing data quality in the pre-instrumental period. They can, however, not account for all types of errors, such as for instance frequency dependent deficiencies in the instrumental or documentary evidence index data. One such type of error is related to the original assignment of index values in each particular month, e.g., index +1 or 0. This type of error is not explicitly accounted for here, but is to some extent included implicitly through the calibration regression relationship and the variable error inflation factors.

To highlight low-frequency trends, Figs. 10 and 11 include smoothed time series and uncertainty is shown also for the smoothed data as yellow bands defined as smoothed CEU.regr data ± 2 SE. The width of these error bands has been adapted according to the degree of smoothing (by decreasing SE), following Briffa et al. (2002) and Gouirand et al. (2008). By looking at how much the smoothed CEU.scal series (thick blue curves) deviate from the middle of the yellow bands, one can visualize how CEU.scal differs from CEU.regr.

In all seasons, the smoothed CEU temperature series have their highest values in the most recent years, highlighting the unusualness of the current warming in a five century context. This is in good agreement with previous European temperature reconstructions, e.g. Luterbacher et al. (2004, 2007) and Xoplaki et al. (2005). To further illustrate this, we plot the upper and lower bounds of the smoothed 2 SE error bands as horizontal black lines in Figs. 10 and 11. The range between these two lines can be viewed as the pre-instrumental range of 20-year smoothed temperatures, including the estimated data uncertainty.

Fig. 10 Spliced Central European temperature reconstructions AD 1500–2007 (based on variance and mean adjusted proxy data 1500–1759, instrumental data 1760–2007), expressed as anomalies from the 1961–1990 average, with Gaussian low-pass filtered ($\sigma = 6$) data approximately corresponding to 20-year moving averages. The error bands are approximate 95% confidence intervals. Dashed horizontal lines show the highest and lowest level for the low-frequency error bars ▶



Temperature anomaly (°C)
Year A.D.
Annual
DJF
MAM
JJA
SON
1500
1550
1600
1650
1700
1750
1800
1850
1900
1950
2000

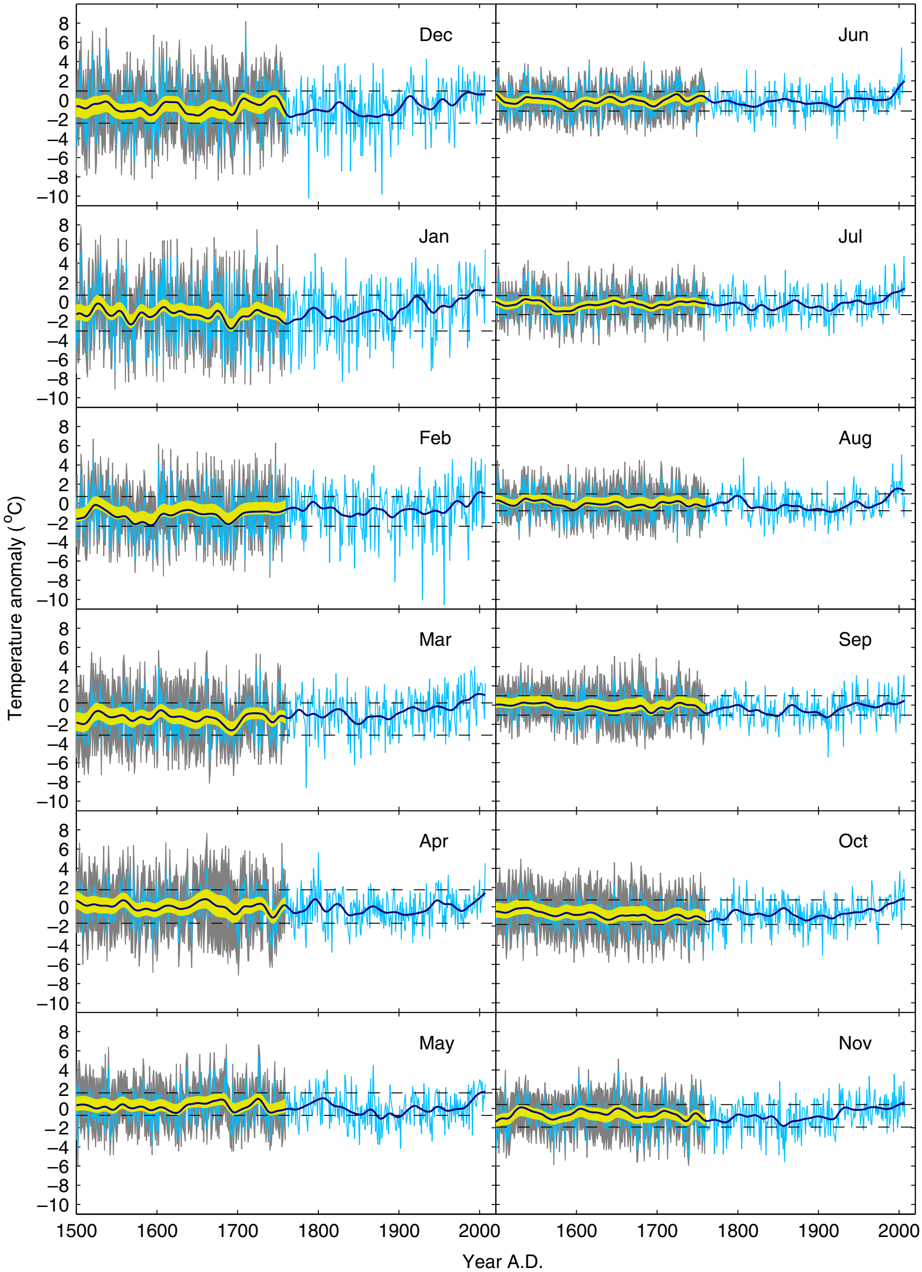


Fig. 11 Monthly temperature reconstructions for Central Europe. Graphical presentation as in Fig. 10

The smoothed annual temperatures rose above the upper level in the late 1980s and have since then remained at a high level. The end points of the smoothed curves can not be directly compared to values in the middle of the series because smoothed values at the end are derived only from past values. However, even the unsmoothed data highlight the recent warming; 18 of the last 20 years in the unsmoothed temperature series lie above the upper bound of the 2 SE error for pre-instrumental smoothed data. The same 18 years also lie above the upper bound of smoothed CEU.scal ±2.7 SE (not shown), approximately corresponding to a 99% confidence interval of 20-year mean temperatures. Among the seasons, the recent warming is most outstanding in JJA (15 of the last 20 years lie above the 2.7 SE upper error bound for smoothed data), followed by MAM (13 years), DJF (11 years) and SON (6 years). The recent warming is less easily detected in the monthly data, due to their larger variance than seasonal averages. Nevertheless, 6 months (January, February, March, May, July, and August) have at least 11 of the last 20 years lying above the 2.7 SE upper bounds for smoothed data.

