Quaternary Science Reviews 46 (2012) 66–79

Contents lists available at SciVerse ScienceDirect

Quaternary Science Reviews

journal homepage: www.elsevier.com/locate/quascirev

Multi-archive summer temperature reconstruction for the European Alps, AD 1053–1996

Mathias Trachsel [a,b,*], Christian Kamenik [a,b], Martin Grosjean [a,b], Danny McCarroll [c], Anders Moberg [d], Rudolf Brázdil [e], Ulf Büntgen [a,f], Petr Dobrovolný [e], Jan Esper [g], David C. Frank [a,f], Michael Friedrich [h], Rüdiger Glaser [i], Isabelle Larocque-Tobler [a,b], Kurt Nicolussi [j], Dirk Riemann [i]

[a] Oeschger Center for Climate Change Research, University of Bern, Zähringerstrasse 25, 3012 Bern, Switzerland
[b] Department of Geography, University of Bern, Erlachstrasse 9a, 3012 Bern, Switzerland
[c] Department of Geography, University of Wales Swansea, Singleton Park Swansea, SA2 8PP, UK
[d] Department of Physical Geography and Quaternary Geology, Stockholm University, 106 91 Stockholm, Sweden
[e] Institute of Geography, Masaryk University, 611 37 Brno, Czech Republic
[f] Swiss Federal Institute for Forest, Snow, and Landscape Research (WSL), Zürcherstrasse 111, 8903 Birmensdorf, Switzerland
[g] Department of Geography, Johannes Gutenberg University, 55099 Mainz, Germany
[h] Institute of Botany, Hohenheim University, Garbenstrasse 30, 70593 Stuttgart, Germany
[i] Institute of Physical Geography, University of Freiburg, 79085 Freiburg, Germany
[j] Institute of Geography, University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria

ARTICLE INFO

Article history:
Received 19 September 2011
Received in revised form
20 April 2012
Accepted 25 April 2012
Available online xxx

Keywords:
Paleoclimate
Alps
Last millennium
Multi-proxy
Climate reconstruction
Climate

ABSTRACT

We present a multi-archive, multi-proxy summer temperature reconstruction for the European Alps covering the period AD 1053–1996 using tree-ring and lake sediment data. The new reconstruction is based on nine different calibration approaches and errors were estimated conservatively. Summer temperatures of the last millennium are characterised by two warm (AD 1053–1171 and 1823–1996) and two cold phases (AD 1172–1379 and 1573–1822). Highest pre-industrial summer temperatures of the 12th century were 0.3 °C warmer than the 20th century mean but 0.35 °C colder than proxy derived temperatures at the end of the 20th century. The lowest temperatures at the end of the 16th century were ~1 °C lower than the 20th century mean.

1. Introduction

To assess the anthropogenic fingerprint of recent climate change, detailed insight into the evolution of pre-industrial (natural) climate during the last millennium is essential (Hegerl et al., 2011). While attention has focussed mainly on hemispheric-scale annual temperature reconstructions for the last 1000 years (e.g. Moberg et al., 2005; Frank et al., 2007a; Mann et al., 2008), regional scale and seasonally-resolved information is required to critically test the ability of General Circulation Models (GCM) to reconstruct the climate of the past, and therefore to assess their reliability for predicting the climate of the future (e.g. McCarroll, 2010; Yamazaki et al., 2011). Spatial and seasonal variation in response to natural and anthropogenic forcing is also key to defining the sensitivity of the climate system to rising levels of greenhouse gases and to predicting the ecological, social and economic consequences of future climate change (Frank et al., 2010).

In the recent past a number of reconstructions of summer temperature for the Greater Alpine region (GAR) have been presented. Climate field reconstructions are available for the last 500 years (e.g. Luterbacher et al., 2004; Casty et al., 2005). Reconstructions based on documentary data covering the last 500 or 1000 years were presented recently (Glaser and Riemann, 2009; Dobrovolný et al., 2010). Further, several tree-ring based reconstructions (e.g. Büntgen et al., 2005, 2006, 2011; Corona et al., 2010, 2011) and high-resolution quantitative reconstructions from lake sediments (e.g. Larocque-Tobler et al., 2010; Trachsel et al., 2010)

* Corresponding author. Bjerknes Centre for Climate Research, Allégaten 55, 5007 Bergen, Norway.
E-mail address: mathias.trachsel@uni.no (M. Trachsel).

0277-3791/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.quascirev.2012.04.021

are available, and parts of the GAR are covered by the multi-proxy reconstruction of Guiot et al. (2005).

Reconstructions covering small areas are usually based on only one type of climate proxy, which makes them prone to proxy-specific restrictions such as the 'observational bias' of documentary data (Brázdil et al., 2005, 2010), the 'segment-length curse' of tree rings (Cook et al., 1995) or dating uncertainties of lake sediments. A multi-archive climate reconstruction is potentially less affected by these limitations and may provide a less biased understanding of pre-industrial climate evolution.

In this study, we present a multi-archive summer temperature reconstruction for the Greater Alpine Region spanning the period AD 1053–1996. The reconstruction is based on a set of climate proxies from lake sediments and tree-rings. Estimates of uncertainty are essential for detection and attribution studies (e.g. Hegerl et al., 2006) as well as data-model comparison (e.g. Graham et al., 2007). Here we use sensitivity analysis, thereby including uncertainty due to differences between the proxies during the reconstruction period (e.g. Moberg et al., 2005) as well as the uncertainties due to calibration (e.g. Rutherford et al., 2005). Our sensitivity analysis involves excluding (i) specific frequency bands of proxy time series (sensu Moberg et al., 2005), (ii) proxies that inherit an associated dating uncertainty and (iii) proxy series at random. We compare our reconstructions with tree-ring (Büntgen et al., 2011) and independent documentary-based reconstructions on a regional scale (Glaser and Riemann, 2009; Dobrovolný et al., 2010) and with multi-proxy reconstructions on hemispheric scale (e.g. Frank et al., 2007a; Hegerl et al., 2007; Mann et al., 2008).

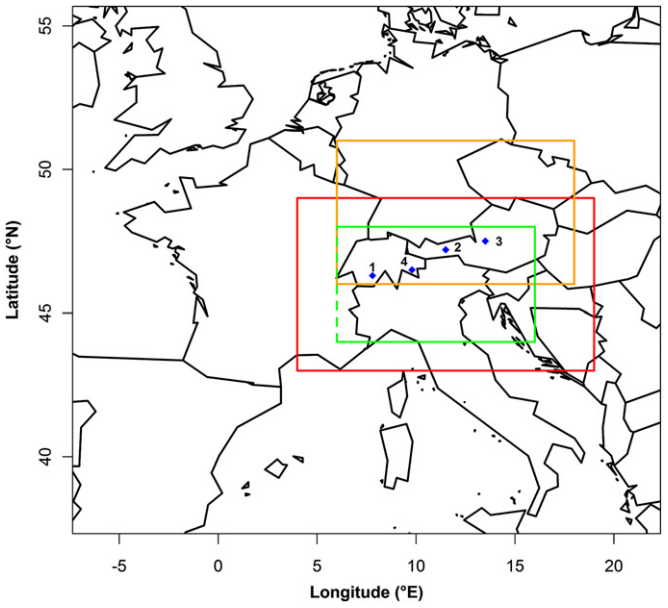

Fig. 1. Locations of proxies used for our reconstructions (blue dots): 1. Lötschental; 2. Tyrol; 3. Dachstein; 4. Lake Silvaplana. Red, orange, light green rectangles, respectively, enclose the extension of the gridded data set of Böhm et al. (2010), the area for which the temperature reconstruction of Dobrovolný et al. (2010) is representative, and the areas for which Büntgen et al. (2009) compiled data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2. Materials and methods

2.1. Data

From the large number of datasets produced within the European Union project Millennium (http://www.ncdc.noaa.gov/paleo/pubs/millennium/millennium.html), we chose in total eight tree-ring and lake sediment based proxies from four different locations in the Greater Alpine Region (Table 1, Fig. 1). Proxy time series were selected according to their length (they should at least cover the period back to 1200) and their sensitivity to summer (JJA) temperatures in the calibration period (AD 1760–1996). Seven of these proxy series (all except biogenic silica flux) covered the period AD 1053–1996, the period to which we finally restricted our reconstruction.

The raw individual series within the three tree-ring width (trw) chronologies and the two tree-ring maximum density (mxd) chronologies were detrended using Regional Curve Standardisation (RCS, e.g. Esper et al., 2003) to retain as much low-frequency variability as possible in the chronology. One tree-ring record is based on measurements of 'blue intensity' (McCarroll et al., 2002; Campbell et al., 2007, 2011), an inexpensive alternative to X-ray densitometry which yields relative rather than absolute values. These data were detrended using a cubic spline; hence, they were considered useful only at frequencies shorter than about 50-year periods. The lake-sediment records include chironomids (Larocque-Tobler et al., 2010) and biogenic silica flux (bSi; Blass et al., 2007; Trachsel et al., 2010) and rely on probabilistic dates. Filtered time series for all eight series are shown in Fig. 2(a–h) and replications of the tree-ring series are shown in Fig. 3. All data series correlate strongly with summer temperature (Table 2). Significance levels were calculated after correcting degrees of freedom (DF) according to Dawdy and Matalas (1964) (i.e. 'DF' $= (n-2)^*(1-r_1{}^*r_1{}')/(1+r_1{}^*r_1{}')$, where n is the initial sample size and r_1 and $r_1{}'$ are the first serial correlation coefficients of the two time series compared).

To assess the similarity of all proxy series included in this reconstruction, mean inter-series correlations (rbar) were calculated for raw data, for high-pass filtered data (1–31- year) and for 31-year low-pass (Gaussian) filtered data in 100 and 200-year time windows, respectively (Fig. 2i–k). For low-pass filtered (31-year Gaussian) data, inter-series variances between the standardised

Table 1
Proxies used in this study.

