European climate reconstructions back to 1000 AD - a methodological approach

Inauguraldissertation

der Philosophisch-naturwissenschaftlichen Fakultät

der Universität Bern

vorgelegt von

Nadja Riedwyl

von Kehrsatz BE

Leiter der Arbeit:

Prof. Dr. H. Wanner

Geographisches Institut, Universität Bern

European climate reconstructions back to 1000 AD - a methodological approach

Inauguraldissertation

der Philosophisch-naturwissenschaftlichen Fakultät

der Universität Bern

vorgelegt von

Nadja Riedwyl

von Kehrsatz BE

Leiter der Arbeit:

Prof. Dr. H. Wanner

Geographisches Institut, Universität Bern

von der Philosophisch-naturwissenschaftlichen Fakultät angenommen

Bern, den 18. September 2008 Der Dekan:

Prof. Dr. U. Feller

To my grandfather

Summary

This thesis focuses on climate field reconstruction methods, and their application at the seasonal scale for Europe. We present an in-depth analysis of PC regression and RegEM. “A priori” knowledge about their performance is gained through robustness and sensitivity tests in a “surrogate climate”. Furthermore, we evaluate reconstructions for European summer and winter temperatures over the last 500 years comparing PC regression to RegEM and CPS. Finally, based on all these methodological considerations, the thesis presents and interprets an ensemble of European summer temperature reconstructions over the last millennium.

Comparison of climate field reconstruction techniques: Application to Europe

We compare the performance of PC regression and RegEM to reconstruct European summer and winter surface air temperature over the past millennium. Reconstruction is performed within a surrogate climate using the National Center for Atmospheric Research (NCAR) Climate System Model (CSM) 1.4 and the climate model ECHO-G 4, assuming different white and red noise scenarios to define the distortion of pseudoproxy series. We show how sensitivity tests lead to valuable “a priori” information that provides a basis for improving real world proxy reconstructions. Furthermore, we demonstrate that uncertainties inherent to the predictand and predictor data have to be more rigorously taken into account. More skilful results are achieved with RegEM as low frequency variability is better preserved. We further detect seasonal differences in reconstruction skill for the continental scale, as e.g. the target temperature average is more adequately reconstructed for sum-

mer than for winter. Both techniques underestimate the target temperature variations to an increasing extent as more noise is added to the signal, albeit RegEM less than with PC regression. We conclude that climate field reconstruction techniques can be improved and need to be further optimized in future applications.

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

An ensemble of reconstruction results for past European temperature variability back to 1500 is presented. We apply PC regression, RegEM and CPS. The reconstruction results of the three techniques for summer and winter European temperature averages, and spatial fields related to warmest and coldest decades are analyzed and discussed. We show that PC regression and RegEM perform more similarly compared to CPS, and that 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 also very likely to be due to limited spatial and temporal availability of the proxy data. The comparison of PC regression, RegEM and CPS reveals that past temperature variability is likely more variable than indicated by earlier European seasonal temperature reconstructions, still indicating 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.

European summer temperature variability over the last millennium

We present reconstructions of European summer temperature variability over the last millennium. Reconstruction is performed using PC regression, RegEM, and additionally CPS. The combination of three reconstruction techniques and compilation of long and continuous proxy series provide the basis for these new results, and for a detailed analysis of European millennial summer temperature amplitudes. Their robustness is tested by cross-validation

with the leave-one-out algorithm. Furthermore, the performances of PC regression and RegEM are compared focusing on reconstructed temperature fields averaged over periods of accordance and discordance. We show that a rather diverse picture of summer temperature variability over the last millennium is very likely caused by a lack of coherence in the temperature signals inherent to the proxy data. We conclude that the "Medieval Warm Period" is not noticeable in the results, whereas some evidence is provided for the "Little Ice Age".

Contents

Abbreviations

AD	Anno Domini
AOGCM	Atmosphere Ocean General Circulation Model
CE	Coefficient of Efficiency
CFR	Climate Field Reconstruction
CPS	Composite-Plus-Scaling
DJF	December, January, February
i.i.d.	independent and identically distributed
JJA	June, July, August
LIA	Little Ice Age
MWP	Medieval Warm Period
NAO	North Atlantic Oszillation
NH	Northern Hemisphere
OLS	Ordinary Least Squares
PC	Principal Component
RE	Reduction of Error
RegEM	Regularized Expectation Maximization
RRMSE	Relative Root Mean Squared Error
TTLS	Truncated Total Least Squares
SE	Standard Errors
SNR	Signal-to-Noise Ratio

