GEOPHYSICAL RESEARCH LETTERS, VOL. 35, L20707, doi:10.1029/2008GL035395, 2008

An ensemble of European summer and winter temperature reconstructions back to 1500

N. Riedwyl,[1] J. Luterbacher,[1] and H. Wanner[1]

Received 22 July 2008; revised 8 September 2008; accepted 18 September 2008; published 25 October 2008.

[1] An ensemble of statistical methods is applied to reconstruct European temperature variability back to 1500. We apply principal component (PC) regression, regularized expectation maximization (RegEM) and composite-plus-scaling (CPS) to multi-proxy data. The reconstructions for summer and winter European temperature averages, and spatial fields related to warmest and coldest decades are analyzed and discussed. PC regression and RegEM perform more similarly compared to CPS, and more robust reconstructions are achieved for winter than for summer. We conclude that temperature reconstructions can not be improved significantly by replacing the reconstruction technique only. Discordances are very likely to be due to limited spatial and temporal availability of the proxy data. The comparison reveals that seasonal temperature variability is likely more variable than indicated earlier, still pointing out the exceptional warmth of the late 20th century. However, further evidence is needed, as the summer reconstruction results of the three techniques are not yet fully coherent. **Citation:** Riedwyl, N., J. Luterbacher, and H. Wanner (2008), An ensemble of European summer and winter temperature reconstructions back to 1500, *Geophys. Res. Lett.*, *35*, L20707, doi:10.1029/2008GL035395.

1. Introduction

[2] Spatio temporally high resolved reconstruction of past climate variability is of high importance in the discussion on current climate change [*Jansen et al.*, 2007; *Mann et al.*, 2008]. The question whether caveats of reconstruction techniques can lead to biased conclusions about past temperature variations, is crucial and needs to be addressed. Many studies focus on testing climate reconstruction techniques in "a surrogate climate" [*Mann and Rutherford*, 2002; *von Storch et al.*, 2004, 2008; *Rutherford et al.*, 2005; *Küttel et al.*, 2007; *Lee et al.*, 2008; *Mann et al.*, 2007; *Moberg et al.*, 2008; *Riedwyl et al.*, 2008] using pseudoproxies, i.e., proxies derived from climate model simulations. This "laboratory" is useful to get "a priori" knowledge about the performance of reconstruction techniques. However, these studies do not allow final conclusions concerning reconstruction techniques applied to real world instrumental and proxy data. Therefore, in this contribution we use real data as a complement to the lessons learnt from "a surrogate climate", and to better take into account the impact of real world conditions. In particular the much more finely resolved real gridded target surface air temperature field in contrast to the lower resolved hemispheric climate model fields is considered. Furthermore, the heterogeneity and limited spatial and temporal availability of the real proxy data, which is idealized often using pseudoproxies [*Riedwyl et al.*, 2008] is accounted for here.

[3] *Mann et al.* [2008] compare hemispheric and global reconstructions based on the composite-plus-scaling (CPS) approach and regularized expectation maximization (RegEM)-based estimation procedure. We compare multivariate principal component (PC) regression, the classical method used to reconstruct past European climate [e.g., *Luterbacher et al.*, 2004; *Xoplaki et al.*, 2005; *Casty et al.*, 2005; *Pauling et al.*, 2006], CPS [e.g., *Jones and Mann*, 2004; *Esper et al.*, 2005], and RegEM [*Schneider*, 2001; *Rutherford et al.*, 2005; *Mann et al.*, 2007]. Following on previous studies, we further explore RegEM with truncated total least squares (TTLS) as regularization scheme, and provide more evidence for the usefulness of the errors-in-variables approach TTLS at the seasonal European scale. We show that the analysis of an ensemble of reconstructions leads to a more in-depth understanding of the reliability and robustness of existing European temperature reconstructions back to 1500.

