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

M. Trachsel a,b,*, M. Grosjean a,b, I. Larocque-Tobler a,b,c, M. Schwikowski b,d, A. Blass e, M. Sturm e

a Department of Geography, University of Bern, Erlachstrasse 9a, 3012 Bern, Switzerland
b Oeschger Center for Climate Change Research, University of Bern, Zähringerstrasse 25, 3012 Bern, Switzerland
c INRS-ETE, 490 De La Couronne, Québec G1K 9A9, Canada
d Department of Chemistry and Biochemistry, University of Bern & Paul Scherrer Institut, 5232 Villigen, Switzerland
e Surface Waters, EAWAG, Überlandstrasse 133, 8600 Dübendorf, Switzerland

1. Introduction

High-resolution, well calibrated temperature reconstructions are fundamental to put the current warming into a wider perspective. In this context, lake sediments are very important because they may provide very long records (e.g. Sturm and Lotter, 1995; Zolitschka et al., 2000) and preserve the signals of low-frequency climate variability (e.g. Moberg et al., 2005; Jansen et al., 2007). Typically, quantitative temperature reconstructions from lake sediments are based on biological proxies (e.g. chironomids, Heiri et al., 2003) using a transfer function (e.g. Birks, 1998). In the recent past considerable efforts have been made to quantitatively assess the range, trends and amplitude of paleoclimate variability.
from high-resolution (annually resolved) geochemical proxies in lake sediments using methods from tree ring research (calibration-in-time approach; Francus et al., 2002; Kalugn et al., 2007; Blass et al., 2007a, b; McKay et al., 2008; Trachsel et al., 2008; von Gunten et al., 2009; Kaufman et al., 2009). The quantitative assessment of paleoclimate parameters from geochemical lake sediment proxies may, however, be challenged by the complex and often non-linear responses of lacustrine systems to climate change (e.g. Ohlendorf, 1998), possible anthropogenic influences (e.g. Gobet et al., 2004; Geiszoëtt et al., 2009), difficulties with precise dating of lake sediments (e.g. Goslar et al., 2009; von Gunten et al., 2009), long-term trends due to the natural lake evolution (Blass et al., 2007b), and the appropriate choice of the statistical calibration and reconstruction method. This latter problem is particularly often overlooked, but it affects substantially the characteristics and amplitudes of the paleoclimate reconstructions (e.g. Esper et al., 2005; Bürger et al., 2006; Riedwyl et al., 2008; Kamenik et al., 2009).

In this study we present a 800-year long annually resolved record of biogenic silica (bSi) from proglacial Lake Silvaplana, eastern Swiss Alps. The sediment of Lake Silvaplana is annually laminated (varved; Leemann and Niessen, 1994; Ohlendorf et al., 1997; Blass et al., 2007a) and provides an excellent age control. Blass et al. (2007b) established that, in this lake, biogenic silica (bSi) flux is a proxy for summer (JA) and June–November temperatures, which corresponds to the ice-free period of the lake (Livingstone, 1997).

In this study we place particular emphasis on testing systematically the robustness of trends and amplitudes of reconstructed climate variability in different frequency domains. Correct amplitudes of variability are fundamental to assess, for instance, the (regional) climate sensitivity to a particular forcing or whether or not specific periods were warmer/colder than the last few decades (Bradley et al., 2003). Since the amplitude of a climate reconstruction is depending on the choice of the statistical reconstruction method (e.g. Esper et al., 2005), we calibrate our data set with six different regression methods (inverse regression, inverse prediction, generalised least squares, Major Axis regression, Ranged Major Axis regression and Standard Major Axis regression). The goal is to assess the influence that the choice of the method has on the amplitudes of the calibrated time series and calibration statistics (root-mean-squared error of prediction RMSEP, reduction of error RE and coefficient of efficiency CE) and, finally, the reconstruction. The regression methods are briefly reviewed and the calibration statistics are discussed in the light of our goal to assess the ‘real’ amplitude of decadal- to centennial-scale climate variability over the past 800 years.

Potential non-climatic trends in paleolimnological data sets and their influence on climate reconstructions are generally very difficult to assess and to quantify. Blass et al. (2007b) found a low-frequency (> 100 year domain), non-climatic trend in the biogenic silica flux data of Lake Silvaplana. This non-climatic trend has been removed and the bSi-based decadal-scale temperature variability is compared with two fully independent summer temperature reconstructions: the multi-proxy climate field reconstruction and the tree-ring late-wood density reconstructions (Casy et al., 2005; Büntgen et al., 2006).