Temperatures of the widely discussed warm summer of 2003 (e.g. Luterbacher et al. 2004; Schär et al. 2004) exceed the pre-instrumental unsmoothed CEU.regr + 2 SE values in June, August and JJA. Furthermore, July of 2006 was warmer than the +2 SE level for unsmoothed data. Apart from these few values, however, no recent individual monthly or seasonal temperatures have risen above the +2 SE error level for unsmoothed pre-instrumental temperatures.

The coldest periods within the last five centuries seem to have occurred in the pre-instrumental period. The smoothed annual, DJF and MAM data have their minima in the 1690s in agreement with recent analysis for the whole of Europe (Luterbacher et al. 2004, 2007; Xoplaki et al. 2005). For SON temperatures, accounting for the −2 SE error, the minimum appears to be in the 1760s while for JJA it is in the 1590s. All four seasons, however, have cold smoothed temperatures around the 1690s and together contribute to the cold level of annual mean temperatures at this time. This corresponds well to the Maunder Minimum (AD 1645–1715) of weak solar activity (Eddy 1976), which is hypothesized to have caused colder temperatures in Central Europe and some other regions (e.g. Shindell et al. 2001).

One potential problem with interpreting the long-term trends and low-frequency variations in the new CEU temperature reconstruction is related to the method used to derive the underlying index data. The choice of indices involves somewhat subjective interpretation of the information from documentary evidence. The original weather-related descriptions in these sources were written by people who were influenced by their perception of climate conditions, which were determined by their personal experience of climate during their respective life times. It is thus a risk that, for example, a ‘very cold winter’ to an individual who lived in one period was not equally cold as another ‘very cold winter’ was to another individual who lived in a period when climate on average was different. This implies a risk that, in particular, low-frequency temperature variations are likely to be suppressed in amplitude. Further investigation of how serious this problem is would be an important task for future research.

6 Discussion

In the following discussion the main features of the new CEU temperature reconstruction are compared with existing reconstructions in the European context. Furthermore, we investigate the potential of the new index data for Poland and the Carpathian Basin for developing temperature reconstructions for these regions.

6.1 Comparison with other European temperature reconstructions

Comparison with appropriate other European temperature reconstructions reveals to what extent the new CEU reconstruction adds new information and illuminates properties of both the new and existing reconstructions. We compare our results with two available reconstructions:

1. Seasonal mean temperatures 1500–2002 averaged over 5–18° E, 45–53° N, from Xoplaki et al. (2005; spring and autumn months) and Luterbacher et al. (2004, winter and summer months). Hereafter referred to as LUT.
2. DJF and JJA mean temperatures for the Low Countries 1500–2000 (Shabalova and van Engelen 2003). Hereafter referred to as LC.

Figure 12 compares the CEU, LUT and LC seasonal temperatures, plotted as anomalies with respect to the 1961–1990 climatology. For CEU, we use the same spliced version as in Figs. 10 and 11. Overall, the three series show similar features, but also some notable differences. Running correlations (31-year windows) between CEU and LUT (Fig. 13) indicate the relationship between these series. Very strong correlations ($r > 0.9$) in all seasons after AD 1800 are hardly surprising, given that both series are derived from instrumental temperature data during that period. Similarly strong correlations before 1650 undoubtedly reveal that CEU and LUT are to a large extent derived from similar proxy data in this period, implying that early German and Swiss documentary data dominate the early period variability in the selected subset from Xoplaki et al. (2005). Somewhat weaker correlations between CEU and LUT are seen during 1650–1800, reflecting a larger degree of independency between the raw data; CEU is derived from documentary data before 1760 while LUT is increasingly dominated by instrumental data from 1659. In particular, the new CZ documentary data in CEU are not used in LUT.

A large drop in correlation between CEU and LUT in summer around 1750 has a counterpart in the correlation between CEU and LC and also (not shown) between CEU and the Central England temperature series (Manley 1974). This reflects the fact that CEU and LUT are essentially derived from different data around this period (proxy vs. instrumental), and also possibly suggests a period when CEU and/or LUT summer data appear to be less reliable. We also show running correlations with LUT where CEU proxy data are used during 1760–1854 (dashed lines in Fig. 13). These correlations are not as weak as those for JJA around 1750, but clearly weaker than those when instrumental CEU data are used 1760–1854, as might be expected from the calibration and verification statistics in Section 4.

Correlations between CEU and LC are constantly weaker than those between CEU and LUT, which is related to the fact that LC represents a region that is distinct from CEU. Nevertheless, the correlations are always significant at the 0.025 level for a one-sided test, apart from around 1750 in the summer. Correlations between LUT and LC are mostly above 0.9 after around 1700 in DJF, but decreases somewhat in the two first centuries, i.e. in the proxy data part of the two series. LUT vs. LC correlations in JJA fluctuate between around 0.6 and 0.9 throughout the records.