2. Data and Methods

[4] The predictand is the European surface air temperature field taken from *Mitchell and Jones* [2005] at 0.5°×0.5° resolution. Europe is represented by the area 24.5°W to 39.75°E and 35.25°N to 69.75°N (land area only) according to earlier studies [*Luterbacher et al.*, 2004; *Xoplaki et al.*, 2005]. The predictors (see *Luterbacher et al.* [2004] for an overview), i.e., the proxy data consist of a combination of early instrumental temperature records, documentary proxy evidence, ice core- and sea ice data, albeit with some additional information that has become later available (tree ring series and grape harvest dates). In order to use maximal predictor information for each year in the reconstruction, different proxy networks are used, i.e., 136-predictor networks in summer and 128 in winter. For PC regression and RegEM separate reconstructions are performed for each single network. For CPS the composite, i.e., the average normalized series, of the available proxy data is calculated first, and then used for reconstruction. PC regression is applied as outlined by *Luterbacher et al.* [2004] and *Riedwyl et al.* [2008]. Predictors and predictand are transformed to their principal components (PC) using truncated singular value decomposition (TSVD) with the truncation levels according to *Luterbacher et al.* [2004]. RegEM is a covariance-based iterative algorithm, replacing missing values with plausible values (imputation), as described in

[1]Oeschger Centre for Climate Change Research and Institute of Geography, Climatology and Meteorology, University of Bern, Bern, Switzerland.

0094-8276/08/2008GL035395

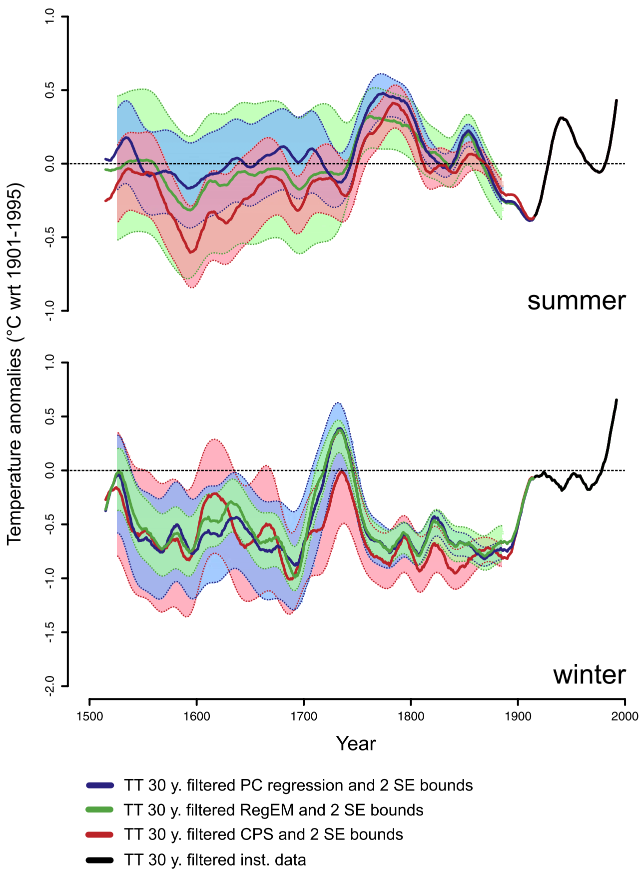

Figure 1. 30-year Gaussian filtered European summer and winter average temperature anomalies (wrt the 1901 to 1995 calibration period) back to 1500, PC regression (blue line, corresponding 2 standard errors (SE) blue shaded), RegEM (green line, corresponding 2 SE green shaded) and CPS (red line, corresponding 2 SE red shaded). Instrumental surface air temperature data (1901 to 2002 [*Mitchell and Jones*, 2005] and 2003 to 2007 [*Hansen et al.*, 2001]) in black.

detail by *Schneider* [2001], *Rutherford et al.* [2005], and *Mann et al.* [2007]. We apply the non-hybrid version of RegEM with TTLS as regularization scheme to take into account under-determined settings [*Mann et al.*, 2007]. Furthermore, the predictand is represented by its leading PC [*Mann et al.*, 2007]. The truncation levels for the TTLS regularization of the covariance and the number of predictand PC to be kept are based on the estimate of the noise continuum to the log-eigenvalue spectrum [*Mann et al.*, 2007; *Wilks*, 1995]. In few cases, the predictor networks are very small, and too many PCs tend to be retained by the above mentioned criteria. We therefore set the truncation levels in these cases as low as possible, i.e., to 1.