Establishing the low-frequency (centennial-scale) domain of climate variability is arguably one of the most difficult but important challenges (Moberg et al., 2005; Jansen et al., 2007). In the following we combine the high-resolution bSi-based decadal-scale reconstruction with a lower-resolution bSi-based summer temperature reconstruction from the same lake and sediment core (Laroque-Tober et al., 2010). The chironomid-based temperature reconstruction has been established using a transfer function (Heiri et al., 2003), which is much more robust against long-term trends. As a final product we present a combined summer temperature reconstruction for the Alps for the past 800 years that has assessed skills in the sub-decadal- to multi-centennial range of climate variability.

2. Regional setting

Lake Silvaplana (Lej da Silvaplana, Fig. 1) is located in the Engadine, south-eastern Swiss Alps, at an altitude of about 1800 m a.s.l. The lake has a maximum depth of 77 m and a volume of 127 × 106 m3 (Blass et al., 2007a and references therein), and is usually ice-covered between January and May. Lake Silvaplana is a dimictic lake with a rather short period of spring mixing after the ice break-up that is followed by a period of summer stratification lasting from June to November. Strong local valley winds develop on sunny days between around 11 h and late afternoon if the synoptic-scale upper airflow is weak. This results in a generally well-mixed epilimnion during the summer stagnation (Blass et al., 2007b). The second overturn of the year occurs usually in December. The mean water residence time is < 1 year because of the high rate of inflow of glacial meltwater during summer. Lake Silvaplana is oligotrophic with < 10 μg l–1 orthophosphate in the water column (Bigler et al., 2007). Oxygen concentrations are consistently > 5 mg l–1 in near-bottom waters and the pH is around 7.8 (Bigler et al., 2007).

The catchment area of the lake extends over 129 km2. In 1998, about 6 km2 (5%) of the catchment area were glaciated (Blass et al., 2007a). The most important inflow is the Fedaela River, which has a mean discharge rate of 1.5 m3 s–1, is fed mainly by glacial meltwater and carries a high load of suspended sediment. A second inflow is the Inn River, which connects Lake Silvaplana to Lake Sils. This river has a mean discharge rate of 2 m3 s–1 but carries almost no suspended sediment. The discharge rates of the Vallun and Surlej Rivers are 0.7 and 0.3 m3 s–1, respectively.

The Engadine is an inner-alpine dry valley with a mean annual precipitation rate of 978 mm (SMA, 2002; climatology 1961–1990). The annual precipitation maximum occurs in August (121 mm) and the minimum in February (42 mm). Thunderstorms are relatively infrequent (20 days per year, SMA, 2002). The area lies in a meteorological boundary zone and receives precipitation mainly from the south (Brunetti et al., 2006). Monthly mean temperatures range from –7.8 °C in January to 10.8 °C in July.

3. Materials and methods

3.1. Material and chronology, climate data

Two piston cores of 9 m length were recovered in winter 2005/2006 to extend two previously recovered freeze cores (Blass et al., 2007b). Using marker layers, the piston cores were stratigraphically correlated to the freeze cores. The sediment cores were frozen with liquid nitrogen and individual varves were sampled in a freeze laboratory (–12 °C). Biogenic silica (bSi) concentration in the sediment was determined using alkaline leaching (Mortlock and Froelich, 1989) and ICP-OES, and corrected for lithogenic Si according to Ohlendorf and Sturm (2007). The annual Mass Accumulation Rate (MAR) was calculated from varve thickness, water and organic matter content according to Berner (1971) and Niessen et al. (1992). For further details see Blass et al. (2007b).

The existing chronology from AD 1580 to AD 1949 (Blass et al., 2007a; varve counting and documented flood layers) was refined with additional historically documented floods from the Upper Engadine (Caviezel, 2007). The chronology was then extended back to AD 1177 using varve counting of thin sections and digital images, and two historically documented very large flood events AD 1566
Fig. 1. Overview map of the Lake Silvaplana catchment area, including the coring position, meteorological station and the glacier extent during the Little Ice Age and in 1991 (redrawn from Blass et al., 2007a). The spatial correlation of Engadine summer temperatures is shown in the inset map on the right (Trachsel et al., 2008). The significant ($p < 0.1$, red, $r > 0.6$) correlations of 9-year smoothed bSi flux data with 9-year smoothed gridded summer temperatures (CRU TS3) are shown on the left. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
and AD 1177 (Sturm and Matter, 1978; Blass et al., 2007b; Fig. 2), which confirmed the varve counting.