Given the overall reasonable or very strong correlations between the three series, it is not surprising that the overall fluctuations in the time series (Fig. 12) are quite similar. The perhaps most notable difference between CEU and the two other series is that CEU is colder in much of the instrumental period before the 1961–1990

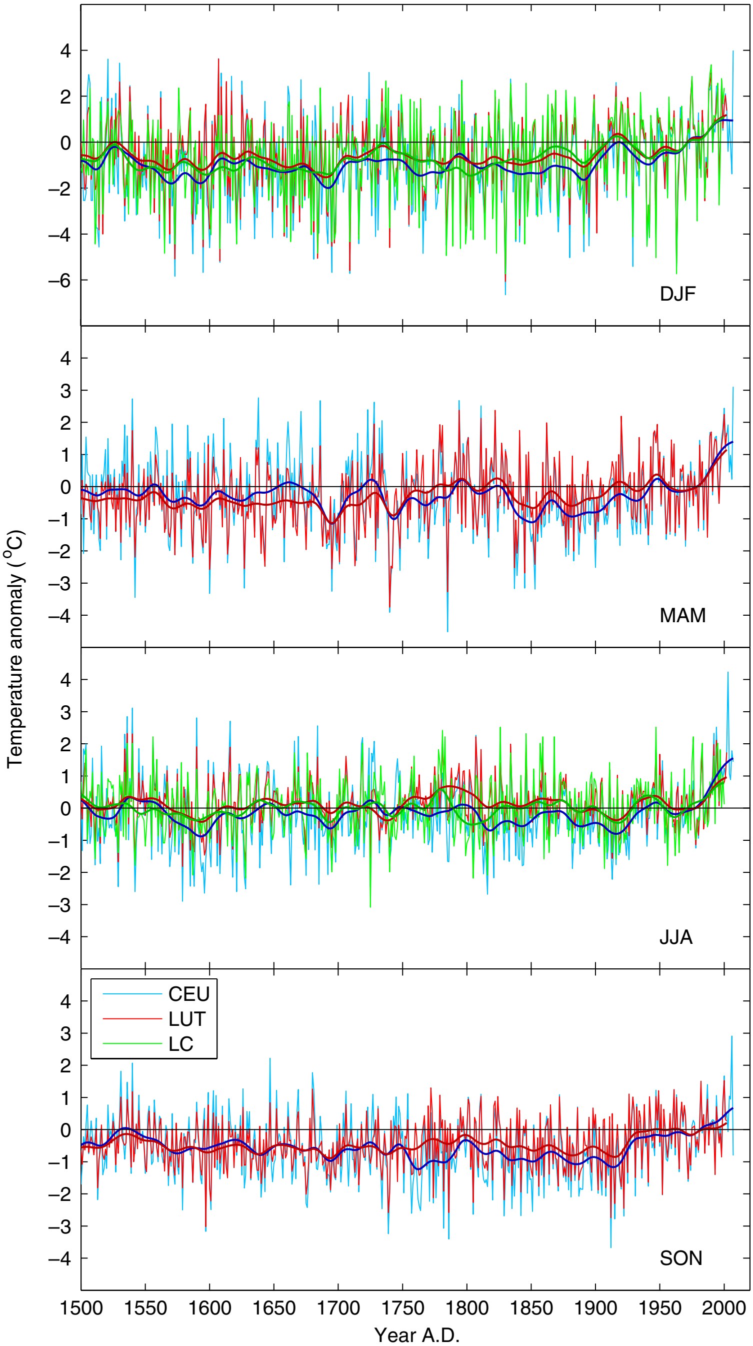


Fig. 12 Comparison of seasonal European temperature reconstructions (anomalies from 1961–1990 average). *CEU* spliced reconstruction as in Fig. 10, *LUT* subset 5–18°E, 45–53°N from Luterbacher et al. (2004) and Xoplaki et al. (2005), *LC* Low Countries (van Engelen et al. 2000). A Gaussian low-pass filter ($\sigma = 6$) is applied

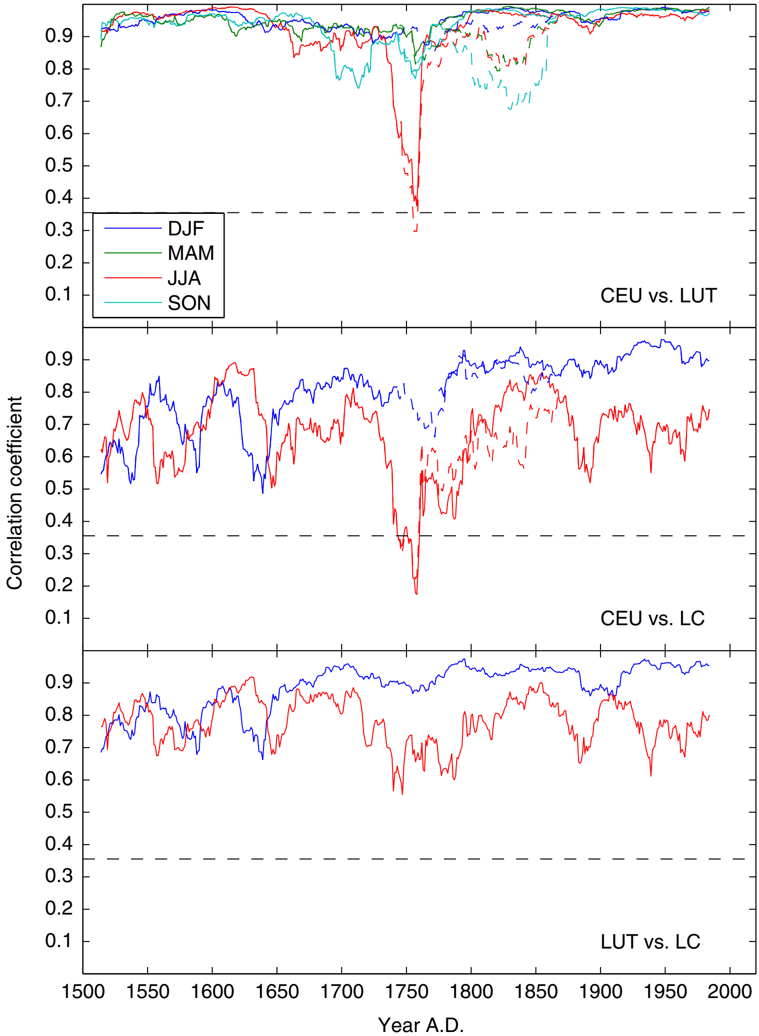


Fig. 13 Running Pearson correlations (31-year windows) between seasonal European temperature reconstructions. LUT and LC defined as in Fig. 12. *Dashed* (*solid*) *curves* in upper and middle subplot represent proxy (instrumental) CEU data. *Horizontal dashed black lines* indicate the level of a one-tailed significance test at the 0.025 level ($r \sim 0.36$)

reference period. This is undoubtedly a result of differences between the homogenized temperature records used in CEU compared to those used in LUT and LC. In particular, temperatures in the early part of the instrumental period are colder in CEU than in LUT due to the EI-bias correction in CEU data. For example, the smoothed JJA temperatures in LUT are warmer than the 1961–1990 average during around 1750–1870, whereas those for CEU are mostly colder than the reference period. Consequently, the long-term warming trend since the nineteenth century is more pronounced in CEU compared to LUT (and LC), not only in summer but also in all seasons.