[5] With PC regression and RegEM the European summer and winter temperature averages are computed given the reconstructions of the underlying spatial field. Using CPS [e.g., *Esper et al.*, 2002, 2005; *Jones and Mann*, 2004; *Moberg et al.*, 2005] the temperature averages are reconstructed directly by centering and scaling the proxy data composite according to the calibration (1901 to 1995) average and standard deviation of the predictand.

[6] For the time series plots, we only focus on lower frequency variations, whereas the reconstruction skill scores are calculated for the un-filtered data. Thus, for the European averages only the 30-year Gaussian filtered reconstruction results with associated uncertainties (filtered 2 standard errors SE) are shown. For PC regression and CPS the SE refer to the prediction intervals, for RegEM they also relate to the imputation error of its iterative algorithm. Thus, for CPS the SE are solely based on the residuals, whereas for PC regression the unexplained variance of the regression is considered as well. Furthermore, the SE of RegEM represent the average imputation error of the imputed PC values. The SE for the 30-year filtered reconstruction results are calculated as in the work by *Xoplaki et al.* [2005].

[7] Furthermore, we compare the spatial temperature anomaly averages of the warmest ten-summer means and coldest ten-winter means as well as extreme single summer and winter years, using PC regression and RegEM.

[8] Verification is performed for the largest predictor networks available by the end of the 19th century, as well as for the available proxies used to reconstruct the warmest summer and the coldest winter decades. We calculate the reduction of error (RE) and the coefficient of efficiency (CE) [*Cook et al.*, 1994], as well as the relative root mean squared error (RRMSE) [*Lee et al.*, 2008], using the period 1901 to 1960 for calibration, and 1961 to 1995 for verification. The closer RE and CE values get to 1 and the RRMSE values to 0, the higher is the skill of the reconstruction results.

3. Results and Discussion

3.1. Analysis of the Temperature Average Reconstructions

[9] Figure 1 shows the comparison of PC regression (blue), RegEM (green) and CPS (red) for European summer (top) and winter (bottom) average temperature anomaly reconstructions back to 1500 and associated uncertainties. For summer the three reconstruction results agree on pronounced positive anomalies in the late 18th century. Furthermore, the CPS result exhibits maximal negative temperature anomalies at the end of the 16th century, which are less pronounced for RegEM and PC regression. Mainly before 1700, the variances of the reconstruction results deviate from each other: Temperature variability is about the same for PC regression and RegEM, but larger for CPS, which illustrates a fundamental difference between scaling, i.e., CPS, and the regression based techniques PC regression and RegEM. PC regression seems to "loose" variability over the reconstruction period [*Küttel et al.*, 2007; *Riedwyl et al.*, 2008], RegEM with TTLS is less affected, and CPS, by definition, fully retains the variability within the calibration period. The central estimate obtained with the CPS method sometimes lies outside the 2 SE bounds for PC regression, and vice versa. However, the PC regression and CPS central estimates lie within the filtered 2 SE bounds of the RegEM result. Furthermore, the RegEM central estimate is just outside the 2 SE bounds of PC regression and CPS at the end of the 18th century.

[10] For winter (Figure 1, bottom) the PC regression and RegEM results show distinctly positive temperature anomalies for the mid 18th century and negative anomalies at the

Table 1. RE, CE and RRMSE Skill Scores for the Results Shown in Figure 1[a]

	PC		RegEM		CPS	
	Summer	Winter	Summer	Winter	Summer	Winter
Reduction of Error (RE)						
Maximal set	0.9838	0.991	0.828	0.938	0.649	0.927
Selected subset	0.773	0.752	0.670	0.693	0.434	0.451
Coefficient of Efficiency (CE)						
Maximal set	0.9836	0.990	0.826	0.936	0.644	0.925
Selected subset	0.770	0.746	0.665	0.685	0.426	0.437
Relative Root Mean Squared Error (RRMSE)						
Maximal set	0.128	0.098	0.418	0.253	0.597	0.274
Selected subset	0.480	0.504	0.579	0.561	0.757	0.750

[a]The calibration period is 1901 to 1960, the verification period 1961 to 1995. The skill scores are calculated for reconstructions using the maximal proxy data set available by the end of the 19th century (173 proxies in summer, and 170 in winter), and the proxy data available to reconstruct the warmest summer and coldest winter decade only (selected subset, i.e., 30 proxies in summer and 11 in winter).