For the calibration of the lake sediment proxies, we used monthly mean values of the homogenised meteorological data from nearby station Sils Maria (1864–1949, Begert et al., 2005) and early instrumental data from Böhm et al., (2010) average of 9 grid points centred around 46°N 9°E. The period AD 1950–2005 was not used for calibration because of varve counting difficulties and eutrophication effects (Blass et al., 2007b).

In order to account for dating uncertainties (varve-counting errors, e.g. Besonen et al., 2008) we tested the calibration also with 5-year and 9-year running means applied to the annually resolved bSi flux and meteorological data series prior to the correlation analysis. For the same reason (chronological uncertainties) the reconstruction is displayed as the 9-year (15-year) running mean back to AD 1500 (AD 1177).

3.2. Statistical methods

We applied six different regression methods to assess the effect of the method choice on the amplitudes of reconstructed temperatures. The methods were: ordinary least squares (OLS) regression with three variants (i) inverse regression (Venables and Ripley, 2002), (ii) inverse prediction (Sokal and Rohlf, 2001), (iii) generalised least squares regression (Venables and Ripley, 2002), (iv) Major Axis regression (Legendre and Legendre, 1998), (v) Ranged Major Axis regression (Legendre and Legendre, 1998), and (vi) Standard Major Axis regression (Legendre and Legendre, 1998).

OLS regression (i–iii) is referred to as Type I regression. OLS regression is minimising the squared errors on the x or the y axis and is assuming and error-free predictor (independent variable), whereas in Type II regressions (iv–vi; variants of Major Axis regression) the squares of the Euclidian distances are minimised (Legendre and Legendre, 1998). Type II regression should be applied if there is a random error on both variables x and y.

In theory, these six regression methods may only be applied to normally distributed data. We therefore carried out a set of Shapiro-Wilk tests (Shapiro and Wilk, 1965) to test our data for normality.

For the calibrations in the long period (AD 1760–1949; early instrumental data from Böhm et al., 2010), average of 9 grid points centred around 46°N 9°E) the reduction of error (RE) and the coefficient of efficiency (CE) statistics following Cook et al. (1994). When calibrating the 9-year smoothed bSi record for the shorter period with instrumental data (AD, 1864–1949, Blass et al., 2007b; Begert et al., 2005) we did not divide the data set into a calibration and a verification period (cross-validation) due to the low number of independent observations (i.e. degrees of freedom; Trenberth, 1984). In this case, we calculated a RE statistics comparing our temperature reconstruction with the mean observed temperature of the calendar period.

We used Pearson correlation coefficients and the significance levels were calculated after correction for the degrees of freedom DF according to Dawdy and Matalas (1964); the DF-corrected p-value is indicated as \( p_{corr} \).

Statistical analysis was performed using the open-source software R (r-project.org).

3.3. Detrending methods

Blass et al. (2007b) detected a non-climate trend in the bSi flux data and band-pass filtered the data series. Following this method, we used the residuals of a 100-year loess filter Cleveland and Devlin (1988). The span was set to represent 100 years in our annually resolved data set. Subsequently we applied a 9-year running mean (low-pass; calibration period and reconstruction from AD 1500 onwards) or 15-year running mean (reconstruction from AD 1177 onwards) to the residuals to account for the varve counting uncertainties (Blass et al., 2007a). The independent data sets used for comparison and validation (Casty et al., 2005; Büntgen et al., 2006) were treated in the same way.

To represent the full decadal- to centennial-scale spectra of climate variability in our record, we combined the band-pass filtered bSi summer (JJA) temperature data with the low-frequency component of the chironomid-inferred temperature series. Since chironomids are representing mostly July temperatures (Larocque-Tobler et al., 2010) we had to convert the chironomid-inferred July temperature record into summer (JJA) temperature to make it comparable with the bSi record. We used the mean and variance values of instrumental July and summer (JJA) temperatures (1864–2002) to scale the chironomid-inferred July temperatures to JJA temperatures. The scaled chironomid record was then low-pass filtered (100-year loess) to retain the centennial-scale variability.

4. Results

4.1. Chronology

Fig. 2 shows the refined age–depth model AD 1177–1949 with the turbidite layers of the documented floods in 1987, 1951, 1834, 1828 (Blass et al., 2007b) and the new additional floods reported for AD 1868, 1793, 1772, 1706, 1566 and 1177 (corrected ages according to Caviezel, 2007). The 1793 flood coincides with the massive sediment sequence from 90 to 95 cm depth where Blass et al. (2007b) found varve counting to be inconclusive. Here we interpret this sediment sequence as one event layer. This is supported by anomalous grain size medians and mineralogical composition (Blass et al., 2007a; Trachsel et al., 2008). In summary the freeze core reaches back to AD 1607 (previously AD 1577); the entire record covers the time since AD 1177. The flood of AD 1177 caused a huge turbidite deposit of 30 cm thickness. Therefore, we decided to end our record before this hiatus.