In the pre-instrumental period, i.e. before around 1760, the low-frequency components of CEU and LUT are very similar in DJF, JJA and SON, while MAM temperatures are consistently colder in LUT. The amplitude of interannual variations, however, is smaller in the early part of LUT. This difference in reconstructed temperature variance is due to the different calibration methods. The amplitude and mean level in CEU is determined by the variance matching and mean adjustment derived from the direct overlap between documentary and instrumental data (1760–1854). In LUT, the mean and variance of reconstructed temperatures is instead determined by multivariate regression relationships, obtained between the first few principal components of proxy data (predictors) and gridded instrumental data (predictand) in the twentieth century. Moreover, Luterbacher et al. (2004) and Xoplaki et al. (2005) used a nesting approach with successively fewer predictors back in time. Consequently, their proxy data networks tend to explain less variance in the early part of their reconstruction. Hence, as they used a regression calibration, the resulting temperature reconstructions artificially lose variance, most notably in the early period. Moreover, as their calibration methods were designed for spatial field reconstruction, with an emphasis on reconstructing the large scale patterns, they are not optimized for individual grid points.

Running standard deviations in the CEU and LUT series highlight the effect of the different calibration approaches on the variability in the time series. Figure 14 shows running standard deviations for 31-year windows, comparing LUT (and LC) data with the CEU series—separately for instrumental data (CEU.inst), the regression based reconstruction (CEU.regr) and the variance scaled reconstruction (CEU.scal). CEU.scal and CEU.instr have on average the same variance in their period of overlap (1760–1854). The running correlations, however, show different temporal patterns for CEU.scal and CEU.instr within the overlapping period, indicating that details in variance changes observed in the documentary data should be interpreted cautiously.

Differences in variance between CEU.scal and CEU.regr are very small in DJF and MAM but more clearly seen in JJA and SON. Overall, however, this difference is small and does not affect the observed long-term trends in running standard deviations. Moreover, the CEU.scal (or CEU.regr) and CEU.instr series show a long term reduction in variance in DJF and JJA over the past five centuries, with strong multi-decadal to centennial fluctuations superimposed. MAM rather shows a relatively stable variability before 1850, followed by an overall decrease in variance with some fluctuations. SON data do not show any clear long-term trend in variance, but rather fluctuations around a constant mean level. Zooming in on the twentieth century, the trends in variance are mostly negative in DJF and SON, but increasing in MAM and JJA (although not to the high general level before 1850). To summarize, the CEU temperature reconstruction displays an overall decrease in variability over the past five centuries, for all seasons except SON.

In contrast, LUT does not show any long-term decrease in variance in MAM and JJA, but only in DJF data from around AD 1600. Notably, LUT shows substantially less variance in the proxy data period in all seasons. This observation holds whether comparing to CEU.regr or CEU.scal. Interestingly, the instrumental part of LUT has somewhat less variance than CEU.instr in all seasons, but most pronouncedly in JJA

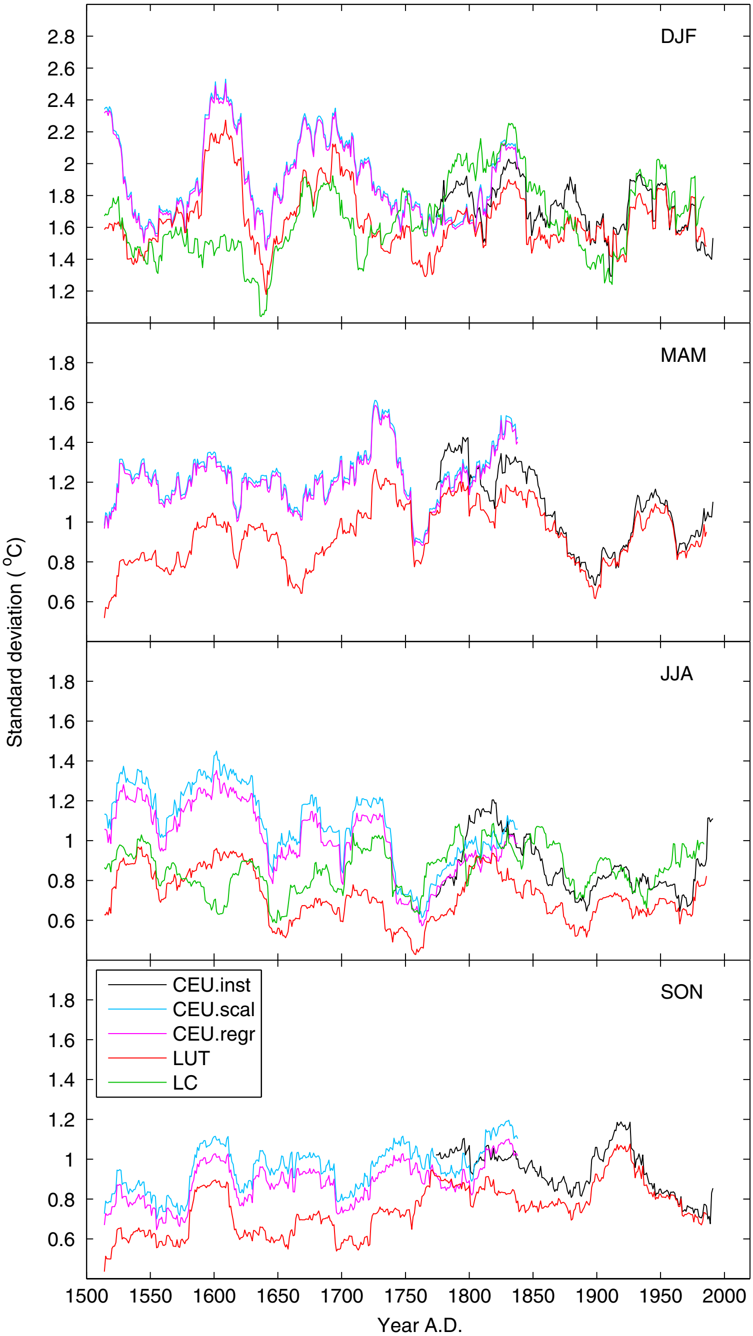


Fig. 14 Running standard deviations (31-year windows) for seasonal European Temperature Reconstructions. *CEU.instr* instrumental data. *CEU.scal* variance adjusted reconstruction. *CEU.regr* regression based reconstruction. LUT and LC as defined in Fig. 12