end of the 17th century. The results of PC regression and RegEM mostly agree on the magnitude of temperature amplitudes. Before 1700 the reconstruction results deviate less from each other than for summer. Also the 2 SE bounds indicate better accordance of the reconstruction results for winter than for summer. The result of RegEM fully lies within the 2 SE bounds of the PC regression result, and vice versa. The CPS central estimates sometimes lie outside the 2 SE bounds of the two other methods. The averaged summer and winter temperature anomalies are coldest for CPS compared to PC regression and RegEM.

[11] Table 1 presents RE, CE and RRMSE scores for the European average temperature time series. PC regression reveals highest skill both for summer and winter, followed by RegEM. CPS has lowest skill. Thus, the validation measures RE, CE and RRMSE penalize large differences between the reconstructed average and the calibration period average, and do not reward the perfect match of the variance of the CPS result with the variance of the calibration period. Furthermore, a probable decrease in variability over the verification period seems not to be penalized by the validation measures [*Riedwyl et al.*, 2008]. There is a decrease in skill (Table 1), if verification is performed for the reconstructions with the two subsets used to reconstruct warmest summer and coldest winter decade, compared to the maximal proxy data sets.

[12] Seasonal differences in the performance are evident between the three methods: For RegEM, the 2 SE bounds for the winter result indicate much smaller imputation SE than for summer. This is likely due to the fact that the predictor series are slightly more continuous for winter, and overall the number of the predictand PCs used for reconstruction is smaller than for summer. Thus, the amount of initial missing values of the input matrix is smaller for winter than for summer, what leads to smaller imputation errors. The uncertainty bounds of CPS differ from those of PC regression and RegEM, in the sense that they are largest in winter, as they directly refer to the European average series, and not to the PC of the underlying spatial field. Furthermore, for CPS and RegEM the skill is higher for winter than for summer (Table 1), which is not the case for PC regression (Table 1, selected subset summer compared to winter). Nevertheless, Figure 1 shows that there is better accordance of the three results for winter regarding periods with maximal positive and negative temperature anomalies. There are significant differences in variance (F-test) between PC regression and CPS, as well as RegEM and CPS for summer, while for winter the differences are not significant. Thus, there seems to be a better coherence of the temperature signals in the winter proxy data. Whereas in winter only early instrumental series, ice core data and documentary evidence are used, the types of proxy data, and the temperature signal inherent vary more for summer.

[13] It has been demonstrated in “a surrogate climate”, that RegEM can provide more skilful results than PC regression, presumably due to its explicit incorporation of errors inherent to proxy data which appears to lead to better capturing of low-frequency variability [*Riedwyl et al.*, 2008]. This advantage is less obvious here. The use of many early instrumental series as predictors leads to rather high signal to noise (SNR) ratios, i.e., a low rate of errors inherent to the predictor data [*Küttel et al.*, 2007]. Therefore, in the case here, PC regression and RegEM seem comparable. Nevertheless, the error assumptions of RegEM with TTLS are more realistic than those of PC regression, presuming noise inherent to the predictors as well.

[14] The strong positive temperature anomalies at the end of the 18th century of the summer results (Figure 1, top) are likely to be an artefact of too warm early instrumental measurements, as has been discussed by *Moberg et al.* [2003], *Frank et al.* [2007], and R. Böhm et al. (The early instrumental warm-bias: A solution for long central European temperature series, submitted to Climatic Change, 2008). The probable warm bias of early instrumental summer temperatures does not invalidate our comparison of the three methods. The pronounced negative winter temperature anomalies at the end of the 17th century (Figure 1, bottom) represent the well known cold of the Maunder Minimum [e.g., *Luterbacher et al.*, 2001, 2004]. We further focus on these two periods for the analysis of the reconstructed European temperature fields.