4.2. Biogenic silica as a temperature proxy: calibration, detrending and reconstruction

Blass et al. (2007b) demonstrated the high correlation of the bSi flux record with instrumental summer and autumn temperature data for the calibration period between 1864 and 1949. As
a consequence of the refined chronology during the calibration period we needed to recalibrate the bSi flux data.

With the new annually resolved data five significant ($p < 0.05$) correlations with meteorological data (Table 1a, Fig. 3) were obtained. The highest correlation was found for T JJA Jason ($r = 0.36$, $p < 0.01$). The 5-year running mean bSi flux was also highly correlated with T JJA Jason ($r = 0.6$, $p_{corr} = 0.012$); the correlation of the 9-year smoothed series with T JJA Jason amounted to $r = 0.69$ ($p_{corr} = 0.043$). These correlations were very similar to the values presented by Blass et al. (2007b) confirming the summer and autumn temperature signal preserved in the bSi flux data from the sediment of Lake Silvaplana. Generally, the accordance between the bSi flux data and the instrumental temperature data is excellent after AD 1878, whereas considerable mismatches in the calibration period are found between AD 1868 and 1878 (Fig. 3).

Since we are calibrating univariate data, the choice of the calibration method does not affect the shape of the reconstruction but the amplitude and, thereby, the calibration statistics (Fig. 4). Table 2 shows the reduction of error RE statistics, the root-mean-square error of prediction RMSEP, the amplitude predicted by the calibration period (1864–1949), Inverse Regression IR performs best (RMSEP is lowest and RE is highest) followed by Standard Major Axis SMA regression. RMSEP of General Least Square GLS, Major Axis MA and Ranged Major Axis RMA regression are very close, but RE performs poorly: RE values of MA and RMA regression are close to zero. The RMSEP of inverse prediction IP is highest and RE is negative.

With regard to the amplitude (Fig. 4), the RMA and MA regression are slightly overestimating the predicted amplitude compared with the observed amplitude of the instrumental data (true values, red line in Fig. 4) whereas SMA regression is slightly underestimating the amplitude. Large differences are found with IP regression (largely overestimating the amplitude), and inverse regression and GLS regression (largely underestimating the amplitude). The ratio between the amplitude and the RMSEP is highest for IP closely followed by RMA, MA and SMA regression. This ratio is considerably lower for IR and the even lower than 1 for the GLS regression, which means that the amplitude is smaller than the RMSEP.

Although in theory, the length and choice of the calibration period should be a random choice, in practice it is determined by the availability of instrumental data. The choice of the calibration period greatly influences the calibration statistics (Table 2). For instance, the reduction of the calibration period to 1874–1949 (instead of 1864–1949) results in large changes of the calibration statistics (data not shown).

According to Blass et al. (2007b) the long-term, non-climate trend in the bSi flux data series needs to be removed. We followed a purely statistical approach and applied a 9–100-year band-pass filter on the raw annual bSi flux data to (i) account for varve-counting errors (9-year low-pass) and (ii) to remove the long-term trend (100-year high-pass; Blass et al., 2007b). In consequence, the target and significance of the bSi-based temperature reconstruction were reduced exclusively to the decadal- and multi-decadal frequency domain of variability. The detrended temperature data are normally distributed for the period 1852–1949, but are differing from a normal distribution 1760–1949 and 1760–1851.

The correlation of the band-pass filtered bSi data with the band-pass filtered local meteorological series of Sils (period 1864–1949) amounts to $r = 0.66$ ($p_{corr} = 0.014$) and $r = 0.67$ ($p_{corr} = 0.023$) for JJASON and JJA temperature, respectively. Correlations with the Böhm et al. (2010) data set (period 1760–1950, mostly early instrumental data) amount to $r = 0.53$ ($p_{corr} = 0.005$) for summer and summer–autumn temperatures.

This longer calibration period contains enough independent observations (Trenberth, 1984) to make a split-period calibration as described by Cook et al. (1994). The results for RMSEP, RE, CE and amplitude statistics for the six statistical methods are shown in Table 2. GLS and IR regressions result in positive RE and CE statistics in all of the three calibration periods; the ratio between the amplitude and RMSEP, however, is always low. With regard to the amplitude, SMA regression matches the closest with the observed amplitude in the three calibration periods (always slightly underestimating the amplitude). Calibrating from 1760 to 1851 all regressions except IP and MA are (largely) underestimating the climate amplitude of the verification period 1852–1949. RMA and SMA are overestimating the amplitude of the verification period 1760–1851 when calibrating 1852–1949 where the amplitude is underestimated.