and SON. This is probably partly because LUT integrates temperature variability over a larger region than CEU.instr, and hence, to a larger degree, spatially smoothes local temperature variations, but partly also because the instrumental portion of LUT before 1900 is affected by some artificial variance reduction due to regression calibration. Running standard deviations in LC shows an increasing rather than decreasing long-term trend in DJF, suggesting an even larger artificial suppression of variance in early proxy data compared to LUT. In JJA, LC shows no clear long-term trend in variance.

6.2 Comparison with temperature index series from Poland and the Carpathian Basin

In Section 2 we described temperature index data derived from documentary evidence from Poland and the Carpathian Basin (Hungary, Slovakia, Western Romania), which were not included in the CEU temperature reconstruction. Here, we calculate Pearson correlation coefficients between each of these two index series and the CEU temperature series. The calculations are split into two periods, before and after 1760. Before 1760 we use the proxy data based on CEU reconstruction but after 1760 we use the instrumental data. No documentary PL data are available after 1760. Results are shown for all months and seasons in Table 4, which also gives the number of years with data and the significance level for the correlations. These are based on one-tailed test as we are only expecting positive correlations. Due to

Table 4 Pearson correlation coefficients (r) between the CEU temperature reconstruction and temperature indices from Poland (PL) and the Carpathian Basin (CB) before 1760 and after 1760

Month (season)	PL pre-1760			CB pre-1760			CB post-1760		
	r	n	Significance levels of correlation	r	n	Significance levels of correlation	r	n	Significance levels of correlation
J	0.33	73	2	0.62	121	3	0.75	105	3
F	0.38	65	2	0.53	109	3	0.57	104	3
M	0.14	61	0	0.41	78	3	0.50	97	3
A	0.24	65	0	0.32	70	2	0.69	90	3
M	0.13	55	0	0.49	60	3	0.62	90	3
J	0.26	60	1	0.33	79	2	0.41	85	3
J	0.18	64	0	0.49	80	3	0.49	92	3
A	0.15	59	0	0.46	77	3	0.53	91	3
S	0.46	51	3	0.28	61	1	0.38	77	3
O	0.21	51	0	0.30	61	1	0.34	76	2
N	0.22	59	0	0.13	68	0	0.47	79	3
D	0.23	65	0	0.44	106	3	0.62	97	3
DJF	0.42	73	3	0.59	91	3	0.73	92	3
MAM	0.32	46	1	0.48	52	3	0.70	80	3
JJA	0.28	183	3	0.54	66	3	0.57	78	3
SON	0.28	42	0	0.21	47	0	0.40	58	3
Ann	0.26	135	2	0.67	32	3	0.68	62	3

The number of years with data (n) and the significance levels of correlation are indicated: *0* not significant at 0.025 level, *1* significant at 0.025 level, *2* significant at 0.01 level, *3* significant at 0.001 level. One sided significance tests are used the mixture of nominal index data and continuous instrumental data, the significance tests are approximate but nevertheless informative.

CB index series correlate, with the exception of September, November and SON, better than PL with CEU proxy data before 1760. Moreover, CB data show significant correlations with CEU in this period in all months and seasons except November and SON. When CB is compared to the instrumental CEU data after 1760, correlations systematically increase and the 0.001 significance level is reached in all months and seasons except October (but the 0.01 level is reached). This clearly indicates a strong potential of using documentary evidence from the CB region as a temperature proxy.

There is potential also for the PL data, but they cannot be conclusively judged from this simple comparison. In particular, the lack of PL documentary evidence in the post-1760 period in this study makes its evaluation difficult. One reason why PL documentary evidence shows weaker correlations with CEU, could be related to real geographic climatological differences between the PL and CEU regions. An overall knowledge of temperature variability in Central Europe can be extended to other studies that used documentary indices defined using different principles. As for Poland, Przybylak et al. (2005) analyzed temperature indices from documentary sources for 1501–1840. Their reconstruction, however, was not developed at a monthly resolution but only for decadal means of winter (DJF) and summer (JJA) temperatures and hence they could not be included in our study. Nevertheless additional information from such analyses of documentary sources has a potential to further refine the current reconstruction or for independent comparison.

7 Conclusions

In this study we have explored the potential for reconstructing monthly temperatures by using temperature indices derived from documentary evidence from five regions in Central Europe—the Czech Republic, Germany, Switzerland, Poland and the Carpathian Basin. We found that index data from the Czech Republic, Germany and Switzerland could be composited together to derive a complete monthly series of temperature reconstructions for the period AD 1500–1854. Polish documentary index data and those from the Carpathian Basin were considered too incomplete at this time, but comparisons with the other documentary series and with instrumental data indicate a clear reconstruction potential also for these data. This potential is most clearly seen for the Carpathian Basin data, which extend well into the instrumental period. Although in previous studies, documentary climate evidence was traditionally not extended into the instrumental period, our analyses (and also Leijonhufvud et al. 2008, 2009) clearly show the importance of having such a period of overlap between documentary and instrumental data, as this is crucial for both calibration and verification of the reconstructions. The CEU and Stockholm (Leijonhufvud et al. 2009) documentary based temperature reconstructions have been used in Luterbacher et al. (2009) to study the circulation dynamics and links to late winter/early spring temperature variability over the past half millennium.