3.2. Analysis of the Temperature Field Reconstructions

[15] PC regression and RegEM agree on the coldest winter decade, 1689 to 1698. However, in summer the warmest decade using PC regression is 1789 to 1798, and in the case of RegEM it is 1774 to 1783. Figure 2 shows the comparison of reconstructed temperature anomaly fields averaged over the warmest summer decade (1789 to

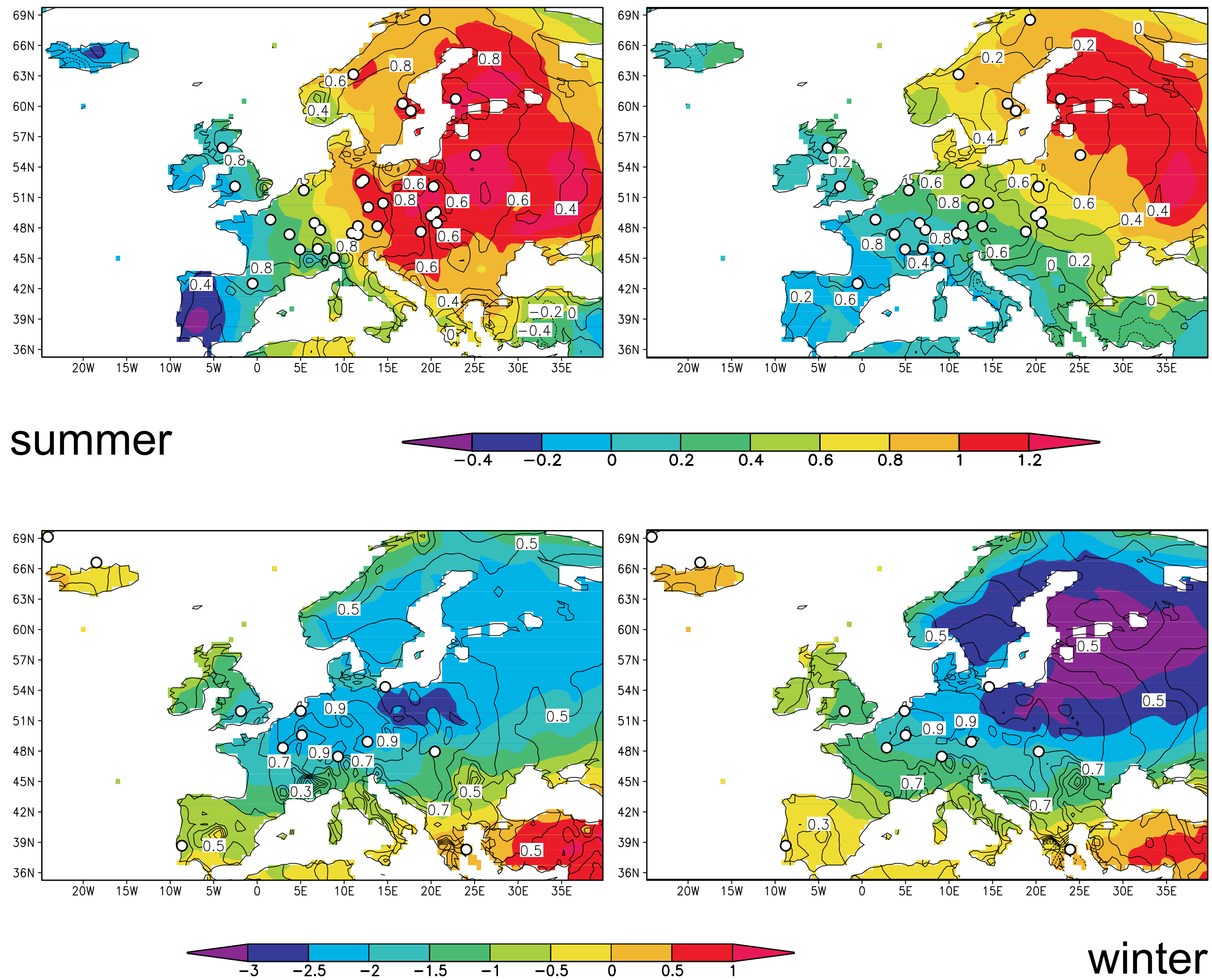

Figure 2. European surface air temperature anomaly (wrt to the 1901 to 1995 calibration period) fields, averaged over the (top) warmest summer decade (1789 to 1798) and (bottom) coldest winter decade (1689 to 1698) using (left) PC regression and (right) RegEM. Contours represent RE skill patterns. White dots indicate the locations of the proxies used for reconstruction.