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Scientific background and motivation

The earth has always witnessed climate change throughout its history. However, time series of atmospheric CO_2 concentration from air bubbles in ice cores reveal highest values presently, relative to the past 650'000 years and are evidence that the present climate change is unusual. Very likely humans are the cause: without taking into account human activities, the exceptionally high concentration of CO_2 in the atmosphere can not be explained (*Jansen and coauthors*, 2007). The forth assessment report (4.AR) of the Intergovernmental Panel on Climate Change (IPCC) proves that an in-depth understanding of past climate is a necessary precondition to understand present climate, and to learn for the future. Thus, in the last few decades, substantial progress has been made in paleoclimatology with regard to the improvement of the understanding of past climate variability prior to the instrumental measurement period.

1.1.1 Models versus reconstructions

Past climate variability and changes are either assessed by climate model simulations or by analyses and reconstructions based on indirect climate information, so called proxies. Past climate signals are recorded in natural archives, which alter their behavior in response to climate (e.g. tree rings, corals, varves, peat, speleothems, ice cores), or in documentary evidence (e.g.

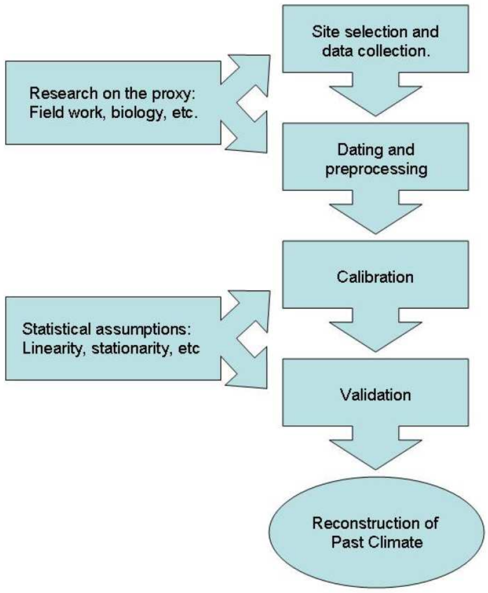

Figure 1.1: *Schematic diagram of the general methodology used to reconstruct past climates (National research Council of the US National Academy, 2006).*

weather reports, diaries, ship logbooks, grape harvest dates). While climate models allow, due to their physical nature, the linkage of cause and effect in past climate, statistical reconstructions describe past climate variability and extremes based on proxies. Thus, paleoclimate reconstructions are crucial to evaluate, how realistically climate models simulate past climatic changes. In turn, the skill and robustness of statistical reconstruction methods can be tested by using climate model simulations as a "surrogate climate".

1.1.2 Methods for climate field reconstruction

Methods for paleoclimate reconstructions range from direct measurements of past climate to reconstructions using single or multiple proxy data. These multiproxy reconstructions can offer more rigorous estimates than a single proxy approach (*Jansen and coauthors*, 2007). Reconstruction is performed with statistical methods, used to estimate past climate variables according to the following principal ideas (Fig. 1.1). Firstly, the period of instrumental measurements is separated into a calibration and a validation part. Then, a statistical relationship between the target, i.e. the instrumental measurements, and the explanatory variables, i.e. the proxy data, is built over the calibration period. Secondly, values of the climate variable of interest, including those during the validation period are then reconstructed by using the statistical relationship to predict past estimates from the proxy data. Skill is attributed to the climate reconstruction if the accuracy of reconstructed and estimated values during the validation period is positively tested (*NRC*, 2006). Thus, skill measurements are calculated in order to indicate how trustful the reconstruction result is.

Commonly used reconstruction methods have the premises of linearity and stationarity in common. It is assumed, that the statistical relationship between target and proxy is linear, and the same throughout the calibration, validation, and reconstruction period. Thus, a presumably nonstationary calibration period potentially introduces biases in the reconstruction process (*Rutherford et al.*, 2003). Results applying covariance-based methods for climate field reconstruction (CFR) indicate that using the 20th century data-rich part of instrumental measurements for calibration does not significantly bias past climate reconstructions (*Rutherford et al.*, 2003). Furthermore, errors integrated in the statistical reconstruction are presumed to be independent and identically distributed (i.i.d) gaussian white noise.