Fig. 6a and b shows the comparison of the bSi flux-based temperature reconstruction with two fully independent climate reconstructions: the climate field reconstruction for the Alpine Region back to AD 1500 (Casty et al., 2005) and a tree-ring based (late-wood density) temperature reconstruction for the Swiss Alps (Büntgen et al., 2006). The three data sets show a very good agreement both in the structure and the amplitude. Despite a considerable mismatch between AD 1530 and 1590, the correlation between the bSi record and the climate field reconstruction is highly significant and amounts to $r = 0.45$, ($p_{corr} = 0.0005$, period AD 1500–1590; $r = 0.5$, $p_{corr} = 0.0009$ for the period AD 1560–1950). Comparing the bSi record with the late-wood density record from AD 1177 to 1530, we find a correlation of $r = 0.54$ ($p_{corr} = 0.01$). The 16th century, however, seems enigmatic: between AD 1500 and 1600 the agreement between tree-ring data, bSi data and the multi-proxy reconstruction is strikingly poor. After

Table 1

<table>
<thead>
<tr>
<th></th>
<th>DIF</th>
<th>MAM</th>
<th>JJA</th>
<th>JAS</th>
<th>JJA Jason</th>
<th>SON</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r$</td>
<td>$r$</td>
<td>$r$</td>
<td>$r$</td>
<td>$r$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bSi flux</td>
<td>0.12</td>
<td>0.05</td>
<td>0.22</td>
<td>0.32</td>
<td>0.36</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>bSi flux 5-year running mean</td>
<td>0.02</td>
<td>0.06</td>
<td>0.56</td>
<td>0.55</td>
<td>0.6</td>
<td>0.51</td>
<td>0.45</td>
</tr>
<tr>
<td>bSi flux 9-year running mean</td>
<td>-0.32</td>
<td>0.08</td>
<td>0.67</td>
<td>0.67</td>
<td>0.69</td>
<td>0.6</td>
<td>0.47</td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Season</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500–1590</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1560–1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1600–1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1760–1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1850–1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JJA (Casty et al., 2005)</td>
<td>0.5</td>
<td>0.56</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>JJA Jason (Casty et al., 2005)</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
<td>0.58</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
AD 1600, the accordance is again very good: between AD 1600 and 1950 the correlation between bSi and the tree-ring record amounts to \( r = 0.42 \) (\( p_{\text{corr}} = 0.03 \)).

4.3. The combined temperature reconstruction

To integrate the low-frequency component of climate variability into our 9–100-year band-pass filtered bSi record, we added the 100-year low-pass component of a chironomid-inferred summer temperature reconstruction from Lake Silvaplana (Fig. 5d, Larocque-Tobler et al., 2010; bSi and chironomid data are from the same core). The RMSEP of the combined record amounts to 0.57 °C. The combined record (Fig. 6c; 9-year smoothed) shows pronounced (multi)decadal-scale variability superposed with long-term trends: warm temperatures around AD 1200 are followed by a cooling with minima around AD 1260 and 1360. Afterwards, cold summers followed with a next minimum around AD 1460. The 16th century appears slightly warmer, particularly at the beginning. The 17th century exhibits two severe minima around AD 1630 and 1690; these were interrupted by a short warm interval around AD 1650. The 19th and 20th century show a long-term warming with pronounced warm decades around AD 1730, AD 1790 and AD 1830. The latter two warm decades were interrupted by a severe cold interval centred on AD 1815–1820. At decadal scale, the highest values are observed in the late 12th century and the lowest values in the late 14th and the late 17th century. The amplitude of the combined record over the past ca 800 years (AD 1180–1949) is in the order of 2.8 °C. The values reconstructed for the warmest pre-industrial decade of the past 800 years (i.e. in the late 12th century) are 0.6 °C warmer than the values for the 1940.

5. Discussion

5.1. Calibration (model choice)

The six tested regression methods entail large differences in amplitude, RMSEP and RE statistics during the calibration period. Each of the methods has its strengths and weaknesses and is, given the distribution and errors of the predictor and predictand, theoretically more or less adequate (Legendre and Legendre, 1998). In theory, total least squares (TLS) regression would be the most appropriate method since it accounts for errors in \( x \) and \( y \) (Venables and Ripley, 2002). The uncertainties of the instrumental data are, however, not known.