Proxy #	Type	Location	Reference	Record length (years AD)	Tree species	Frequency bands used
1	Late-wood density	Lötschental (Switzerland)	Büntgen et al., 2006	755–2004	*Larix decidua*	All
2	Late-wood density	Tyrol (Austria)	Esper et al., 2007	1053–2003	*Picea abies*	All
3	Tree-ring width	Tyrol (Austria)	Büntgen et al., 2005	848–1997	*Pinus cembra*	All
4	Tree-ring width	Alpine Arc	Büntgen et al., 2009	951–2004	*Larix decidua*	All
5	Tree-ring width	Dachstein (Austria)	Friedrich unpubl.	900–1996	*Larix decidua*	All
6	Blue Intensity	Dachstein (Austria)	Friedrich unpubl.	813–1996	Conifer	1–31; 1–11; 11–31; 31–51
7	Chironomids	Silvaplana (Switzerland)	Larocque-Tobler et al., 2010	1007–2002		11–31; 31–51; >31; >51
8	Biogenic Silica	Silvaplana (Switzerland)	Blass et al., 2007; Trachsel et al., 2010	1181–1949		11–31; 31–51

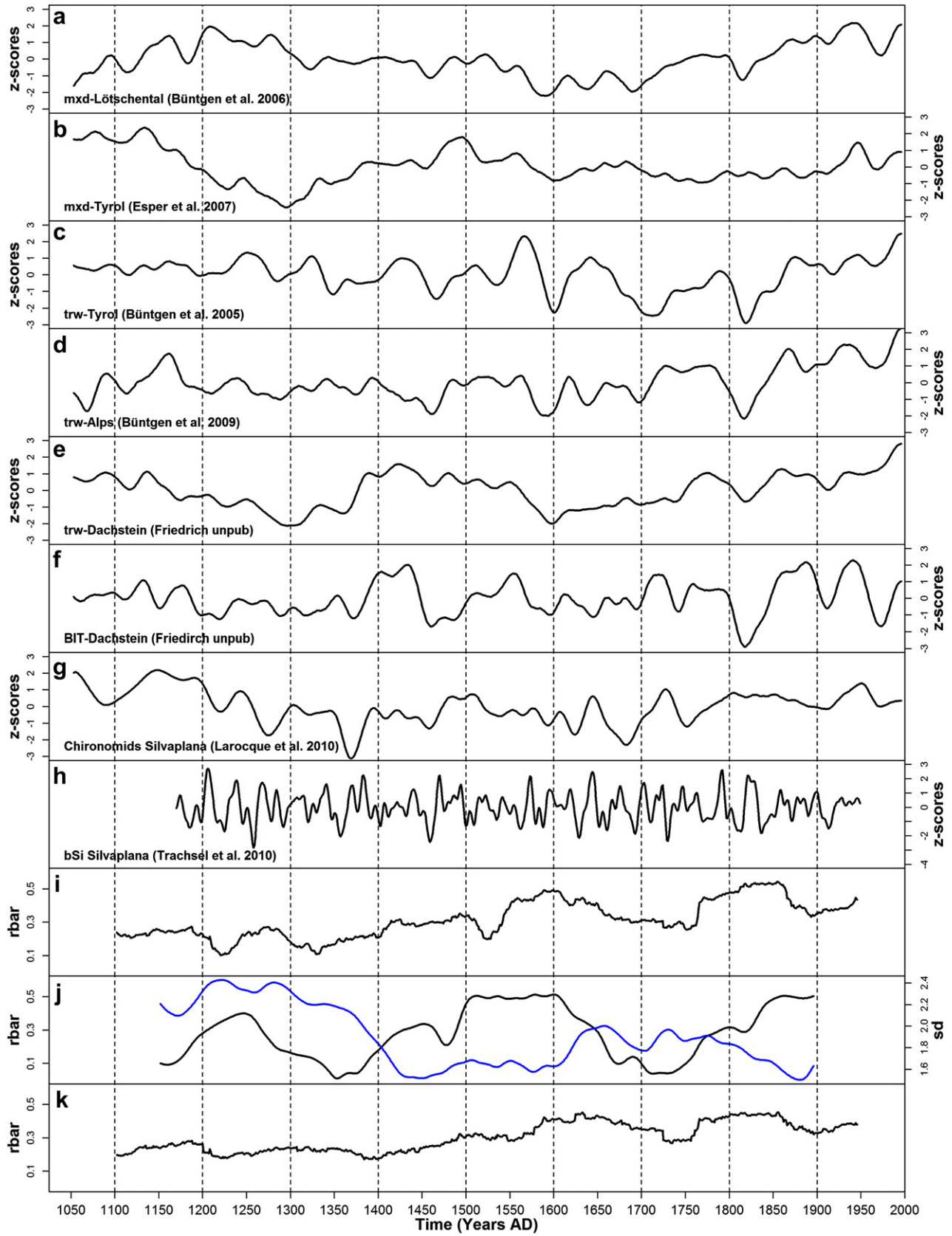

Fig. 2. Normalised proxy data (*z*-scores, 31-years low-pass (Gaussian) filtered) used in this study (order similar to Table 1): (a) Büntgen et al. (2006), (b) Esper et al. (2007), (c) Büntgen et al. (2005), (d) Büntgen et al. (2009), (e) Friedrich unpubl. (tree-ring width) (f) Friedrich unpubl. (blue intensity), (g) Larocque-Tobler et al. (2010) and (h) Trachsel et al. (2010) (frequency constrained to 11–51 years); and (i–k) their mean inter-series correlations (rbar): (i) unfiltered data (100-year time window), (j) 31-year (Gaussian) low-pass filtered data (black, 200-year time window) along with their inter-series standard deviations (sd (blue), 200-year running means of annual values), and (k) 1–31-year (Gaussian) high-pass filtered data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

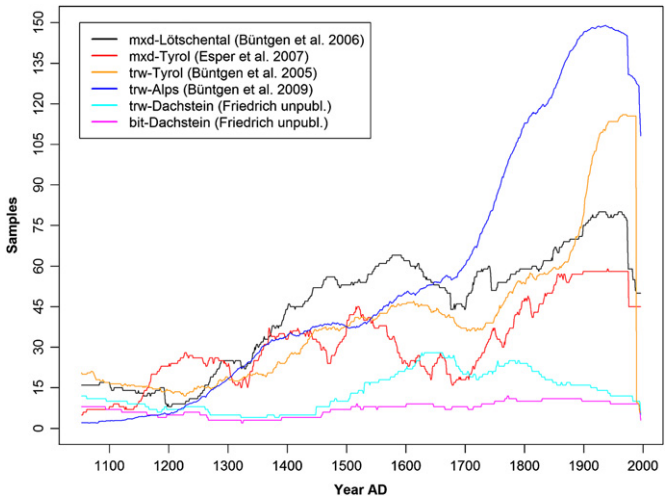

Fig. 3. Sample depth of the six tree-ring series used in this study. Note: The number of trees in the series of Büntgen et al. (2005, 2009) are two and ten times higher respectively than shown in the figure.

proxy series were calculated for each year. We accounted for decreased variance caused by autocorrelation by multiplying the calculated variance by a correction factor consisting of standard error multiplied by the square root of $(1 + r_1)/(1 - r_1)$ where r_1 is the first serial correlation coefficient (Ramsey and Schafer, 2002). The square root of this corrected annual variance (standard deviation) was averaged over 200 years to assess the spread of the low-frequency variability of proxy data.

Since the biogenic silica record covers the period AD 1173–1949, it is included in only one of the reconstructions presented in this study. Chironomids are the only proxies which were calibrated against a spatial temperature gradient ('calibration in space', Larocque-Tobler et al., 2010); all other proxies were calibrated against instrumental meteorological measurements ('calibration in time' see below). Since all these data fall within the period AD 1000–2000, all dates mentioned hereafter are years *Anno Domini* (AD).

2.2. Calibration and reconstruction

Calibration was undertaken, for the period 1760–1996, on JJA mean temperatures averaged over the region 43°–49° N and 4°–19° E using gridded temperature data (Böhm et al., 2010, Fig. 1) which includes early instrumental data (Böhm et al., 2010). All proxies except biogenic silica have data in the whole of this period. We tested calibration without frequency splitting, and after splitting proxies and instrumental data into two (1–31, >31-year) or four (1–11, 11–31, 31–51, >51-year) frequency bands using Gaussian low- and high-pass filters. Using different frequency bands allows us to: (i) include data at non-annual resolution (e.g. Moberg et al., 2005; Guiot et al., 2010; Boucher et al., 2011) and (ii) test whether proxy series are differently weighted during calibration in the different frequency bands. For calibration we applied Composite Plus Scale (CPS, e.g. Riedwyl et al., 2008), Multiple Ordinary Least-Squares Regression (OLSR, e.g. Legendre and Legendre, 1998) and Partial Least-Squares Regression (PLSR, e.g. Martens and Naes, 1989) to test for possible differences in reconstructions introduced by different calibration techniques. Reconstructions obtained for individual frequency bands were rescaled to the calibration target (Böhm et al., 2010) to avoid loss of amplitude associated with inverse regression techniques (e.g. Esper et al., 2005; Bürger et al., 2006). The rescaled reconstructions of the individual frequency bands were summed to obtain reconstructions covering all frequency bands. Altogether, our different calibration approaches resulted in nine different reconstructions of summer temperature for the last millennium. Finally, we averaged the nine reconstructions (thereafter referred to as variants 1–9) to obtain a consensus which was validated with independent reconstructions. Information on the nine individual calibration and reconstruction methods is summarized in Table 3. Differences in methodology and motivations for these methodological tests are outlined below.

To account for potential edge effects when filtering proxy series we performed calibrations (i) excluding all data affected by edge effects and (ii) including proxy series filtered at the beginning/end. When filtering at the beginning/end of time series, we artificially extended the filtered time series (i) by adding the climatology (i.e. the mean value of the last 30 years of the time-series), (ii) by reflecting the time series 'horizontally' about the final data point (i.e. by adding the last 30 years of the time series in reversed order), and (iii) by reflecting the time series 'horizontally' and 'vertically' about the last data point (i.e. a point reflection about the last data point or turning the time series by 180° about the last data point). Subsequently, the three extended data series were filtered and individually weighted (sum of individual weights = 1) in order to obtain lowest Root Mean Square Error of Prediction (RMSEP) compared to raw data, applying the methodology suggested by Mann (2008).

Calibration methods used are based on different statistical frameworks. CPS standardises all proxy time series over the common period and averages them. Then the mean series is scaled (i.e. mean and variance are matched) to the climate variable of interest over the instrumental (calibration) period (e.g. Riedwyl et al., 2008). Since variance depends upon the strength of the correlation among the individual time series as well as the number of records considered, variance was stabilised using the method proposed by Frank et al. (2007a), i.e., calculating mean correlation coefficients among proxies for 45-year running windows. In contrast to CPS, the two regression-based approaches are built on a more thorough statistical framework with underlying assumptions. OLSR, for example, requires independence among explanatory variables; co-linearity may cause instability in parameter estimation (e.g. Legendre and Legendre, 1998). PLSR avoids this problem by reducing a large number of predictors to a small number of components (latent variables) which are then used in place of the original predictors. Unlike similar dimension-reduction techniques, such as Principal Component Regression (PCR), components are chosen to provide maximum correlation with the dependent variable (Martens and Naes, 1989; de Jong, 1993).