Large-scale climate reconstructions range from simple averaging and scaling of proxy data, to more complex CFR methods. For composite-plus-scaling (CPS) (e.g. *Jones and Mann*, 2004; *Esper et al.*, 2005) simply the composite, i.e. the average series of the proxy data, is built and then used for reconstruction by scaling this composite according to the standard deviation of the target during the calibration period. Focusing on CFR methods, which provide not only temporal, but also spatial climatic information, there are mainly two approaches. Multivariate principal component (PC) regression is the traditional method used to reconstruct spatial climatic patterns (*Mann et al.*, 1998, 1999; *Luterbacher et al.*, 2004; *Xoplaki et al.*, 2005; *Casty et al.*, 2005; *Pauling et al.*, 2006). PC regression seeks to reconstruct past climatic fields using truncated principal components (PC) of both the target and the proxy data. The transformation to the PC with truncated singular value

decomposition (TSVD) allows for reduction of dimensionality, still retaining most of the variability contained in the full data sets (*Wilks*, 1995). Furthermore, the first few PC's typically represent large-scale modes, e.g. the North Atlantic Oscillation (*Casty et al.*, 2005a). The regression coefficients during the calibration period are estimated by ordinary least squares (OLS), and then used to "retrodict" past climate values. The disadvantage of OLS is, that if noise is inherent to the explanatory variables in the regression, the coefficient estimates are biased, which leads to a loss in variance in the reconstruction result, and underestimation of amplitudes (*Gleser*, 1992; *Thejll and Schmith*, 2005; *Lee et al.*, 2008). Therefore, recently alternative methods have been introduced (*Allen and Stott*, 2003; *Hegerl et al.*, 2006; *Rutherford et al.*, 2005; *Mann et al.*, 2005, 2007; *Lee et al.*, 2008), not only taking into account the uncertainties of the statistical models, but also the noise inherent to the proxy data. Regularized expectation maximization (RegEM), first described by *Schneider* (2001), is a covariance-based iterative algorithm, which linearly models the relationship between missing and available values (giving plausible ones). The EM algorithm estimates the mean and the covariance of an incomplete data matrix by imputing values for missing ones (*Schneider*, 2001). Thus, with each iteration step, estimates of the mean and the covariance of the input matrix are computed, followed by the calculation of the regression coefficients and the statistical characteristics of the residuals. The iterations are repeated until the convergence criterion is fulfilled (*Schneider*, 2001). For the regularization of the EM algorithm different regularization schemes can be applied. *Mann et al.* (2007) proposed truncated total least squares (TTLS), i.e. to retain only a specific number of PC of the covariance matrix, accounting for i.i.d. gaussian white noise of equal variance inherent to both, the target and the proxies.
There have been criticisms with regard to large-scale climate field reconstructions, predominantly focusing on the validation and robustness of reconstructions (*McIntyre and McKitrick*, 2003; *von Storch et al.*, 2004; *Thejll and Schmith*, 2005; *Bürger and Cubasch*, 2005; *Smerdon and Kaplan*, 2007). With the use of "a millennial surrogate climate" *von Storch et al.* (2004) showed that temperature reconstructions very likely do not fully capture the variance on longer time scales: the bias addressed damps the amplitude of

PC regression based reconstructions, such that temperatures during cooler periods may have been colder than estimated, while periods with comparable warm temperatures as in the calibration period are likely unbiased (*NRC*, 2006). Thus, concerns about application of the reconstruction methods arise from uncertainties in the ability of the statistical methods and algorithms to recognize and reproduce the climatic signal and its variations against the noise at different timescales (*NRC*, 2006). Detailed evaluations and reassessments of climate reconstruction methods have been performed (*Rutherford et al.*, 2003, 2005; *Küttel et al.*, 2007; *Mann et al.*, 2007; *Wahl and Ammann*, 2007; *von Storch et al.*, 2007; *Moberg et al.*, 2007; *Lee et al.*, 2008) and are still needed in order to identify their flaws, and enhance their reliability. The role statistical reconstruction methods play in this context is not trivial (*NRC*, 2006).

1.1.3 The data perspective

Every climate reconstruction represents a compromise with regard to the chosen target, and to a greater extent with regard to the explanatory variables. Many times temperature only is considered, and other climate variables, e.g. pressure or precipitation, which offer complementary insights into long-term mechanisms of the climate system, are neglected. Temperature fields are preferred to reconstruct as they are more homogeneous than the precipitation ones, and offer a more stable and less erratic basis for methodological investigations.
The calibration often involves a simplification of what is very likely a more complex relationship between the the proxies and the target climate variable. The simplification is defensible (*Jansen and coauthors*, 2007). Furthermore, reconstructions implicitly presume rather ideal conditions, e.g. the associated errors being i.i.d. gaussian white noise. Therefore, it is important to understand the limitations of reconstructions, and the area of conflict between the premises of the reconstruction methods, the target instrumental measurements and the heterogeneity and quality of proxy series.