The recent warming, expressed in the new reconstruction, is most pronounced in annual mean temperatures, but is very clear also in winter, spring and summer, but somewhat less so in autumn. For annual mean temperatures, eighteen of the last 20 years (1988–2007) were warmer than the upper bound of a 99% confidence interval for 20-year averages of reconstructed pre-instrumental temperatures. The corresponding numbers of years in the four seasons are: JJA (15), MAM (13), DJF (11), SON (6). Individual seasonal or monthly temperatures are less easily detected as unusual in a five century perspective, when reconstruction uncertainty is considered. June, August and JJA temperatures in 2003 and July in 2006 are the only recent single temperature values that rise above the estimated 95% confidence interval for pre-instrumental data.

The new reconstruction displays a previously unobserved long-term decrease in temperature variability over last five centuries in winter, spring and summer. This long-term decrease in temperature variance is perhaps the most notable feature that the new CEU reconstruction adds to the knowledge from already available reconstructions. Analysis of output from climate model simulations, with realistic forcing histories, could provide a useful way to check if this feature is also seen in models (Zorita et al. 2009).

The new CEU reconstruction offers some important new advances:

1. Monthly temperature reconstructions are provided for the entire period back to AD 1500.
2. The early instrumental data used here are corrected for a systematic warm bias (mainly in the summer) of temperatures measured before screens with proper radiation protection were introduced.
3. Calibration and verification is made directly against overlapping instrumental data.
4. Reliability of proxy data in the pre-instrumental period is quantified by means of the expressed population signal among the individual index series. This information is then included in the error estimation.
5. Due to variance adjustment for changing sample depth, the variance of the reconstructed temperatures is unaffected by variance artifacts related to changes in replication, and, as we calibrate using long continuous composite records, the variance does not decrease backwards in time.

We conclude that the new Central European temperature reconstruction can potentially be used to improve the robustness of current gridded temperature reconstructions for the last 500 years. Even further improvement is expected if more documentary data from other European countries are better explored and made more complete.

Acknowledgements This work was financially supported by EU project FP-6 no. 017008 European Climate of the Past Millennium (MILLENNIUM). Anders Moberg was funded by the Swedish Research Council, VR. Jürg Luterbacher acknowledges support the EU/FP6 project CIRCE (grant 036961) and from the EU/FP7 project ACQWA (grant 212250). We would like to thank Michael Begert (Zurich, Switzerland) for the Basel temperature series. Two anonymous reviewers are acknowledged for their useful and constructive criticism of the manuscript.

References

Alexandersson H (1986) A homogeneity test applied to precipitation data. J Climatol 6:661–675. doi:10.1002/joc.3370060607

Alexandersson H, Eriksson B (1989) Climate fluctuations in Sweden 1860–1987. SMHI reports meteorology and climatology 58. SMHI, Norrköping

Auer I, Böhm R, Jurkovic A, Lipa W, Orlik A, Potzmann R, Schöner W, Ungersböck M, Matulla C, Briffa K, Jones P, Efthymiadis D, Brunetti M, Nanni T, Maugeri M, Mercalli L, Mestre O, Moisselin JM, Begert M, Müller-Westermeier G, Kveton V, Bochnicek O, Stastny P, Lapin M, Szalai S, Szentimrey T, Cegnar T, Dolinar M, Gajic-Capka M, Zaninovic K, Majstorovic Z, Nieplova E (2007) HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region. Int J Climatol 27(1):17–46. doi:10.1002/joc.1377

Balling RC, Vose RS, Weber GR (1998) Analysis of long-term European temperature records: 1751–1995. Clim Res 10(3):193–200. doi:10.3354/cr010193

Bartholy J, Pongrácz R, Molnár Z (2004) Classification and analysis of past climate information based on historical documentary sources for the Carpathian basin. Int J Climatol 24(14):1759–1776. doi:10.1002/joc.1106

Begert M, Schlegel T, Kirchhofer W (2005) Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. Int J Climatol 25(15):65–80. doi:10.1002/joc.1118

Böhm R, Jones PD, Hiebl J, Brunetti M, Frank D, Maugeri M (2009) The early instrumental warm-bias: a solution for long central European temperatures series 1760–2007. Clim Change. doi:10.1007/s10584-009-9649-4

Brázdil R (1996) Reconstructions of past climate from historical sources in the Czech Lands. In: Jones PD, Bradley RS, Jouzel J (eds) Climatic variations and forcing mechanisms of the last 2000 years. Springer, Berlin, pp 409–431

Brázdil R, Pfister C, Wanner H, von Storch H, Luterbacher J (2005) Historical climatology in Europe—the state of the art. Clim Change 70:363–430. doi:10.1007/s10584-005-5924-1

Brázdil R, Černušák T, Řezníčková L (2008) Weather information in the diaries of the Premonstratensian Abbey at Hradisko, in the Czech Republic, 1693–1783. Weather 63(7):201–207. doi:10.1002/wea.264

Briffa KR, Osborn TJ, Schweingruber FH, Jones PD, Shiyatov SG, Vaganov EA (2002) Tree-ring width and density data around the Northern Hemisphere: part 1, local and regional climate signals. Holocene 12:737–757. doi:10.1191/0959683602hl587rp

Brohan P, Kennedy JJ, Haris I, Tett SFB, Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. J Geophys Res 111:D12106. doi:10.1029/2005JD006548

Burri M, Rutishauser T (2008) Pflanzenphänologische Beobachtungen—Dank Johann Jakob Sprünglis Aufzeichnungen zu gesteigerten Ernten und neuen Erkenntnissen in der Klimaforschung. In: Stuber M, Moser P, Gerber-Visser G, Pfister C (eds) Oekonomische und Gemeinnützige Gesellschaft des Kantons Bern 1759–2009. Verlag Paul Haupt, Bern (accepted)

Camuffo D, Jones P (eds) (2002) Improved understanding of past climatic variability from early daily European instrumental sources. Kluwer Academic Publishers, Dordrecht Boston London

Cook ER, Kairiukstis LA (eds) (1990) Methods of dendrochronology: applications in the environmental sciences. Kluwer Academic Publishers, Dordrecht

Cook ER, Briffa KR, Jones PD (1994) Spatial regression methods in dendroclimatology: a review and comparison of two techniques. Int J Climatol 14(4):379–402. doi:10.1002/joc.3370140404