Low-pass filtered proxy series (>31 and >51 years) have high autocorrelations, low degrees of freedom and multi co-linearities. For these frequency bands, we exclusively applied PLSR (instability in OLSR parameter estimation, see above). Only the first PLSR component was used (further components did not improve calibration statistics). In contrast to the frequency bands 11–31, >31 and >51 years, the frequency bands 1–31 and 1–11 years have low autocorrelation and, hence, high degrees of freedom. For PLSR run on these frequency bands, the number of components included in the model was assessed by 10-fold cross-validation (e.g. Efron and Gong, 1983): Only components leading to a reduction of the RMSEP of more than 5% were retained in the model (Birks, 1998).

2.3. Model validation and sensitivity tests

Calibration models were validated by RMSEP (Birks, 1998), coefficient of determination (r^2), Amplitude/RMSEP (Trachsel et al., 2010), and reduction of error (RE) and coefficient of efficiency (CE)

Table 2
Correlations of all proxies with summer temperature (Böhm et al., 2010).

	Esper et al. (2007)	Büntgen et al. (2009)	Friedrich trw RCS	Friedrich BIT spline	Büntgen et al. (2006)	Büntgen et al. (2005)	Larocque-Tobler et al. (2010)	Trachsel et al. (2010)
	r	*r*	*r*	*r*	*r*	*r*	*r*	*r*
Raw	***0.46***	***0.55***	***0.54***	***0.59***	**0.59**	***0.41***	NA	NA
1–31-year	***0.42***	***0.58***	***0.49***	***0.68***	***0.63***	***0.34***	NA	0.4[a]
>31-year	***0.67***	*0.53*	***0.78***	0.32	0.45	***0.62***	0.29	***0.74***[b]

Bold and italic $p < 0.05$.
[a] 11–31-year frequency band.
[b] 31–51-year frequency band.

statistics (Cook et al., 1994) for the recombined series. With autocorrelated time series, RMSEP is underestimated by leave-one-out and k-fold cross-validation (e.g. Telford and Birks, 2009). Therefore, we assessed RMSEP by h-block cross-validation (Burman et al., 1994) wherein the observation for which predictions are made is omitted from calibration, along with *h* observations on either side (Telford and Birks, 2009). Preliminary autocorrelation functions suggested $h = 20$ for 31-year low-pass filtered data and $h = 40$ for 51-year low-pass filtered data. For the high-pass filtered proxy time series that showed no significant autocorrelation, uncertainty of the calibration model was assessed by leave-one-out cross-validation. Errors among proxies within the same frequency band were combined based on Gaussian error propagation (i.e., square root of the sum of the squares). Gaussian error propagation was also applied when combining different frequency bands.

To quantify uncertainty, the overlap of the different reconstructions was determined by applying the 'envelope method' used in the fourth report of the IPCC (Jansen et al., 2007).

If $|Tr_{it} - Tp_{tj}| \leq$ RMSEPi; the score of Tp_{tj} is set to 100%/9 = 11.11%
If RMSEPi $< |Tr_{it}\text{-}Tp_{tj}| \leq 2$*RMSEPi; the score of Tp_{tj} is set to 100%/(2*9) = 5.55%
if 2*RMSEPi $< |Tr_{it}\text{-}Tp_{tj}|$; the score of Tp_{tj} is set to 0%

where $i = 1,2,\ldots, 9$ (a specific reconstruction), $t = 1053, 1054,\ldots, 1996$ (year), Tr is reconstructed temperature; Tp is a temperature that ranges from $j = -2$ °C to $j = 1$ °C wrt 1901–2000 (at an interval of 0.01 °C); and RMSEPi is RMSEP of the ith reconstruction (nomenclature follows Ge et al., 2010). Hence, temperatures within ±1 RMSEP of an individual reconstruction were assigned a 'score' of 11.11%, temperatures between ±1 RMSEP and ±2 RMSEP of an individual reconstruction were assigned a 'score' of 5.55%. After repeating this procedure for all nine reconstructions 'scores' were added. Maximum overlap (100%) was obtained for temperatures falling all within the RMSEP of all nine individual reconstructions.

Significant changes among mean temperatures of the nine reconstructions were detected with CONstrained Incremental Sum-of-Squares (CONISS) clustering (Grimm, 1987), as implemented in the R-package 'rioja' (Juggins, 2009), and by the broken-stick model (Bennett, 1996).

To estimate the robustness of our reconstructions with respect to the proxy data included in reconstructions, we ran sensitivity tests that compared reconstructions including all available proxies with reconstructions omitting one and two proxy time series, respectively. These tests were run on three different reconstruction variants (variants 4, 5 and 7, Table 3). We choose these three variants because they cover the full reconstruction period (1053–1996) and we thereby compared calibration in different frequency bands with calibration of unconstrained data (variants 4 and 5 versus variant 7). Since we further combined proxies with calendar dates (tree rings) and proxies with probabilistic dates (lake sediments), we also tested the effects of including or excluding the latter (variant 4 versus variants 5 and 7).

For comparison of the Alpine reconstruction with other reconstructions covering the same area, we calculated Theil-Sen trend estimates (Theil, 1950; Sen, 1968) for 31-year windows. We thereby compared mid-term fluctuations but disregarded low-frequency variability.

Table 3
Numerical methods and error statistics of the nine calibration/reconstruction variants. Colour codes refer to Fig. 4.

Name	Number of frequency bands	Data included	Nr. of data series (Table 1)	Calibration period (AD)	RMSEP [°C]	r^2	Amplitude/ RMSEP	Reconstruction period (AD)	Frequency band and statistical method along with their colour coding for Fig. 3 (in parenthesis)
Variant 1	2	Tree-rings	1–6	1776–1981	0.76	0.61	6.44	1068–1981	1–31 (PLSR, Fig. 4d: green) and >31 (PLSR, Fig. 4e: blue); Fig. 4h: grey
Variant 2	2	Tree-rings and lake	1–7	1776–1981	0.78	0.62	6.19	1068–1981	1–31 (PLSR, Fig. 4d: black) and >31 (PLSR, Fig. 4e black); Fig. 4h: black
Variant 3	2	Tree-rings	1–6	1761–1996	0.8	0.61	6.26	1053–1996	1–31 (PLSR, Fig. 4d: green) and >31 (PLSR, Fig. 4e cyan); Fig. 4h: cyan
Variant 4	2	Tree-rings and lake	1–7	1761–1996	0.76	0.62	6.74	1053–1996	1–31 (PLSR, Fig. 4d: orange) and >31 (PLSR, Fig. 4e: orange); Fig. 4h: brown
Varriant 5	2	Tree-rings	1–6	1761–1996	0.8	0.61	6.61	1053–1996	1–31 (PLSR, var stab, Fig. 4d: blue) and >31 (PLSR, Fig. 4e: green); Fig. 4h: pink
Variant 6	2	Tree-rings	1–6	1761–1996	0.78	0.58	7.18	1053–1996	1–31 (CPS, var stab, Fig. 4d: cyan) and >31 (PLSR, Fig. 4e: green); Fig. 4h: gold
Variant 7	1	Tree-rings	1–6	1761–1996	0.84	0.48	6.18	1053–1996	None (PLSR; Fig. 4, blue) h:orange
Variant 8	4	Tree-rings and lake	1–7	1761–1996	0.76	0.66	6.91	1053–1996	1–10 (Fig. 4a, blue)/11–31 (Fig. 4b, blue)/31–51 (Fig. 4c, black; all OLSR) and >51 (PLSR, Fig. 4f, black); Fig. 4h: blue
Variant 9	4	Tree-rings and lake	1–8	1761–1949	0.76	0.66	7.16	1173–1949	1–10 (Fig. 4a, blue)/11–31 (Fig. 4b, black)/31–51 ((Fig. 4c, blue; all OLSR) and >51 (PLSR, (Fig. 4f, black); Fig. 4h: coral

All numerical analyses were carried out using R (R Development Core Team, 2011) and its add-on packages *gtools* (Warnes, 2010), *pls* (Wehrens and Mevik, 2007), *rioja* (Juggins, 2009) and *vegan* (Oksanen et al., 2011).

3. Results

3.1. Data

Fig. 2(a–h) shows filtered versions of all the proxy time series used in this study. In the following, we briefly introduce the proxy series on which incomplete information is available in peer-reviewed journals.

The Tyrol-mxd data set (Esper et al., 2007, Fig. 2b) was developed using a total number of 227 P*icea abies* samples from living trees and relict wood spanning the period 1053–2003 with a minimum replication of 4 samples. The mean inter-series correlation is always higher than $r = 0.53$. The trw record of Büntgen et al. (2009, Fig. 2d) is based on 2610 *Larix decidua* series coming from 40 sites located at higher elevations in the European Alps and is a subset of data previously presented by Büntgen et al. (2008). Mean inter-series correlation (calculated in 31-year moving windows) is consistently higher than 0.6 between 1053 and 1996.

The tree-ring width RCS detrended record from Dachstein (Friedrich unpubl., Fig. 2e) is based on 60 *L. decidua* series with a minimum replication of 4 series in the 14th century. The BIT series from Dachstein (Fig. 2f) is based on 20 samples from the three conifer species *L. decidua*, *Picea abies* and *Pinus cembra*.

Comparison of temperature data (Böhm et al., 2010) and the previously unpublished proxy series reveals influence of temperature on all the proxies (Table 2). Correlation coefficients are significant ($p < 0.05$) for all proxies in the 1–31-year band and for the RCS detrended data in the frequency band >31-year. The series of Esper et al. (2007) and Friedrich (unpubl., trw RCS) show the highest correlations of all the proxies in the >31-year frequency band with correlation coefficients of $r = 0.67$ and $r = 0.77$ (both $p < 0.005$), respectively. In the 1–31-year band, highest correlations with temperature data are found for the BIT series ($r = 0.68$). However, the BIT series shows a massive loss in variance back in time.

Mean inter-series correlations are shown in Fig. 2i–k. For raw data and for the frequency band 1–31-year mean correlations increase through time and are highest after 1550. In the >31-year frequency band increased mean inter-series correlations are found centred on 1250, between 1500 and 1625 and after 1850, and low values are found centred on 1350 and between 1650 and 1750. The standard deviation among proxy series is high between 1150 and 1400 (Fig. 2j).