1798) and coldest winter decade (1689 to 1698), using PC regression (left) and RegEM (right). For the warmest summer decade (Figure 2, top), the maximal temperature anomalies are more pronounced using PC regression than using RegEM. There are some similarities between the spatial patterns, e.g., for Central Europe, where most proxy data are available, and less accordance exists for the North East, where the proxy data coverage is sparse. For the coldest winter decade (Figure 2, bottom) the negative temperature anomalies in the North East are more pronounced using RegEM than using PC regression, and again the differences between the two anomaly patterns are most clear where the proxy data network is sparse. The reasons for the differences might be that with PC regression, the PC of predictand and predictors are taken, while with RegEM the PC of the predictand only are considered. Therefore, single summer predictor series may obtain more weight using PC regression, and dominate the periods in which they explain most variance, which is not the case for RegEM. The RE scores for PC regression (Figure 2, left) indicate slightly higher skill than those for RegEM (Figure 2, right) in particular over the data sparse regions. However, the performance of PC regression and RegEM to reconstruct warmest summer and coldest winter decades (Figure 2), as well as extreme years (auxiliary material) are comparable.[1] The spatial field of reconstructed single cold and warm winters and summers are provided in the auxiliary material. PC regression and RegEM agree more for winter (warmest year: 1724; coldest year: 1695) than for summer (warmest year: 1798 for PC regression, 1826 for RegEM; coldest year: 1821).

4. Conclusions

[16] We conclude that temperature reconstructions can not be improved significantly by only replacing the reconstruction technique. Discordances are very likely to be due

[1]Auxiliary materials are available in the HTML. doi:10.1029/2008GL035395.

to the spatial distribution and uncertainties inherent to the proxy data, as well as their limited availability. More robust results are found for winter than for summer. More evidence is still needed in order to get a coherent reconstruction for past European summer temperatures. An ensemble of results can help to improve the reliability and robustness of reconstructed past temperature variability amplitudes. Applying several techniques to reconstruct the same target can reduce the uncertainties, and is an approach worthwhile pursuing consequently in future.

[17] However, related to European summer and winter average reconstructions, we found that PC regression and RegEM perform more similar compared to CPS. This is likely due to the impact of scaling (CPS) in contrast to multivariate regression with regularization schemes (PC regression using TSVD, and RegEM using TTLS).

[18] Testing RegEM with TTLS and PC regression with TSVD shows that both techniques are suitable and promising for reconstructions at the European scale with real instrumental and proxy data. However, the determination of truncation levels, both for PC regression using TSVD, and for RegEM using TTLS is a field for further investigation and exploration. CPS, compared to PC regression and RegEM, offers the advantage to be much more easily applied.

[19] Highest skill for both summer and winter is achieved for PC regression, with RegEM having slightly less good skill scores and CPS having still lower skill. However, both RegEM and CPS reveal reconstruction results with lower, and in the case of CPS clearly more variable values than PC regression. The commonly calculated skill scores for verification seem not to fully capture the performance of the reconstruction techniques [*Riedwyl et al.*, 2008]. The fact that with CPS the variance of the calibration period is fully retained within the reconstruction period is not rewarded by the commonly calculated skill measures.

[20] **Acknowledgments.** This work has been supported by the Swiss National Science Foundation (SNSF) through its National Center of Competence in Research on Climate (NCCR Climate) project PALVAREX 2. We thank the reviewers for their constructive comments and suggestions, which helped to improve the quality of this study.

References

Casty, C., H. Wanner, J. Luterbacher, J. Esper, and R. Böhm (2005), Temperature and precipitation variability in the European Alps since 1500, *Int. J. Climatol.*, *25*, 1855–1880.

Cook, E., K. Briffa, and P. Jones (1994), Spatial regression methods in dendroclimatology: A review and comparison of two techniques, *Int. J. Climatol.*, *14*, 379–402.

Esper, J., E. Cook, and F. Schweingruber (2002), Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability, *Science*, *295*, 2250–2253.

Esper, J., D. C. Frank, R. J. S. Wilson, and K. R. Briffa (2005), Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, *Geophys. Res. Lett.*, *32*, L07711, doi:10.1029/2004GL021236.