Inverse regression IR, the most commonly used regression, minimises the error of the predictand with the least squares method assuming an error-free predictor (the proxy). In our case study, this assumption is obviously not valid and, therefore, IR is not a method of first-choice: according to Legendre and Legendre (1998) IR should not be applied when the uncertainty of \( x \) (here: the proxy) is much larger than the uncertainty of \( y \) (here: the temperature data). IR is also known to underestimate the slope of the regression (parameter \( b \); Legendre and Legendre, 1998; Esper et al., 2005), which underestimates the amplitude of the target (here: temperature variability in the calibration and reconstruction). This is consistent with our results (Fig. 4). Inverse regression results in good RE and CE statistics. This is expected because the

![Fig. 3. Comparison of annual biogenic silica flux (grey) with instrumental temperatures (Sils Maria, black) for the summer season (JJA; left row; Fig. 3a,d,g) and summer/autumn season (JJASON; Fig. 3c,e,i). The top panels show annually resolved data (Fig. 3a–c), the middle panels show 5-year smoothed data (Fig. 3d–f) and the bottom panels show 9-year smoothed data (Fig. 3g–i). Correlation coefficients are displayed in Table 1.](image-url)
predicted values are dragged towards the mean value of the target (Venables and Ripley, 2002) to which they are compared calculating the RE and CE statistics (Cook et al., 1994).

Inverse Prediction IP would be the correct method since the proxy (y) is dependent on temperature (x). IP enhances the slope of the regression (parameter b), and thus the amplitude (Sokal and Rohlf, 2001). This is in line with our results showing that IP does not perform well in reproducing climate variability during the calibration period. In light of the homogenisation problems with the meteorological data used (Begert et al., 2005) it remains also debatable to which extent the assumption is valid that temperature measurements (x) are error-free. Because of the large predicted amplitude, IP yields poor RE and CE results.

The Generalised Least Squares regression GLS is a special inverse regression type (Venables and Ripley, 2002) and is, thereby also known to underestimate the slope of a regression and thus to underestimate the amplitude (Legendre and Legendre, 1998). GLS regression produces a summer temperature reconstruction that exhibits almost no variability, which is not realistic. MA, Ranged Major Axis RMA and Standard Major Axis SMA regressions (Type II regressions) should be used if we can assume an identical and random error of both the predictor and the predictand (here: bSi and temperature) (Legendre and Legendre, 1998). MA regression should be used when both variables are expressed in the same units or are dimensionless, or when the variance of the random errors of both parameters is equal. If the variance of the random errors of both parameters is not equal SMA regression (scaling, geometric mean regression) should be used if the ratio of the variance of the error and the variance of the random errors of both parameters is not equal. SMA regression produces a summer temperature reconstruction that exhibits almost no variability, which is not realistic.
assessments of the amplitude of the reconstruction (slope of the regression model). The method with the best performance in RE/CE statistics (IR) shows poor performance in modelling the amplitude. Thus the question arises: how can the quality of a calibration be evaluated? There are a number of different measures of "quality": the RE and CE calibration statistics (Cook et al., 1994), smallest RMSEP (e.g. Kamenik et al., 2009), the ratio between the amplitude and the RMSEP (which indicates the proportion of uncertainty on the amplitude, e.g. Birks, 1998), or the 'correct' representation of the amplitude of the target variable (Legendre and Legendre, 1998).

Assessing the 'correct' amplitude of climate change in the various frequency domains is arguably one of the prime challenges to detect forced and unforced variability. This is especially the case for the past millennium, when forcing factors are relatively well documented (Jansen et al., 2007) and regional responses to perturbations may be investigated quantitatively (detection and attribution). In consequence, we place utmost care on the 'correct' amplitude of climate variability and favour for our summer temperature reconstruction (Fig. 6) Type II regressions, whereby SMA performs best for the detrended data in the calibration period 1760–1949. Placing too much emphasis on good RE and CE statistics while neglecting the performance of the amplitude might result in problematic climate reconstructions.

5.2. Comparing the biogenic silica flux record with independent temperature reconstructions

The band-pass filtered bSi flux series, the multi-proxy reconstruction for the Alpine area of Casty et al. (2005; AD 1500 onwards) and the late-wood density MXD data of the Alpine summer temperature reconstruction (Büntgen et al., 2006) (Fig. 6a and b) show overall an excellent agreement among each other both in the structure and the amplitudes. All of the three data sets are fully independent. Agreement is expected since the distribution of summer temperatures in Central Europe and the Alps is mostly uniform (Fig. 1 inset map). The residuals, however, might be attributed to differential ecosystem responses to a specific combination of forcings or boundary conditions (e.g. drought, change in nutrient availability between 1868 and 1878 in Lake Silvaplana, Bigler et al., 2007), or to errors in the data sets. In the following, the three different time series are compared, and put into perspective of internal and forced variability as found in ensemble model simulations for European summers of the past 1000 years (Goosse et al., 2005, and references therein; forcings after Jansen et al., 2007).