3.2. Calibration and model validation

Amplitude/RMSEP values indicate skill for all calibration models, with amplitudes more than six times larger than the errors (Table 3). Cross-validated RMSEP of summarised calibrations range between 0.76 °C and 0.84 °C (Table 3). The lowest RMSEP are found for variants 4 (two frequency bands) and 8 (four frequency bands), followed by the three remaining reconstructions based on calibration in two frequency bands. The unconstrained calibration of tree-ring data (variant 7) shows highest RMSEP (0.84 °C).

All nine calibration approaches underestimated (were colder than) the instrumental target between 1760 and 1850 and overestimated (were warmer than) it between 1860 and 1950 (Fig. 4f, Table 4). From 1760 to 1850, the offset between the reconstructions and the instrumental target was smaller for reconstructions with frequency-dependent calibration than for the reconstruction without frequency restrictions (differences in means: −0.23 °C and −0.38 °C, respectively, Table 4). Accordingly, RE and CE were higher for calibration in frequency bands than for unconstrained calibration (Table 5). The offset was largest between 1810 and 1821. It was most pronounced in the calibration based on CPS (Table 4). The overestimation 1860–1950 was larger than the underestimation 1760–1850 in six out of nine variants (Table 4).

Calibrations of individual frequency bands are shown in Fig. 4. Frequency constrained calibrations (Fig. 4a–d) showed similar progression and amplitudes as the calibration target, except for frequency bands >31 and >51-years (Fig. 4e, f), where multi-decadal changes generally agreed with the calibration target, but considerable differences in the centennial trends were found. The instrumental target was underestimated between 1760 and 1850 and overestimated between 1860 and 1950. Similarly, unconstrained tree-ring data underestimated temperatures from 1760 to 1850 and overestimated temperatures between 1860 and 1950.

3.3. Reconstruction

One unfiltered reconstruction (variant 5) is shown in Fig. 5a and the nine low-pass filtered reconstructions are presented in Fig. 5b. The highest temperatures prior to the 20th century were recorded in the 12th century, and were 0.3 °C higher than the 20th century mean. The lowest temperatures were recorded at the end of the 16th century and in the 14th century, and were 1 °C lower than the 20th century mean. Low-pass filtered (31-year Gaussian) reconstructed temperatures were highest (0.66 °C) in the late 20th century and exceeded the upper uncertainty limit (highest reconstructed temperature +2*RMSEP) of all reconstructions. When considering the target data by Böhm et al. (2010) available up to 2008 (including high temperatures in the 1990s and in the early 2000s), a 31-year filtered series not affected by edge effects is obtained up to the year 1993 and reaches a positive anomaly of 0.99 °C (wrt 1901–2000) in 1993 which is 0.75 °C higher than the warmest value (31-year filtered series) observed in the 12th century (Fig. 5b).

CONISS and the broken-stick model divided the nine (non-smoothed) summer temperature reconstructions into seven zones (periods) (Fig. 5c), delimited by the years 1171, 1379, 1573, 1610, 1812 and 1822. The first period is characterised by temperatures with mean values close to the mean of the 20th century (mean value 0 °C wrt 1901–2000). Temperature was lower in the second period which is characterized by a negative climate anomaly with mean values 0.57 °C lower than the 20th century mean. The third period was again warmer (−0.2 °C wrt 1901–2000) but interrupted by a cold anomaly around 1450. In the fourth period temperature minima of the last millennium were reached around 1600, showing negative anomalies of −0.9 °C compared to the 20th century. The fifth period is defined by temperatures 0.47 °C below the mean of the 20th century. This phase was followed by a cold anomaly about 1810–1830 (−1.63 °C wrt 1901–2000). The most recent period is comparable to the 20th century mean (−0.05 °C wrt 1901–2000).

When calculating the correlation between the consensus reconstruction (Fig. 5c) and the 31-year low-pass filtered input data, we find highest correlations for the series by Friedrich (unpubl., trw RCS, $r = 0.80$), Esper et al. (2007, $r = 0.70$) and Büntgen et al. (2009, $r = 0.68$). Each of the other input series correlate with the consensus reconstruction with a coefficient of $r = 0.55$.

Differences among the eight reconstructions based on frequency splitting are minor. In contrast, considerable differences exist between these eight reconstructions and the reconstruction based on tree-ring data without frequency splitting. Values in the late

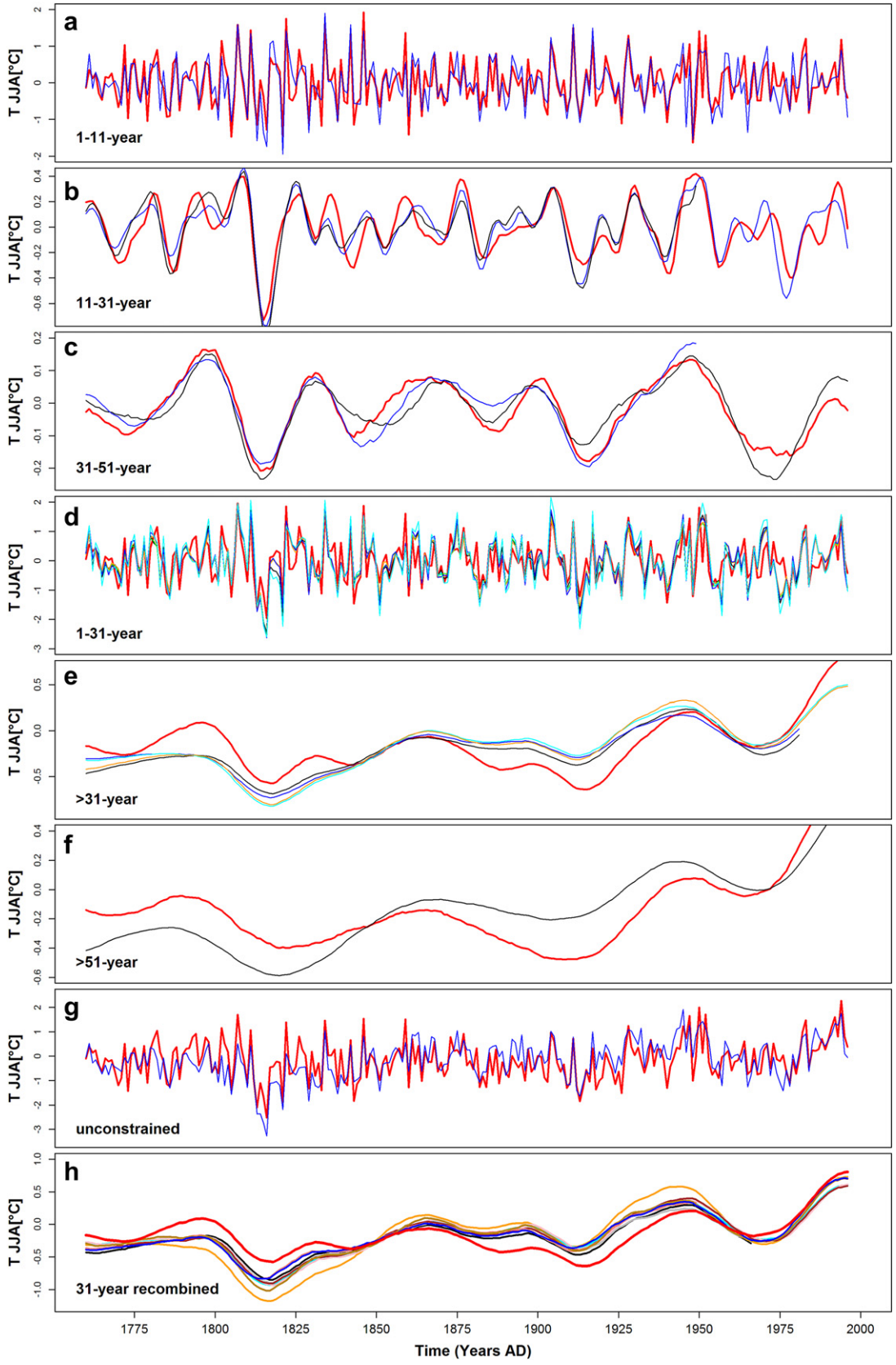

Fig. 4. Calibration in individual frequency bands: (a) 1–11-year, (b) 11–31-year, (c) 31–51-year, (d) 1–31-year, (e) >31-year, (f) >51-year. Panel (g) shows the calibration of unconstrained tree-ring data. In panel (h) the nine low pass-filtered (>31-year Gaussian) final reconstructions are shown. Instrumental data of Böhm et al. (2010) is shown in red. Colour codes of individual series are presented in Table 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Mean offsets among reconstructions and instrumental data (Böhm et al., 2010) for selected time windows.

Reconstruction variant	Raw data				Low-pass filtered data (30-year Gaussian)			
	1760–1850	1810–1821	1860–1950	1950–1996	1760–1850	1810–1821	1860–1950	1950–1996
1	−0.21	−0.92	0.20	NA	−0.20	−0.30	0.18	NA
2	−0.22	−0.87	0.17	NA	−0.21	−0.26	0.15	NA
3	−0.23	−0.88	0.25	−0.06	−0.22	−0.34	0.23	−0.07
4	−0.24	−0.86	0.24	−0.06	−0.23	−0.32	0.23	−0.07
5	−0.23	−0.90	0.26	−0.06	−0.21	−0.34	0.23	−0.07
6	−0.24	−1.18	0.26	−0.06	−0.23	−0.42	0.24	−0.06
7	−0.38	−0.96	0.37	−0.03	−0.37	−0.60	0.34	−0.03
8	−0.22	−0.73	0.22	−0.04	−0.20	−0.26	0.21	−0.03
9	−0.20	−0.74	0.22	NA	−0.19	−0.25	0.21	NA

13th and early 14th centuries and in the 16th and 17th centuries were lower in the single-band tree-ring reconstruction (variant 7) than in the reconstructions using frequency splitting.

Reconstructions including chironomids (variants 4 and 8) were indifferent to the number of frequency bands (two or four) or statistical model (OLSR or PLSR, Fig. 5b). Differences between reconstructions that include lake sediment data (mainly chironomids, variant 4 and 8), with associated age uncertainties, and reconstructions only including tree-ring data (variants 3, 5, and 6) occurred around 1360 and during the Maunder Minimum (1670–1715). Series including chironomids showed a more pronounced cooling than reconstructions based on frequency-weighted tree-ring series.

Variance stabilisation in the frequency band 1–31 years based on PLSR calibration (variant 5) did not affect the low-frequency behaviour (>31-year) of the reconstruction. In contrast, the reconstruction based on variance stabilised CPS in the 1–31-year band (variant 6) produced lower temperatures in the cold phases 1810–1821 and around 1600 than all other reconstructions based on frequency splitting (Fig. 5b).