Dobrovolný P, Brázdil R, Valášek H, Kotyza O, Macková J, Halíčková M (2009) A standard paleoclimatological approach to temperature reconstruction in historical climatology: an example from the Czech Republic, AD 1718–2007. Int J Climatol 29(10):1478–1492. doi:10.1002/joc.1789

Eddy JA (1976) The Maunder minimum. Science 192:1189–1202. doi:10.1126/science.192.4245.1189

Esper J, Frank DC, Wilson RJS, Briffa KR (2005) Effect of scaling and regression on reconstructed amplitude for the past millennium. Geophys Res Lett 32:L07711. doi:10.1029/2004GL021236

Flohn H (1979) Zwei bayerische Wetterkalender aus der Reformationszeit. Wiss Mitt Meteor Inst München 35:173–177

Frank D, Büntgen U, Böhm R, Maugeri M, Esper J (2007a) Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. Quat Sci Rev 26:3298–3310. doi:10.1016/j.quascirev.2007.08.002

Frank D, Esper J, Cook ER (2007b) Adjustment for proxy number and coherence in a large-scale temperature reconstruction. Geophys Res Lett 34:L16709. doi:10.1029/2007GL030571

Glaser R (1997) On the course of temperature in central Europe since the year 1000 AD. Hist Soc Res 22:59–87

Glaser R (2001) Klimageschichte Mitteleuropas. 1000 Jahre Wetter, Klima, Katastrophen. Primus Verlag, Darmstadt

Glaser R (2008) Klimageschichte Mitteleuropas. 1200 Jahre Wetter, Klima, Katastrophen. Primus Verlag, Darmstadt

Glaser R, Riemann D (2009) A thousand year record of climate variation for Central Europe at a monthly resolution. J Quat Sci 24(5):437–449

Glaser R, Stangl H (2004) Climate and floods in Central Europe since AD 1000: data, methods, results and consequences. Surv Geophys 25:485–510. doi:10.1007/s10712-004-6201-y

Gouirand I, Linderholm H, Moberg A, Wohlfarth B (2008) On the spatiotemporal characteristics of Fennoscandian tree-ring based summer temperature reconstructions. Theor Appl Climatol 91:1–25. doi:10.1007/s00704-007-0311-7

Guiot J, Nicault A, Rathgeber C, Edouard JL, Guibal F, Pichard G, Till C (2005) Last-millenium summer temperature variations in western Europe based on proxy data. Holocene 15:489–500. doi:10.1191/0959683605hl819rp

Jones PD (2001) Early European instrumental record. In: Jones PD, Ogilvie AEJ, Davies TD, Briffa K (eds) History and climate: memories of the future? Kluwer Academic Press, New York, pp 55–77

Jones PD, Briffa KR (2006) Unusual climate in northwest Europe during the period 1730 to 1745 based on instrumental and documentary data. Clim Change 79:361–379. doi:10.1007/s10584-006-9078-6

Jones PD, Osborn TJ, Briffa KR, Folland CK, Horton EB, Alexander LV, Parker DE, Rayner NA (2001) Adjusting for sampling density in grid box land and ocean surface temperature series. J Geophys Res 106(D4):3371–3380

Leijonhufvud L, Wilson R, Moberg A (2008) Documentary data as proxy variables for Stockholm late winter to early spring temperatures in the 18th and 19th centuries. Holocene 18:333–343. doi:10.1177/0959683607086770

Leijonhufvud L, Wilson R, Moberg A, Söderberg J, Retsö D, Söderlind U (2009) Five centuries of winter/spring temperatures in Stockholm reconstructed from documentary evidence and instrumental observations. Clim Change. doi:10.1007/s10584-009-9650-y

Luterbacher J, Xoplaki E, Dietrich D, Rickli R, Jacobeit J, Beck C, Gyalistras D, Schmutz C, Wanner H (2002) Reconstruction of sea-level pressure fields over the Eastern North Atlantic and Europe back to 1500. Clim Dyn 18:545–561

Luterbacher J, Dietrich D, Xoplaki E, Grosjean M, Wanner H (2004) European seasonal and annual temperature variability, trends and extremes since 1500. Science 303:1499–1503. doi:10.1126/science.1093877

Luterbacher J, Liniger MA, Menzel A, Estrella N, Della-Marta PM, Pfister C, Rutishauser T, Xoplaki E (2007) The exceptional European warmth of autumn 2006 and winter 2007: historical context, the underlying dynamics and its phenological impacts. Geophys Res Lett 34:L12704. doi:10.1029/2007GL029951

Luterbacher J, Koenig SJ, Franke J, van der Schrier G, Della-Marta PM, Jacobeit J, Küttel M, Gonzalez-Rouco F, Zorita E, Xoplaki E, Stössel M, Rutishauser T, Wanner H, Pfister C, Brázdil R, Dobrovolný P, Camuffo D, Bertolin C, Moberg A, Leijonhufvud L, Söderberg J, Allan R, Wilson R, Wheeler D, Barriendos M, Glaser R, Riemann D, Nordli PØ, Limanówka D, van Engelen A, Zerefos CS (2009) Circulation dynamics and its influence on European and Mediterranean January–April climate over the past half millennium: results and insights from instrumental data, documentary evidence and coupled climate models. Clim Change (this volume)

Manley G (1974) Central England temperatures: monthly means 1659 to 1973. Q J Royal Meteorol Soc 100:389–405. doi:10.1002/qj.49710042511

Mann ME, Gille E, Bradley RS, Hughes MK, Overpeck JT, Keimig FT, Gross W (2000) Global temperature patterns in past centuries: an interactive presentation. Earth Interact 4:1–29. doi:10.1175/1087-3562(2000)004<0001:GTPIPC>2.3.CO;2

Meier N, Rutishauser T, Pfister C, Wanner H, Luterbacher J (2007) Grape harvest dates as a proxy for Swiss April to August temperature reconstructions back to AD 1480. Geophys Res Lett 34:L20705. doi:10.1029/2007GL031381

Osborn TJ, Briffa KB, Jones PD (1997) Adjusting variance for sample size in tree-ring chronologies and other regional mean timeseries. Dendrochronologia 15:89–99