The characteristic features of decadal-scale climate variability between AD 1500 and 1949 are consistently found in the three archives: the cold anomaly at the end of the 16th century (e.g. Dobrovolny et al., 2010) coincided with marked negative volcanic forcing: the very cold summers in the Late Maunder Minimum centred on AD 1690 (Dobrovolny et al., 2010) coincided with low solar irradiance and strong negative volcanic forcing; the same combination is found for the very pronounced cold spell during the Dalton Minimum and the negative volcanic forcing around 1815–1820 (Tambora eruption). Negative temperature anomalies are found during the Damon solar minimum around 1910; the relatively warm phases in the middle of the 18th century (Dobrovolny et al., 2010) are also well captured by all the three records.

Interesting is the disagreement among the three records in the 16th century. The cold anomaly centred on AD 1505 and the warm anomaly around AD 1525 are found in the tree ring and lake sediment data while the multi-proxy data (Casty et al., 2005) show an opposite structure. The cold decade around AD 1545 is found in the tree ring and multi-proxy series, while the sediments suggest a warming. All the three archives differ vastly until AD 1595 when...
they line up again during a cold anomaly with strong negative volcanic forcing. The 16th century was a period with broad absence of negative volcanic forcing and relatively low solar irradiance. We may speculate that the three archives responded with different sensitivities to this particular combination of forcings. The different behaviour of the tree-ring record between 1635 and 1665 is also enigmatic, where both the sediment record and the multi-proxy record show consistent decadal trends and negative temperature anomalies coinciding with negative volcanic forcing peaking around AD 1630 and 1660. The multi-proxy and sediment records seem to respond more sensitively to volcanic forcing, at least during that period. The same accordance between the multi-proxy and the sediment records is found during the cold anomaly around 1775 but not recorded in the tree-ring MXD data. The minor discrepancies between 1780 and 1810 were attributed to an anomalously high NAO index (Blass et al., 2007b) which influences temperatures during summer in the Southern Alps (location of Lake Silvaplana).

The sediment record and the tree-ring record compare well between AD 1177 and 1500. The temporal offset between both records (sediments lead the tree rings by 4–6 years) is likely attributable to varve-counting errors since there is no stratigraphic flood marker in that section of the core that could help to constrain the sediment chronology in more detail (Fig. 2). On purpose we did not match the wiggles and keep both records fully independent. The recorded cold anomaly centred on AD 1453; is very well represented: the Kuwae eruption produced the third largest negative volcanic forcing of the past 1000 years (Jansen et al., 2007), which is accurately captured in the ensemble model simulations (Goosse et al., 2005). A very pronounced cooling is recorded after the volcanic eruption of AD 1258, one of the two strongest negative forcings of the past 1000 years. A further negative temperature anomaly around 1220 is also attributable to volcanic forcing. It seems that all of the volcanic eruptions with major negative forcing are represented as pronounced short-lived pulses with negative summer temperature anomalies in the decadal- to multi-decadal-scale reconstructions suggesting that volcanic forcing is a major source of forced variability during summer in the (multi)decadal frequency band. This is consistent with model experiments (Goosse et al., 2005) and observations of composite temperature responses to volcanic forcing in Europe (Fischer et al., 2007). Of particular interest is the massive and prolonged cooling recorded in the sediments around AD 1360 (Fig. 6b and c). This cooling is hardly visible in the tree ring series. Solar irradiance is around average and volcanic forcing is unimportant, and the model simulations show only an insignificant cooling (Goosse et al., 2005). However, this cooling is real, massive and regionally very well documented in the enormous and very rapid advances of the Great Aletsch and the Gorner Glacier to maximum Little Ice Age stages around AD 1370–1390 (Holzhauser et al., 2005; the response lag of 2–3 decades to the initial cooling is characteristic for the glaciers).

In the absence of sufficiently strong forcings, we consider a reduced Atlantic Meridional Overturning Circulation during that time (Bianchi and McCave, 1999) as a possible explanation. Reduced AMOC leads to generally cool conditions in Europe, which is consistent with our observations but needs further investigation. We may speculate whether this phenomenon might be an
expression of internal variability and serve as a case for stochastic resonance (Alley et al., 2001).