3.4. Sensitivity tests

In general, the leave-out reconstructions appeared to be similar to the reconstructions presented earlier (referred to as 'full reconstructions'; Fig. 6). The highest reconstructed temperatures (31-year low-pass filter) were always recorded in the 20th century.

Between 1050 and 1400, most reconstructions show pronounced low-frequency variability (amplitude >1 °C). When excluding the record of Friedrich (unpubl., trw RCS) and especially that of Esper et al. (2007) reconstructions exhibit reduced low-frequency climate variability in the Medieval Period (amplitude <0.75 °C, <0.5 °C, respectively, Fig. 6).

The leave-out approach reveals phases of agreement and disagreement in decadal-scale climate variability. General agreement is detected for the phases 1050–1200, 1400–1620 and 1670 to the present. In these three phases the reconstructions differ in amplitude but show similar decadal-scale fluctuations (Fig. 6). In contrast, between 1200 and 1400 and in the 50 years between 1620 and 1670 no accordance of decadal-scale fluctuations of the leave-out reconstructions was found (Fig. 6d and e).

Comparing calibration in frequency bands (Fig. 6a,b,d,e) with calibration of unconstrained data (Fig. 6c and f), we find the reduction of amplitude predicted for the last millennium for the approaches with frequency splitting (variants 4 and 5). The spread among leave-out reconstructions is also reduced when applying the split frequency approach. For example, for the year 1070 the range of the reconstructions amounts to 0.8 °C for the frequency constrained reconstructions (Fig. 6a and b) compared to 1.5 °C for the calibration of raw data (Fig. 6c).

4. Discussion

4.1. Reconstructions

The Greater Alpine region represents a relatively small portion of the Northern Hemisphere extra-tropics. Still, there is general agreement between five northern hemisphere reconstructions and the reconstruction presented in this study, which does not include any proxy data used in the Northern Hemisphere reconstructions (Fig. 7a). Reconstructions indicate warmer conditions at the beginning of the last millennium, a subsequent cooling with coldest temperatures mostly in the 17th century and a warming trend between 1600 and 1800. Although we only compare values between 1053 and 1980, the correlations between the reconstructions (31-year Gaussian low-pass filter) are all significant ($p < 0.05$) and range between 0.56 and 0.62. This suggests common low-frequency signal among the five selected large-scale reconstructions and our Alpine reconstruction.

There was also agreement on the regional scale. For the period 1053–1996 the fully independent reconstruction by Glaser and Riemann (2009), the reconstruction of Büntgen et al. (2011), which includes some data from the reconstruction presented in this study, and the reconstruction presented here show common variance back to 1550 (Fig. 7b–e). They show a warm period around 1575, a pronounced cold phase around 1600, the cold phase of the Late Maunder Minimum (1675–1710) and a subsequent warming culminating around 1780. Focussing on the last 500 years, we also found reasonable agreement with decadal-

Table 5
Reduction of Error (RE) and Coefficient of Efficiency (CE) statistics of the nine calibration variants for the indicated calibration and verification periods.

Calibration period	Validation period		Variant 1[a]	Variant 2[a]	Variant 3	Variant 4	Variant 5	Variant 6	Variant 7	Variant 8	Variant 9[b]
1760–1877	1878–1996	RE	0.31	0.32	0.44	0.44	0.43	0.42	0.14	0.50	0.39
1760–1877	1878–1996	CE	0.31	0.32	0.43	0.43	0.43	0.41	0.13	0.50	0.39
1878–1996	1760–1877	RE	0.52	0.55	0.48	0.52	0.54	0.50	0.16	0.55	0.50
1878–1996	1760–1877	CE	0.52	0.55	0.48	0.51	0.53	0.49	0.15	0.54	0.49

[a] Calibration periods 1776–1877 and 1878–1981, respectively and verification periods 1878–1981 and 1776–1877, respectively.
[b] Calibration periods 1760–1855 and 1856–1949, respectively and verification periods 1878–1981 and 1776–1877, respectively.

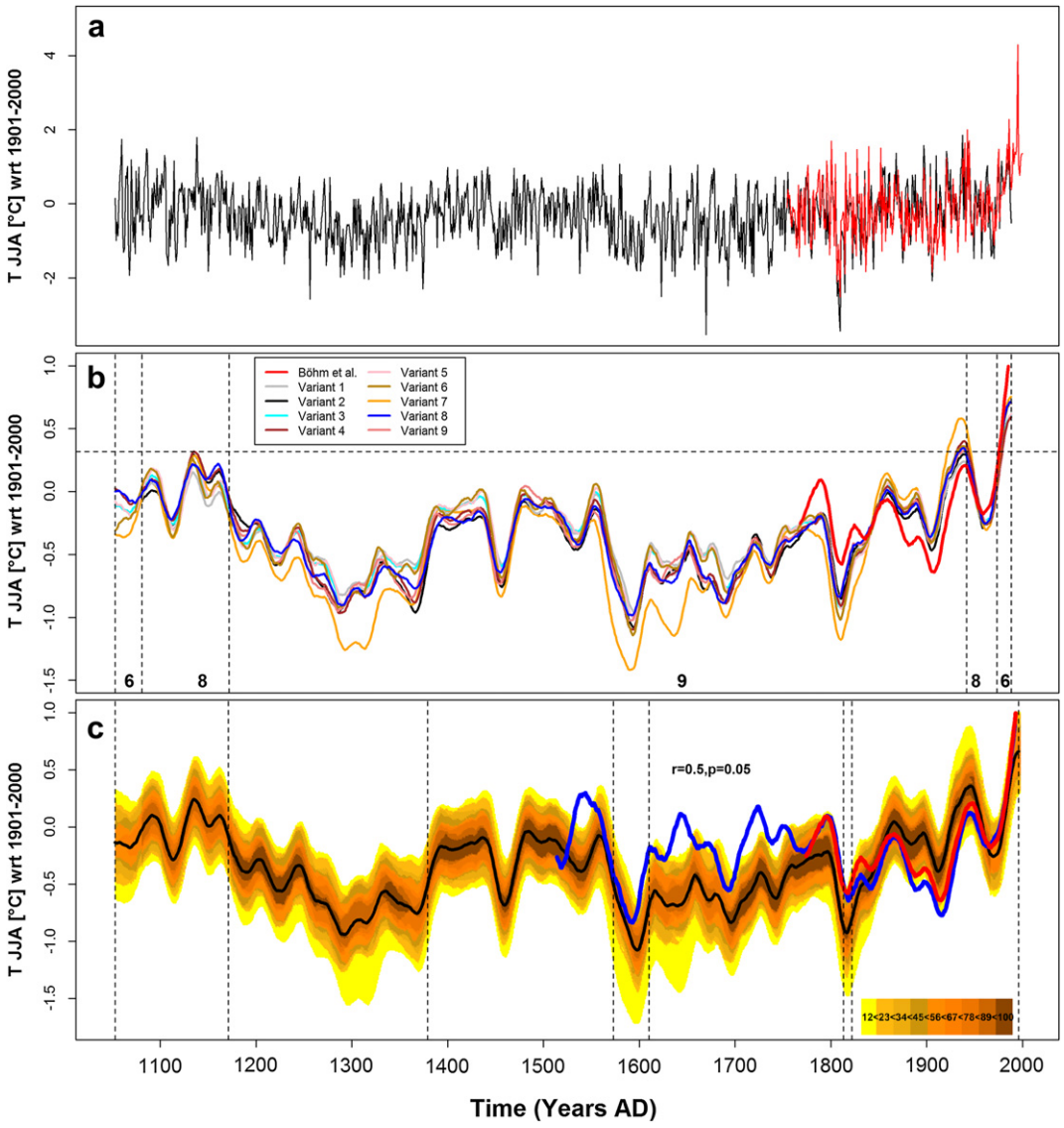

Fig. 5. Reconstructions covering the period 1053–1996 (for details on variants see Table 2) along with instrumental data (Böhm et al., 2010, red): (a) Annually resolved reconstruction (variant 5, black), (b) Nine low-pass filtered (31-year Gaussian) reconstructions, dashed horizontal line indicates the warmest temperatures reconstructed in the 12th century, dashed vertical lines indicate changes in number of reconstructions (corresponding to numbers indicated), (c) overlap of the nine reconstructions (see text for explanation). Documentary data of Dobrovolný et al. (2010) are indicated in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

scale climate variability reported for Central Europe (Dobrovolný et al., 2010): $r = 0.50$ ($p < 0.05$) for 31-year low-pass (Gaussian) filtered temperatures, although the two reconstructions had an offset of 0.45 °C between 1600 and 1760 (Fig. 5c). Part of this offset might be explained by limitations inherent to documentary data, which might not pick up the full range of low-frequency climate variability (see e.g. Brázdil et al., 2010; Leijonhufvud et al., 2010; Zorita et al., 2010). Part of this offset might, however, also be caused by variants 1–9 that overestimate low-frequency climate variability in the calibration and, potentially, reconstruction periods. There was, for example, disagreement between natural proxies and documentary proxies between 1520 and 1550. Part of this disagreement might be explained by numerical restrictions (see above). Trachsel et al. (2010), however, hypothesised that this disagreement might also be caused by the absence of strong forcing during the first half of the 16th century, resulting in a weak proxy signal.

Disagreement was also found prior to 1550. The timing of minimum and maximum temperatures between 1050 and 1500 in five recent reconstructions covering the Alpine area differed considerably. While Corona et al. (2010) found maximum temperatures between 1210 and 1235, Corona et al. (2011) reconstructed temperature maxima at 1235 and 1400. Glaser and Riemann (2009) found maximum temperatures at 1170 and 1275 whereas Büntgen et al. (2011) reconstructed maximum temperatures at 1160, 1440 and 1565. In our consensus reconstruction maximum temperatures were found between 1125 and 1160. The disagreement before 1450 might be to some extent caused by reduced replication of tree-ring series (Fig. 3, e.g. Frank et al., 2007a).