Frank, D., U. Büntgen, R. Böhm, M. Maugeri, and J. Esper (2007), Warmer early instrumental measurements versus colder reconstructed temperatures: Shooting at a moving target, *Quat. Sci. Rev.*, *26*, 3298–3310.

Hansen, J., R. Ruedy, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl (2001), A closer look at United States and global surface temperature change, *J. Geophys. Res.*, *106*, 23,947–23,963.

Jansen, E., et al. (2007), Paleoclimate, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the International Panel on Climate Change*, edited by S. Solomon et al., pp. 433–497, Cambridge Univ. Press, Cambridge, U. K.

Jones, P. D., and M. E. Mann (2004), Climate over past millennia, *Rev. Geophys.*, *42*, RG2002, doi:10.1029/2003RG000143.

Küttel, M., J. Luterbacher, E. Zorita, E. Xoplaki, N. Riedwyl, and H. Wanner (2007), Testing a European winter surface temperature reconstruction in a surrogate climate, *Geophys. Res. Lett.*, *34*, L07710, doi:10.1029/2006GL027907.

Lee, T., F. Zwiers, and M. Tsao (2008), Evaluation of proxy-based millennial reconstruction methods, *Clim. Dyn.*, *31*, 263–281.

Luterbacher, J., R. Rickli, E. Xoplaki, C. Tinguely, C. Beck, C. Pfister, and H. Wanner (2001), The Late Maunder Minimum (1675–1715)—A key period for studying decadal scale climate change in Europe, *Clim. Change*, *49*, 441–462.

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

Mann, M. E., and S. Rutherford (2002), Climate reconstruction using 'pseudoproxies', *Geophys. Res. Lett.*, *29*(10), 1501, doi:10.1029/2001GL014554.

Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann (2007), Robustness of proxy-based climate field reconstruction methods, *J. Geophys. Res.*, *112*, D12109, doi:10.1029/2006JD008272.

Mann, M. E., Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, S. Rutherford, and F. Ni (2008), Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *Proc. Natl. Acad. Sci. U. S. A.*, *105*, 13,252–13,257.

Mitchell, T., and P. Jones (2005), An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *Int. J. Climatol.*, *25*, 693–712.

Moberg, A., H. Alexandersson, H. Bergstroem, and P. Jones (2003), Were southern Swedish summer temperatures before 1860 as warm as measured?, *Int. J. Climatol.*, *23*, 1495–1521.

Moberg, A., D. Sonechkin, K. Holmgren, N. Datsenko, and W. Karlen (2005), Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature*, *433*, 613–617.

Moberg, A., R. Mohammad, and T. Mauritsen (2008), Analysis of the Moberg et al. (2005) hemispheric temperature reconstruction, *Clim. Dyn.*, in press.

Pauling, A., J. Luterbacher, C. Casty, and H. Wanner (2006), Five hundred years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation, *Clim. Dyn.*, *26*, 387–405.

Riedwyl, N., M. Küttel, J. Luterbacher, and H. Wanner (2008), Comparison of climate field reconstruction techniques: Application to Europe, *Clim. Dyn.*, in press.

Rutherford, S., M. Mann, T. Osborn, R. Bradley, K. Briffa, M. Hughes, and P. Jones (2005), Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to method, predictor network, target season, and target domain, *J. Clim.*, *18*, 2308–2329.

Schneider, T. (2001), Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values, *J. Clim.*, *14*, 853–871.

von Storch, H., E. Zorita, J. Jones, Y. Dimitriev, F. Gonzalez-Rouco, and S. Tett (2004), Reconstructing past climate from noisy data, *Science*, *306*, 679–682.

von Storch, H., E. Zorita, and F. Gonzalez-Rouco (2008), Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation, *J. Earth Sci.*, in press.

Wilks, D. (1995), *Statistical Methods in the Atmospheric Sciences*, Academic, San Diego, Calif.

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

N. Riedwyl, J. Luterbacher, and H. Wanner, Oeschger Centre for Climate Change Research, University of Bern, Hallerstrasse 12, CH-3012 Bern, Switzerland. (riedwyl@giub.unibe.ch)