The combined summer temperature record of Lake Silvaplana (9–100-year band-pass biogenic Si and 100-year low-pass chironomid; Fig. 6c) shows similar characteristics in the multi-decadal and centennial variability domain compared with the tree-ring late-wood density inferred summer temperature series (Büntgen et al., 2006). The warmest decades superposed on the long-term centennial temperature trends, are found in both records around AD 1200 and in the late 20th century, whereas the coldest decades are observed around AD 1690, in the 17th century in general. The first decade of the 21st century are 0.13 °C warmer than the warmest decade of the pre-industrial time for the past 800 years, which is not significant given the reconstruction errors. The overall amplitude between the warmest and coldest decades during the past 800 years amounts to Δ2.8 °C in both the Alpine tree ring (Büntgen et al., 2006) and the sediment records (this study). This amplitude at the regional scale is about at least two times larger than other hemispheric or global reconstructions (Jansen et al., 2007): for instance the 9-year smoothed Northern Hemispheric reconstruction by Moberg et al. (2005) suggests an amplitude of 1 °C between 1177 and 1950 compared to 2.8 °C in this study. Our results confirm the finding by Hunt (2004) suggesting that the amplitude of climate variability is much enhanced at the regional scale compared with hemispheric or global scales. This has, indeed, significant implications when impacts of climate change are concerned: natural and managed ecosystems experience enhanced variability of the local or regional climate.

6. Conclusions

In this study we present multi-decadal to sub-centennial biogenic silica (bSi)-inferred summer (JJA) and summer–autumn (JJASON) temperature reconstructions back to AD 1177. We combined the decadal-scale bSi record with the low-frequency component (100-year low-pass) of chironomid-inferred summer temperatures to produce a record that has defined skill in a broad spectrum of climate variability.

One of the most critical issues in past climate research and climate reconstructions is the correct assessment of amplitudes of

---

Fig. 6. (a) 100-year high-pass filtered and 9-year smoothed (9–100-year band-pass filtered) bSi flux-based temperature reconstruction AD 1500–1949 (blue, light blue indicating the RMSEP) and multi-proxy summer temperature reconstruction (red; Casty et al., 2005, data filtered in the same way). (b) 15–100-year band-pass filtered bSi temperature reconstruction AD 1177–1949 (blue), Casty et al. (2005) JJA temperature reconstruction (red) and tree-ring based Alpine summer temperature reconstruction (Büntgen et al., 2006) (black). (c) Combined chironomid and bSi flux-based summer temperature reconstruction (9-year running mean, red, light red indicating the uncertainty) and 9-year running mean of summer (JJA) temperature from adjacent station Sils Maria (black). All anomalies are given with reference to the period 1961–1990. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
climate variability. Here we tested systematically six different regression methods (Type I regressions: Inverse Prediction, Inverse Regression and General Least Squares; Type II regressions: Major Axis, Ranged Axis and Standard Major Axis). We used a number of indicators to assess the quality of the calibration: RE and CE statistics, the ratio amplitude/RMSEP and the correct representation of the modelled amplitude compared with observed data. The six methods revealed very different results whereby those methods that show good skills in the RE and CE statistics (IR) perform very poorly with regard to the amplitude, and vice versa (Type II regressions). There seems to be a trade-off, and the optimum model choice depends on the objective of the research. Obviously the correct amplitude is the prime target for climate reconstructions; in our case, i.e. for band-pass filtered data, a regression model based on Standard Major Axis regression performed best. We used this regression model for the climate reconstruction and conclude that a focus on RE and CE statistics alone might by myopic and eventually lead to problematic reconstructions.

Our sediment-based summer temperature reconstruction reveals very good agreement with two independent summer temperature reconstructions (multi-proxy reconstruction from Casty et al., 2005 and tree-ring inferred Alpine temperatures from Büntgen et al., 2006) both in the structure and the amplitudes. We found that all the major episodes of negative volcanic forcing are represented in negative summer temperature anomalies. This is particularly pronounced in periods when negative volcanic forcing coincides with low solar irradiance as e.g. during the Maunder Minimum around AD 1690, the Spörer Minimum around AD 1450, and possibly at the very end of the Wolf Minimum around AD 1360. Volcanic forcing seems to be a major source of summer temperature variability in the decadal-scale frequency band. The warmest and coldest decades of the past 800 years coincide in the sediment record and the tree-ring record, showing that the decadal-scale variability superposed on the centennial-scale variability is robust. The current decades are only slightly warmer (ca 0.13 °C) than the warmest pre-industrial decade since AD 1180. Warm season temperature amplitudes between the warmest and the coldest decades of the past 800 years coincide in the sediment varved sediments. The Holocene 17, 51–63.

References