Theil-Sen trend estimates for 31-year time periods that summarise mid-term fluctuations and disregard low-frequency variability showed general agreement back to 1450, warming centred around 1375 and cooling centred around 1185, 1280 and 1350 for the reconstructions of Glaser and Riemann (2009), the

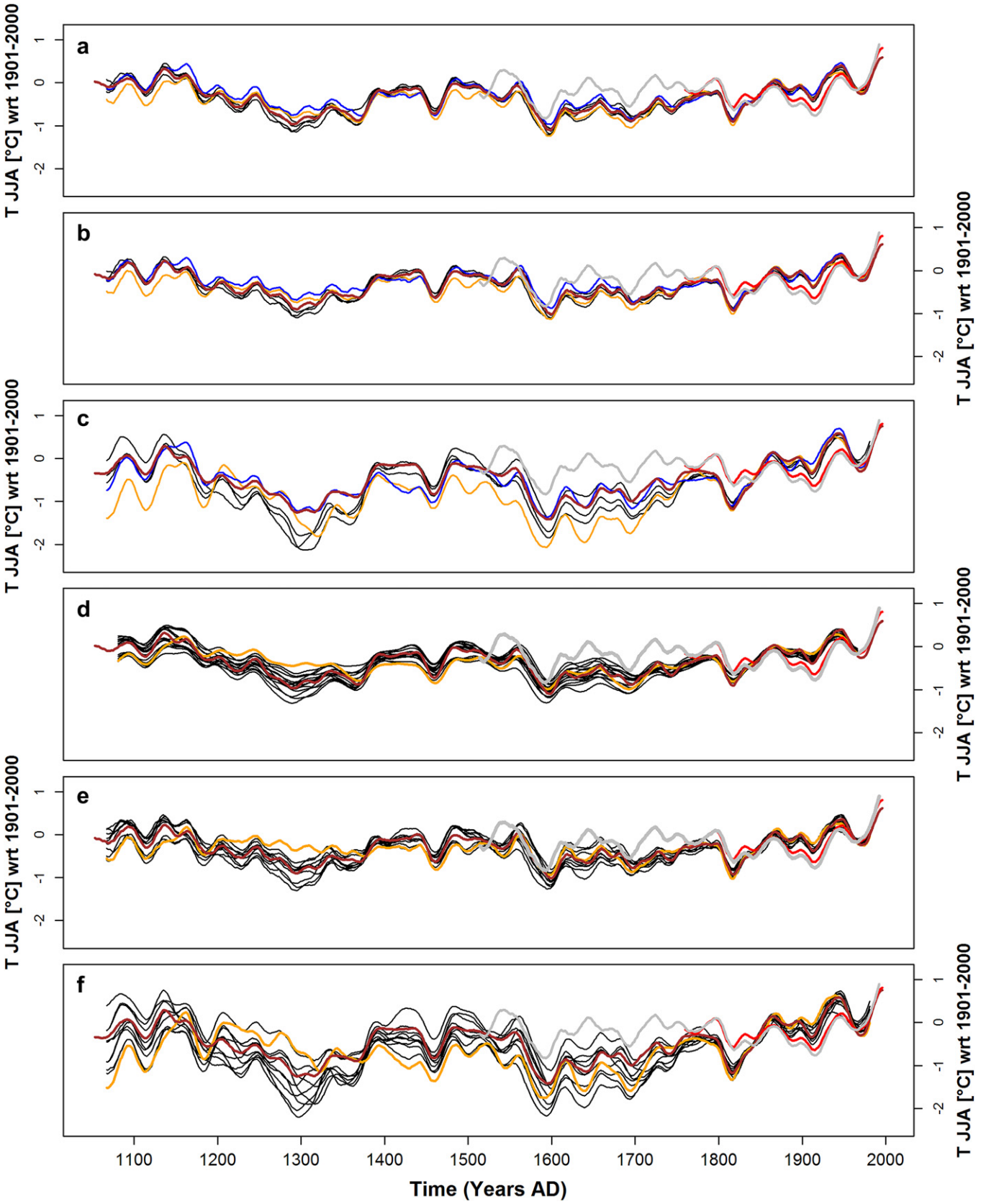

Fig. 6. Results of sensitivity experiments: comparison of reconstructions including all available proxies with reconstructions omitting one (a–c) and two (d–f) proxy time series, respectively. The data shown are low-pass filtered (31-year Gaussian) leave-one and leave-two-out reconstructions of variants 4 (a,d), 5 (b,e) and 7 (c,f). All panels: brown lines: reconstructions when including all proxies; red lines: instrumental data of Böhm et al. (2010); grey lines: documentary data of Dobrovolný et al. (2010); (panels a–c): orange and blue lines: reconstructions when leaving out the proxies of Esper et al. (2007) and Friedrich (unpubl., trw RCS), respectively. (d–f) Orange line: reconstruction when leaving out the reconstructions of Esper et al. (2007) and Friedrich (unpubl., trw RCS). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

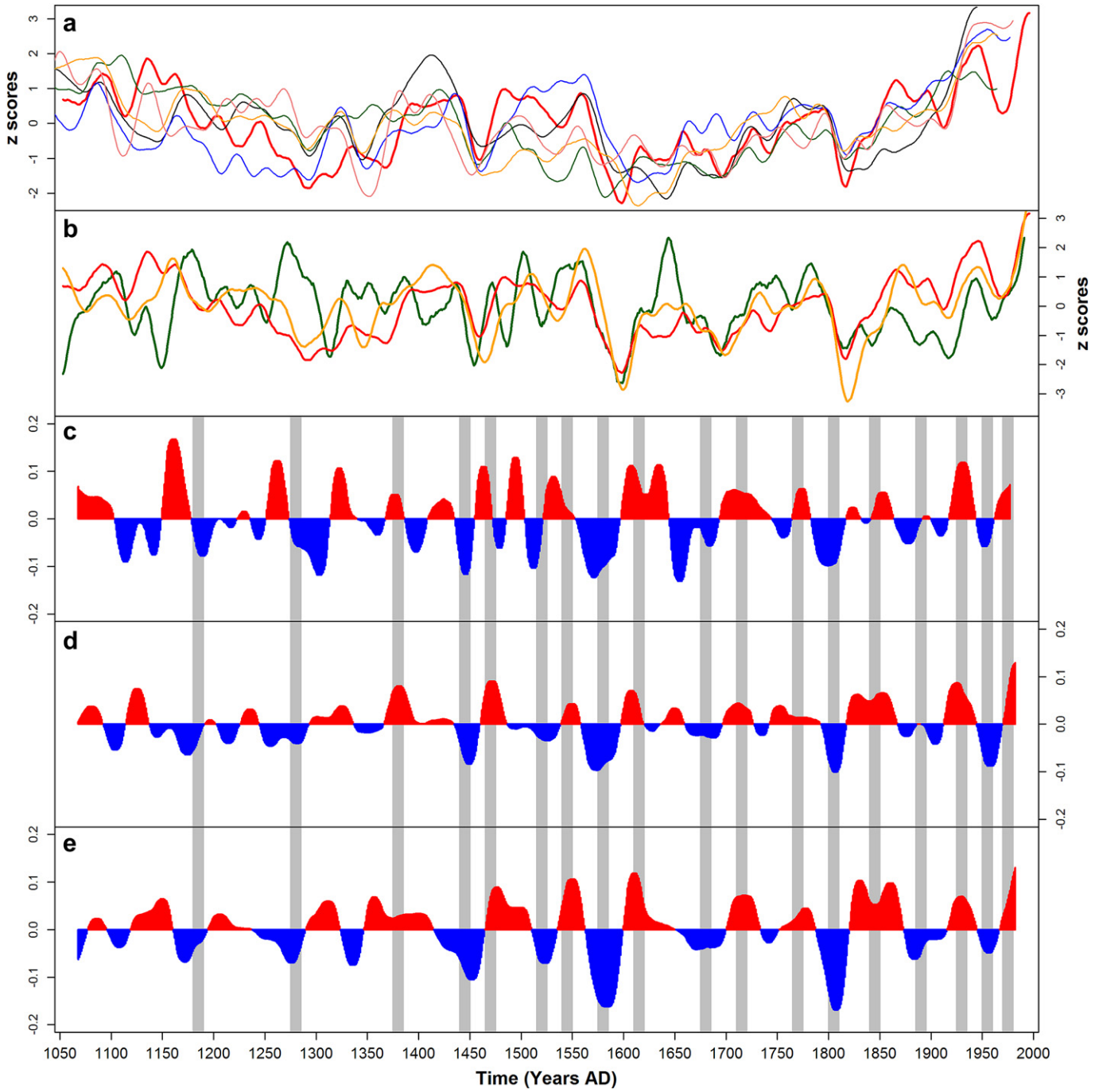

Fig. 7. Temperature reconstructions for hemispheric and regional scales. (a) Northern Hemisphere reconstructions of Juckes et al. (2007, green), Frank et al. (2007a, blue), Moberg et al. (2005, orange), Hegerl et al. (2007, black) and Mann et al. (2008, black) and Alpine consensus reconstruction (this study, red) (b) Alpine reconstruction (consensus, red), reconstructions of Büntgen et al. (2011, orange) and Glaser and Riemann (2009, darkgreen); 31-year Theil-Sen trend estimates of *z*-scores of the Glaser and Riemann (2009) reconstruction (c), the Alpine reconstruction (d), and the Büntgen et al. (2011) reconstruction (e); grey bars indicate phases of agreement between Theil-Sen trend estimates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reconstruction of Büntgen et al. (2011) and the reconstruction presented here (Fig. 7c–e).

A comparison to glacier length variations is problematic, because glaciers do not respond to summer temperature alone (e.g. Oerlemans, 2001) and respond with delays. This might be a reason why a comparison between phases of glacier advance and retreat in the Alps (Holzhauser et al., 2005) and summer temperature reconstructions for the Alpine area did not clearly favour any of the aforementioned reconstructions. Glaciers advancing throughout the 14th century do not support reconstructions that show warm temperatures for this phase (Corona et al., 2010, 2011, and partly; Glaser and Riemann, 2009), whereas glacier high stands at the end of the 12th century cast some doubt on maximum temperatures reconstructed in the 12th century (Büntgen et al., 2011, and our consensus reconstruction) and are supporting the data by Glaser and Riemann (2009) who reconstruct a cold phase around 1160 and a subsequent warming that might have stopped the glacier advance.

Differences between reconstructions may originate from two sources: (i) differences in proxy data used and (ii) differences in calibration methods (Juckes et al., 2007). Since the sensitivity tests did not affect the overall shape of the reconstruction and had no influence on the above average temperatures reconstructed for the 12th century (see next section), it is most likely differences in proxy data that caused disagreement between 1053 and 1600.

4.2. Calibration and sensitivity tests

Differences among the three different statistical methods (CPS, OLSR and PLSR) were minor (Table 3, Figs. 4h and 5b). The similarity of reconstructions obtained from different calibration methods is indicative of a stable calibration. Stability of calibration and reconstructions points to high accordance among proxies during the calibration period and to similar weights assigned to proxies by OLSR and PLSR.