Target instrumental records

In most reconstructions, the target are instrumental measurement records. Their length and statistical properties determine the calibration period. However, only since the late 19th century the number and distribution of instrumental measurement stations have been large enough to estimate temperatures for the entire globe (*NRC*, 2006). Associated uncertainties of instrumental measurements arise from the geographic distribution, land-use changes, poor sampling and inhomogeneities (*Brohan et al.*, 2006). Errors inherent to instrumental records can deteriorate the calibration and reduce the significance of reconstructions. Improvements in methods to construct data bases of gridded global climate observations are therefore crucial (*Mitchell and Jones*, 2005).

Documentary proxy evidence and early instrumental measurements

Documentary proxies are of high precision, and provide information that relates directly to natural hazards with a socio-ecomic impact (*Pfister et al.*, 1999; *Glaser*, 2001; *Brázdil et al.*, 2005). Documentary evidence tends to best record extreme events, thus compromising not continuous, but discrete series of indexes related to climate, often lacking information about past climatic mean states. The most limiting factor in their use for statistical reconstruction, is the missing information during the calibration period. Obviously, when instrumental measurements started, the necessity of compiling documentary sources on climate stopped.
As in Europe instrumental measurements started early, here they provide a basis for comparison with natural proxy records. Not surprisingly, when early instrumental records are used for reconstruction, they correlate highly with the target. The limitations of early measurements relate to inhomogeneities. This has proved to be of particular importance for reconstructing summer temperatures. Instrumental measurements might have been too warm due to awkwardly exposure of thermometers (*Frank et al.*, 2007; *Böhm et al.*, 2008). The role of early instrumental measurement in millennial reconstructions is minor, compared to the crucial role of documentary evidence, before 1500 AD.

Tree ring records

Tree rings (ring widths and maximum late wood density) play a dominant role in paleoclimate reconstructions. They agree well with statistical premises, as they are continuous, precisely dated at annual and higher resolution, and found in many locations (*Jones and coauthers*, 2008). Although new methods have been introduced to process tree ring data appropriately for lower frequency reconstruction (*Esper et al.*, 2002; *Büntgen et al.*, 2006), often precise information about the sampling and the pre-processing of available tree ring data is missing, which introduces uncertainties in the process of reconstruction. Furthermore, the different variance spectra of different tree species are mostly not taken into account by the commonly used statistical reconstruction methods (*Jones and coauthers*, 2008). Finally, a problem of "divergence" between some tree ring data and the observed temperature trend at the end of the 20th century has been discovered (*Briffa et al.*, 1998; *Wilson et al.*, 2007), influencing considerably the calibration with and interpretation of tree rings.

Ice core isotopic records

Ice cores are a prominent proxy for reconstruction of past climate. The ice core variable, that has widely been used, is $\delta^{18}O$, the isotopic composition of water (e.g. *Grootes and Stuiver*, 1997; *Johnsen and Vinther*, 1998). Ice cores are very limited in their availability, their importance is emphasized as they are found in crucial locations (e.g. Greenland, Antarctica) where no other proxy information is available. There is no simple temperature-isotope effect, as there are exchanges between water at the surface, water droplets as well as atmospheric water vapor which have to be considered in detail (*Bradley*, 1999). Biases introduced due to non-climatic noise, e.g. surface roughness or rapid ice flow, can not yet be adequately integrated in the reconstruction process by statistical methods (*Jones and coauthers*, 2008). Furthermore, annual markers in the ice do very likely not correspond to fixed calendar dates, which introduces further uncertainties to reconstructions using ice core data (*Jones and coauthers*, 2008).

Other proxies

Many, many other proxies are available, e.g. corals, speleothems, lake and peat sediments, as well as marine sediments. Most of these proxies are difficult to incorporate in highly temporally resolved climate reconstruction. On the one hand they can not yet be dated accurately enough, on the other specific statistical methods need to be developed. Although generally not considered for multiproxy reconstruction, they provide valuable archives and records to interpret past climate variability and change also at lower frequencies. Furthermore, e.g. coral records also include information about the tropical ocean climate, and build a basis for the extension of land-only climate reconstructions. Lately, also temperature-depth profiles from boreholes are used for geothermal past climate reconstruction given a decreasing resolution in depth and in time (*González-Rouco et al.*, 2008). Also new proxy types are developed steadily, e.g. tree ring isotope records (*Treydte et al.*, 2007), which further enrich the paleoclimate proxy archives for reconstruction.