Pauling A, Luterbacher J, Casty C, Wanner H (2006) Five hundred years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation. Clim Dyn 26(4):387–405. doi:10.1007/s00382-005-0090-8

Peterson TC, Easterling DR, Karl TR, Groisman P, Auer I, Böhm R, Plummer N, Nicholis N, Torok S, Vincent L, Tuomenvirta H, Salinger J, Förland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D (1998) Homogeneity adjustments of in situ climate data: a review. Int J

Climatol 18(13):1493–1517. doi:10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T

Pfister C (1981) Die Fluktuationen der Weinmosterträge im schweizerischen Weinland vom 16. bis ins frühe 19. Jahrhundert. Klimatische Ursachen und sozioökonomische Bedeutung. Schweiz. Zeitschrift für Geschichte 31:455–491

Pfister C (1984) Das Klima der Schweiz von 1525 bis 1860 und seine Bedeutung in der Geschichte von Bevölkerung und Landwirtschaft, Academia Helvetica 5, 2 vols. Verlag Paul Haupt, Bern

Pfister C (1992) Monthly temperature and precipitation in central Europe 1525–1979: quantifying documentary evidence on weather and its effects. In: Bradley RS, Jones PD (eds) Climate since AD 1500. Routledge, London, pp 118–143

Pfister C (1999) Wetternachhersage. 500 Jahre Klimavariationen und Naturkatastrophen (1496–1995). Verlag Paul Haupt, Bern

Pfister C, Brázdil R (1999) Climatic variability in sixteenth-century Europe and its social dimension: a synthesis. Clim Change 43(1):5–53. doi:10.1023/A:1005585931899

Pfister C, Brázdil R, Glaser R, Barriendos M, Camuffo D, Deutsch M, Dobrovolný P, Enzi S, Guidoboni E, Kotyza O, Militzer S, Rácz L, Rodrigo FS (1999) Documentary evidence on climate in sixteenth-century Europe. Clim Change 43:55–110. doi:10.1023/A:1005540707792

Pfister C, Luterbacher J, Wanner H, Wheeler D, Brázdil R, Ge Q, Hao Z, Moberg A, Grab S, Rosario del Prieto M (2008) Documentary evidence as climate proxies. In: PAGES (Past Global Changes). Proxy-specific White Paper produced from the PAGES/CLIVAR workshop, Trieste, 2008. http://www.pages-igbp.org/cgibin/WebObjects/products.woa/wa/product?id=331. Accessed 8 May 2009

Przybylak R, Majorowicz J, Wójcik G, Zielski A, Chorążyczewski W, Marciniak K, Nowosad W, Oliński P, Syta K (2005) Temperature changes in Poland from the 16th to the 20th centuries. Int J Climatol 25(6):773–791. doi:10.1002/joc.1149

Rácz L (1999) Climate history of Hungary since 16th century: past, present and future. Centre for Regional studies of Hungarian Academy of Sciences, Pecs

Rutherford S, Mann ME, Osborn TJ, Bradley RS, Briffa KR, Hughes MK, Jones PD (2005) Proxy-based Northern Hemisphere surface temperature reconstructions: sensitivity to method, predictor network, target season, and target domain. J Clim 18:2308–2329. doi:10.1175/JCLI3351.1

Schär C, Vidale PL, Lüthi D, Frei C, Häberli C, Liniger MA, Appenzeller C (2004) The role of increasing temperature variability in European summer heatwaves. Nature 427(6972):332–336. doi:10.1038/nature02300

Shabalova MV, van Engelen AFV (2003) Evaluation of a reconstruction of winter and summer temperatures in the Low countries, AD 764–1998. Clim Change 58(1–2):219–242. doi:10.1023/A:1023474032539

Shindell DT, Schmidt GA, Mann ME, Rind D, Waple A (2001) Solar forcing of regional climate change during the Maunder minimum. Science 294(5549):2149–2152. doi:10.1126/science.1064363

Tuomenvirta H, Alexandersson H, Drebs A, Frich P, Nordli PØ (2000) Trends in Nordic and Arctic temperature extremes and ranges. J Clim 13(5):977–990. doi:10.1175/1520-0442(2000)013<0977:TINAAT>2.0.CO;2

van Engelen AFV, Buisman J, IJnsen F (2000) Reconstruction of the low countries temperature series AD 764–1998. In: Mikami T (ed) Proceedings of the international conference on climate change and variability—past, present and future. Tokyo Metropolitan University, Tokyo, pp 151–157

van Engelen AFV, Buisman J, IJnsen F (2001) A millennium of weather, winds and water in the low Countries. In: Jones PD, Ogilvie AEJ, Davies TD, Briffa KR (eds) History and climate: memories of the future? Kluwer Academic Press, New York, pp 101–124

van Vliet LJ, Young IT, Verbeek PW (1998) Recursive Gaussian derivative filters. In: Proceedings of the 14th international conference on pattern recognition, Brisbane, 16–20 August 1998

von Storch H, Zwiers FW (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge

Wahl E, Ammann C (2007) Robustness of the Mann, Bradley, Hughes reconstruction of the Northern Hemisphere surface temperatures: examination of criticisms based on the nature and processing of proxy climate evidence. Clim Change 85(1–2):33–69. doi:10.1007/s10584-006-9105-7

Wigley TML, Briffa KR, Jones PD (1984) On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. J Clim Appl Meteorol 23(2):201–213. doi:10.1175/1520-0450(1984)023<0201:OTAVOC>2.0.CO;2

Wilson R, Tudhope A, Brohan P, Briffa K, Osborn T, Tett S (2006) 250-years of reconstructed and modeled tropical temperatures. J Geophys Res 111:C10007. doi:10.1029/2005JC003188

Xoplaki E, Luterbacher J, Paeth H, Dietrich D, Steiner N, Grosjean M, Wanner H (2005) European spring and autumn temperature variability and change of extremes over the last half millennium. Geophys Res Lett 32:L15713. doi:10.1029/2005GL023424

Zorita E, Moberg A, Leijonhufvud L, Wilson R, Brázdil R, Dobrovolný P, Luterbacher J, Pfister C, Glaser R, Söderberg J, González-Rouco F (2009) European temperature records of the past five centuries based on documentary information compared to climate simulations. Clim Change (this volume)