Cross-validated RMSEP favoured approaches based on frequency splitting over the approach without frequency splitting. Calibration in specific frequency bands possibly alters the weight assigned to individual proxy series which in turn alters the RMSEP. Calibration in different frequency bands is, however, challenged by the increase in serial autocorrelation and, hence, the loss of degrees of freedom introduced by low-pass filtering (e.g. Trenberth, 1984; Yiou, 2010). Since RMSEP of series calibrated in two frequency bands was similar to RMSEP for calibration in four frequency bands (Table 3), we follow the rule of parsimony and favour calibration in only two frequency bands.

The result of the leave-out experiment points to increased uncertainty of the reconstructions prior to 1400 (Fig. 5a–f). Differences in amplitude when excluding specific proxy series (Friedrich unpubl., trw RCS and Esper et al., 2007) highlight the importance of data included in calibration/reconstruction approaches. Hence highly accurate proxy records remain the fundamental prerequisite for meaningful (multi-archive) climate reconstructions. In general the leave-out experiment revealed the strong dependence of the reconstruction on the data included during calibration. Since the calibrations did not differ considerably, this points to inconsistencies among proxies included in the reconstruction between 1053 and 1760 and to increased inconsistencies between 1050 and 1450, which is confirmed by lower mean correlations and higher standard deviations in the >31-year band (Fig. 2j). These changes might be partly caused by reduced replication of tree-ring series (Fig. 3), which in turn results in increased uncertainty of the tree-ring series (e.g. Frank et al., 2007a).

Between 1200 and 1400 the leave-out approach further revealed differences of decadal-scale variability of proxy series which points to inconsistencies in the higher frequency (<31-year) variability of the proxy data, which seems at least in part to be affected by the reduced replication inherent to tree-ring records used in this study.

In our study, calibrating in different frequency bands consistently resulted in reduced low-frequency climate variability. The reduction of low-frequency with frequency-dependent calibration in our study was caused by the reduced offset between the reconstructions and the data of Böhm et al. (2010), which was achieved by calibration in frequency bands. The effect of this offset on the low-frequency component of climate reconstructions is discussed in detail by Frank et al. (2007b). It alters the amplitude of low-frequency climate variability by as much as 1 °C (also found in this study, data not shown). In the leave-out experiments the lowest reconstructed temperatures in cold anomalies were 0.5 °C warmer when calibrating in different frequency bands compared to unconstrained calibrations.

5. Conclusions

In this study we presented a multi-archive reconstruction of summer temperature for the European Alps covering the last millennium based on nine different calibration approaches and conservative, but not all-embracing error estimation. Warmest summers 0.3 °C warmer than the 20th century mean occurred between 1050 and 1200. These temperatures were, however, 0.35 °C lower than temperatures in the last decade of the 20th century, though uncertainties during the early period of our reconstruction are likely increased due to reduced replication (number of individual measurement series) inherent to some of the proxy records used here.

The lowest temperatures occurred at the turn of the 16th to the 17th century and were about 1 °C colder than the 20th century mean (31-year low-pass filtered data). Comparison to hemispheric-scale reconstructions revealed similarities in multi-decadal to centennial scale climate variability.

Sensitivities of our multi-archive reconstructions to different calibration approaches and the inclusion or exclusion of proxies were carefully assessed. These sensitivity experiments suggested higher reconstruction uncertainties prior to 1400, mainly caused by inconsistencies among proxies included in this study. Hence, production of high-quality, highly precise and highly accurate proxies remains one of the key tasks in palaeoclimatology.

Acknowledgements

Funding was provided by the EU-FP6 IP 'Millennium' (017008) and the Swiss National Science Foundation (NCCR Climate, Grant NF 200021-116005/1 and a personal grant to MT). AM was funded by grants from the Swedish Research Council, VR. JE was funded by Mainz Excellence Cluster and Palaeoweather Group. We would like to thank three anonymous reviewers for comments that greatly improved the clarity of this manuscript. Gridded temperature data was downloaded from: http://www.zamg.ac.at/histalp/content/view/36/1/index.html. This is publication no. A397 from the Bjerknes Centre for Climate Research.

Appendix. A. Sensitivity to the early instrumental warm-bias.

All calibration approaches underestimated (were too low) temperatures from 1760 to 1850 and overestimated (were too high) temperatures between 1860 and 1950. We exemplify the sensitivity of our calibration models to this bias with variant seven (tree-rings only, single-band). The calibration period was restricted to 1950–1996, i.e. the time-period where calibrated proxies and instrumental target are in close agreement and free of offset (Fig. 4, Table 3). The reconstruction focused on the period from 1760 to 1950, either including or excluding the pronounced offset between 1810 and 1821. We still found a mean offset of −0.49 °C between 1760 and 1850, but rather different mean offsets of +0.36 °C and of −0.39 °C between 1860 and 1950 after including or excluding the years 1810–1821, respectively.

The offset between early instrumental data (1760–1840) and proxy-based reconstructions is discussed in Frank et al. (2007b). It seems that in many early instrumental series summer temperatures are overestimated because thermometers were not adequately shaded. Early instrumental data for the greater Alpine region have been re-homogenised by Böhm et al. (2010), but there is still some room for uncertainty and a remaining warm-bias in the early instrumental summer temperature series cannot be ruled-out. A certain amount of the offset might as well be attributable to uncertainties in proxy data. For example, the largest offset (1810–1821) was probably caused by tree-ring width data, which

are known to underestimate the temperature of cold events (i.e. show too cold temperatures) and to show inadequately long cold phases because of biological persistence (e.g. Frank et al., 2007b). Using the data of Böhm et al. (2010) as calibration target, we still found an offset between early instrumental and proxy data, as do Corona et al. (2011). This offset was reduced when applying our calibration approach with different frequency bands. In this case, the offset becomes considerably lower than described by Corona et al. (2011; we found an offset of 0.2 °C compared to their 0.5 °C).

In contrast to the 18th and early 19th century, the measurement network became denser after 1860 and the temperature measurements are considered to be more accurate (e.g. Frank et al., 2007b). Still, we found an offset (overestimation) between reconstructions and instrumental data from 1860 to 1950 (similar to Büntgen et al., 2006). This offset was as large as or even larger than the offset with early instrumental data (Table 4). Calibration based on tree-ring data and no frequency separation (variant 7) in the period 1950–1996 resulted in similar offsets for the periods 1760–1850 and 1860–1950 (although calibration based on unconstrained tree-ring data had poor calibration statistics, it was the only calibration approach allowing for such a comparison, since it retained enough degrees of freedom for the period 1950–1996).

Although we can not conclude from this experiment that the early instrumental period correction applied by Böhm et al. (2010) is accurate, we can clearly show that the early instrumental data is not the only possible source of error. The offset between instrumental data and 'reconstruction' between 1860 and 1950 in our experiment as well suggests proxy data as a possible source of the offset.

References

Bennett, K.D., 1996. Determination of the number of zones in a biostratigraphical sequence. New Phytologist 132, 155–170.

Birks, H.J.B., 1998. Numerical tools in palaeolimnology: progress, potentials, and problems. Journal of Paleolimnology 20, 307–332.

Blass, A., Bigler, C., Grosjean, M., Sturm, M., 2007. Decadal-scale autumn temperature reconstruction back to AD 1580 inferred from the varved sediments of Lake Silvaplana (southeastern Swiss Alps). Quaternary Research 68, 184–195.

Böhm, R., Jones, P.D., Hiebl, J., Brunetti, M., Frank, D., Maugeri, M., 2010. The early instrumental warm bias: a solution for long central European temperatures series 1760–2007. Climatic Change 101, 41–67.

Boucher, E., Guiot, J., Chapron, E., 2011. A millennial multi-proxy reconstruction of summer PDSI for Southern South America. Climate of the Past 7, 957–974.

Brázdil, R., Pfister, C., Wanner, H., von Storch, H., Luterbacher, J., 2005. Historical climatology in Europe – the state of the art. Climatic Change 70, 363–430.

Brázdil, R., Dobrovolný, P., Luterbacher, J., Moberg, A., Pfister, C., Wheeler, D., Zorita, E., 2010. European climate of the past 500 years: new challenges for historical climatology. Climatic Change 101, 7–40.

Büntgen, U., Esper, J., Frank, D.C., Nicolussi, K., Schmidhalter, M., 2005. A 1052-year tree-ring proxy for Alpine summer temperatures. Climate Dynamics 25, 14153.

Büntgen, U., Frank, D.C., Nievergelt, D., Esper, J., 2006. Summer temperature variations in the European Alps, A.D. 755–2004. Journal of Climate 19, 5606–5623.

Büntgen, U., Frank, D.C., Wilson, R.J.S., Carrer, M., Urbinati, C., Esper, J., 2008. Testing for tree-ring divergence in the European Alps. Global Change Biology 14, 2243–2453.

Büntgen, U., Frank, D., Carrer, M., Urbinati, C., Esper, J., 2009. Improving Alpine summer temperature reconstructions by increasing sample size. In: Kazcka, R., et al. (Eds.), Tree Rings in Archaeology, Climatology and Ecology (TRACE), vol. 7, pp. 36–43.

Büntgen, U., Tegel, W., Nicolussi, K., McCormick, M., Frank, D., Trouet, V., Kaplan, J.O., Herzig, F., Heussner, K.U., Wanner, H., Luterbacher, J., Esper, J., 2011. 2500 years Of European climate variability and human susceptibility. Science 331, 578–582.

Bürger, G., Fast, I., Cubasch, U., 2006. Climate reconstruction by regression – 32 variations on a theme. Tellus Series A: Dynamic Meteorology and Oceanography 58, 227–235.

Burman, P., Chow, E., Nolan, D., 1994. A cross-validatory method for dependent data. Biometrika 81, 351–358.

Campbell, R., McCarroll, D., Loader, N.J., Grudd, H., Robertson, I., Jalkanen, R., 2007. Blue reflectance in *Pinus sylvestris*: application, validation and climatic sensitivity of a new palaeoclimate proxy for tree ring research. The Holocene 17, 821–828.

Campbell, R., McCarroll, D., Robertson, I., Loader, N.J., Grudd, H., Gunnarson, B., 2011. Blue intensity in *Pinus sylvestris* tree rings: a manual for a new palaeoclimate proxy. Tree-Ring Research 67, 127–134.

Casty, C., Wanner, H., Luterbacher, J., Esper, J., Böhm, R., 2005. Temperature and precipitation variability in the European Alps since AD 1500. International Journal of Climatology 25, 1855–1880.

Cook, E.R., Briffa, K.R., Jones, P.D., 1994. Spatial regression methods in dendroclimatology: a review and comparison of two techniques. International Journal of Climatology 14, 379–402.

Cook, E.R., Briffa, K.R., Meko, D.M., Graybill, A., Funkhouser, G., 1995. The 'segment length curse' in long tree-ring chronology development for palaeoclimatic studies. Holocene 5, 229–237.

Corona, C., Guiot, J., Edouard, J.L., Chalié, F., Büntgen, U., Nola, P., Urbinati, C., 2010. Millennium-long summer temperature variations in the European Alps as reconstructed from tree rings. Climate of the Past 6, 379–400.

Corona, C., Edouard, J.-L., Guibal, F., Guiot, J., Bernard, S., Thomas, A., Denelle, N., 2011. Long-term summer (AD 751–2008) temperature fluctuation in the French Alps based on tree-ring data. Boreas 40, 351–366.

Dawdy, D.R., Matalas, N.C., 1964. Statistical and probability analysis of hydrologic data, part III: analysis of variance, covariance and time series. In: Chow, V.T. (Ed.), Handbook of Applied Hydrology, A Compendium of Water-Resources Technology. McGraw-Hill, New York, pp. 8.68–8.90.

de Jong, S., 1993. PLS fits closer than PCR. Journal of Chemometrics 7, 551–557.

Dobrovolný, P., Moberg, A., Brázdil, R., Pfister, C., Glaser, R., Wilson, R., van Engelen, A., Limanowka, D., Kiss, A., Halíčková, M., Macková, J., Riemann, D., Luterbacher, J., Böhm, R., 2010. Monthly and seasonal temperature reconstructions for Central Europe derived from documentary evidence and instrumental records since AD 1500. Climatic Change 101, 69–107.

Efron, B., Gong, G., 1983. A leisurely look at the Bootstrap, the Jackknife, and cross-validation. The American Statistician 37, 36–48.

Esper, J., Cook, E.R., Krusic, P.J., Peters, K., Schweingruber, F.H., 2003. Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies. Tree-Ring Research 59, 81–98.

Esper, J., Frank, D.C., Wilson, R.J.S., Briffa, K.R., 2005. Effect of scaling and regression on reconstructed temperature amplitude for the past millennium. Geophysical Research Letters 32, L07711.

Esper, J., Büntgen, U., Frank, D., Pichler, T., Nicolussi, K., 2007. Updating the Tyrol treering dataset. In: Haneca, K., et al. (Eds.), Tree Rings in Archaeology, Climatology and Ecology (TRACE), vol. 5, pp. 80–85.

Frank, D., Esper, J., Cook, E.R., 2007a. Adjustment for proxy number and coherence in a large-scale temperature reconstruction. Geophysical Research Letters 34, L16709.

Frank, D., Büntgen, U., Böhm, R., Maugeri, M., Esper, J., 2007b. Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. Quaternary Science Reviews 26, 3298–3310.

Frank, D.C., Esper, J., Raible, C.C., Büntgen, U., Trouet, V., Stocker, B., Joos, F., 2010. Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate. Nature 463, 527–530.

Ge, Q.S., Zheng, J.Y., Hao, Z.X., Shao, X.M., Wang, W.C., Luterbacher, J., 2010. Temperature variation through 2000 years in China: an uncertainty analysis of reconstruction and regional difference. Geophysical Research Letters 37, L03703.

Glaser, R., Riemann, D., 2009. A thousand-year record of temperature variations for Germany and Central Europe based on documentary data. Journal of Quaternary Science 24, 437–449.

Graham, N.E., Hughes, M.K., Ammann, C.M., Cobb, K.M., Hoerling, M.P., Kennett, D.J., Kennett, J.P., Rein, B., Stott, L., Wigand, P.E., Xu, T., 2007. Tropical Pacific – mid-latitude teleconnections in medieval times. Climatic Change 83, 241–285.

Grimm, E.C., 1987. CONISS: a FORTRAN 77 program for stratigraphically constrained cluster analysis by the method of incremental sum of squares. Computers and Geosciences 13, 13–35.

Guiot, J., Corona, C., ESCARSEL members, 2010. Growing season temperatures in Europe and climate forcings over the past 1400 Years. PLOS One 5, 1–15.

Guiot, J., Nicault, A., Rathgeber, C., Edouard, J.L., Guibal, F., Pichard, G., Till, C., 2005. Last-millennium summer-temperature variations in western Europe based on proxy data. The Holocene 15, 489–500.

Hegerl, G., Crowley, T., Hyde, W., Frame, D., 2006. Climate sensitivity constrained by temperature reconstructions over the past seven centuries. Nature 440, 1029–1032.

Hegerl, G.C., Crowley, T.J., Allen, M., Hyde, W.T., Pollack, H.N., Smerdon, J., Zorita, E., 2007. Detection of human influence on a new, validated 1500-year temperature reconstruction. Journal of Climate 20, 650–666.

Hegerl, G., Luterbacher, J., González-Rouco, F., Tett, S.F.B., Crowley, T., Xoplaki, E., 2011. Influence of human and natural forcing on European seasonal temperatures. Nature Geoscience 4, 99–103.

Holzhauser, H., Magny, M., Zumbühl, H.J., 2005. Glacier and lake-level variations in west-central Europe over the last 3500 years. Holocene 15, 789–801.

Jansen, E., Overpeck, J., Briffa, K.R., Duplessy, J.C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W.R., Rahmstorf, S., Ramesh, R., Raynaud, D., Rind, D., Solomina, O., Villalba, R., Zhang, D., 2007. Palaeoclimate. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Juckes, M.N., Allen, M.R., Briffa, K.R., C, J., Hegerl, G.C., Moberg, A., Osborn, T.J., Weber, S.L., 2007. Millennial temperature reconstruction intercomparison and evaluation. Climate of the Past 3, 591–609.
Juggins, S., 2009. RIOJA: analysis of Quaternary Science data, version 0.5–6. http://cran.r-project.org/web/packages/rioja/index.html.
Larocque-Tobler, I., Grosjean, M., Heiri, O., Trachsel, M., Kamenik, C., 2010. 1000 years Of climate change reconstructed from chironomid subfossils preserved in varved-lake Silvaplana, Engadine, Switzerland. Quaternary Science Reviews 29, 1940–1949.
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.
Legendre, P., Legendre, L., 1998. Numerical Ecology, second ed. Elsevier, Amsterdam, 853 pp.
Leijonhufvud, L., Wilson, R.J.S., Moberg, A., Söderberg, J., Retsö, D., Söderlind, U., 2010. Five centuries of Stockholm winter/spring temperatures reconstructed from documentary evidence and instrumental observations. Climatic Change 101, 109–141.
Mann, M.E., 2008. Smoothing of climate time series revisited. Geophysical Research Letters 35, L16708.
Mann, M.E., Zhang, Z., Hughes, M.K., Bradley, R.S., Miller, S.K., Rutherford, S., Ni, F., 2008. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. Proceedings of the National Academy of Sciences of the United States of America 105, 13252–13257.
Martens, H., Naes, T., 1989. Multivariate Calibration. Wiley, Chichester, 423 pp.
McCarroll, D., Pettigrew, E., Luckman, A., Guibal, F., Edouard, J.L., 2002. Blue reflectance provides a surrogate for latewood density of high-latitude pine tree rings. Arctic, Antarctic and Alpine Research 34, 450–453.
McCarroll, D., 2010. Future climate change and the British Quaternary research community. Quaternary Science Reviews 29, 1661–1672.
Moberg, A., Sonechkin, D.M., Holmgren, K., Datsenko, N.M., Karlén, W., 2005. Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. Nature 433, 613–617.
Oerlemans, J., 2001. Glaciers and Climate Change. A.A. Balkema, Lisse, etc, 148 pp.
Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2011. vegan: community ecology package, version1.17-11. http://cran.r-project.org/web/packages/vegan/index.html.
R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
Ramsey, F., Schafer, D., 2002. The Statistical Sleuth: A Course in Methods of Data Analysis. Duxbury Press, Belmont.
Riedwyl, N., Luterbacher, J., Wanner, H., 2008. An ensemble of European summer and winter temperature reconstructions back to 1500. Geophysical Research Letters 35, L20707.
Rutherford, S., Mann, M.E., Osborn, T.J., Bradley, R.S., Briffa, K.R., Hughes, M.K., Jones, P.D., 2005. Proxy-based Northern Hemisphere surface temperature reconstructions: sensitivity to method, predictor network, target season, and target domain. Journal of Climate 18, 2308–2329.
Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association 63, 1379–1389.
Telford, R., Birks, H.J.B., 2009. Evaluation of transfer functions in spatially structured environments. Quaternary Science Reviews 28, 1309–1316.
Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis. I, II, III. Koninklijke nederlandse akademie van wetenschappen Proceedings 53, 386–392. 521–525, 1397–1412.
Trachsel, M., Grosjean, M., Larocque-Tobler, I., Schwikowski, M., Blass, A., Sturm, M., 2010. Quantitative summer temperature reconstruction derived from a combined biogenic Si and chironomid record from varved sediments of Lake Silvaplana (south-eastern Swiss Alps) back to AD 1177. Quaternary Science Reviews 29, 2719–2730.
Trenberth, K.E., 1984. Some effects of finite sample size and persistence on meteorological statistics. Part I: Autocorrelations. Monthly Weather Review 112, 2359–2368.
Warnes, G.R., 2010. gtools: various R programming tools, version 2.6.2. http://cran.r-project.org/web/packages/gtools/index.html.
Wehrens, R., Mevik, B.-H., 2007. pls: Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR), version 2.1-0. http://cran.r-project.org/web/packages/pls/index.html.
Yamazaki, Y.H., Hind, A., Moberg, A., Yamazaki, K., Aina, T., Thurston, M., Bellouin, N., Hanappe, P., Allen, M., McCarroll, D., 2011. Can we constrain future climate predictions with European paleoclimate data of the last millennium? Geophysical Research Abstracts 13, EGU2011–12569.
Yiou, P., 2010. Statistical issues about solar–climate relations. Climate of the Past 6, 565–573.
Zorita, E., Moberg, A., Leijonhufvud, L., Wilson, R., Brázdil, R., Dobrovolný, P., Luterbacher, J., Böhm, R., Pfister, C., Glaser, R., Söderberg, J., González-Rouco, F., 2010. European temperature records of the past five centuries based on documentary information compared to climate simulations. Climatic Change 101, 143–168.