1.1.4 Northern hemispheric temperature reconstructions over the last millennium

Since the late 1990s paleoclimatologists have reconstructed annual to decadal surface air temperature variability over the northern hemisphere (NH). Many NH annual average temperature reconstructions covering approximately the last millennium have been presented (*Jones et al.*, 1998; *Mann et al.*, 1998, 1999; *Briffa et al.*, 2001; *Esper et al.*, 2002; *Mann and Jones*, 2003; *Cook et al.*, 2004; *Moberg et al.*, 2005; *Rutherford et al.*, 2005; *DArrigo et al.*, 2006; *Hegerl et al.*, 2006). Hereby different variants of CPS (e.g *Esper et al.*, 2002; *Moberg et al.*, 2005; *Hegerl et al.*, 2006; *DArrigo et al.*, 2006), as well as inverse multivariate PC regression (*Mann et al.*, 1998, 1999) have been applied. There are differences between these NH reconstructions particularly with regard to the magnitude of the past coolings in the 12th to 14th and 17th to 19th centuries. The reconstruction of *Moberg et al.* (2005) reveals persistent warm conditions comparable to the mid 20th century, while the others mostly exhibit a small maximum just before the year 1000 AD. However, none of them show temperatures before the 20th century being warmer

than levels reached for the last two decades of the 20th century (*Jansen and coauthors*, 2007). Besides NH paleoclimate reconstructions, a range of increasingly complex atmosphere-ocean general circulation models (AOGCMs), earth system models of intermediate complexity (EMICs) and energy balance models (EBMs) have been used to simulate NH temperatures over the last 500 and 1000 years using both natural and anthropogenic forcings (e.g. *Crowley et al.*, 2003; *González-Rouco et al.*, 2006; *Goosse et al.*, 2006; *Ammann et al.*, 2007). The comparisons of energy balance climate model simulations and observations disclose that much of the pre-anthropogenic decadal-scale temperature variations on the NH can be explained by changes in volcanism and solar irradiance, while only approximately 25 percent of the temperature increase in the 20th century can be attributed to natural variability (*Crowley*, 2000). Furthermore, *Mann* (2007b) concludes that natural forcings (solar and volcanic) explain reasonably major large-scale average temperature changes of the past millennium until the 19th century. However, only greenhouse gas and sulfate aerosols can explain the recent anomalous warming displayed in the NH temperature reconstructions for the late 20th century.
Recent studies emphasizing regional to continental dynamics of past climate variability on seasonal to centennial timescales, more and more have gained importance (http://www.pages.unibe.ch). At regional scale other aspects than at the hemispheric scale in the response to forcing become important, e.g. the influence of internal variability on climate variations being more dominant (*Shindell et al.*, 2003, 2004; *Bengtsson et al.*, 2006). Furthermore, *Goosse et al.* (2006) suggests land surface forcing has induced Medieval summer warmth around the 11th century in Europe, comparable to the recent warmth of the late 20th century. And *Mann* (2007b) concludes that atmospheric circulation patterns may play a crucial role in understanding variability patterns of specific regions. Thus, in contrast to the NH, where the lowering of solar irradiance during the "Maunder Minimum" in the 17th century has only lead to a moderate decrease of the average temperature, in Europe a substantial cooling was induced, due to a tendency towards the negative phase of the NAO (*Luterbacher et al.*, 2001; *Mann*, 2007b). Finally, it is important to detect the seasonal response to explosive volcanism, on smaller scales than hemispheric (*Fischer et al.*, 2007).

Smaller and seasonal scales imply different methodological aspects to be considered in the reconstruction. On the one hand, the magnitude of regional climate variability amplitudes is often greater than at the hemispheric scale (*Luterbacher et al.*, 2004), on the other, the dimensionality of the "problem" decreases considerably. Thus, CFR methods have to be tested and evaluated according to the spatial and temporal scales they are applied to.

1.2 European climate reconstructions over the last millennium

Many reconstructions at the European scale put the current climate change into the context of past climate variability covering approximately the past 500 years (*Luterbacher et al.*, 2004, 2006; *Xoplaki et al.*, 2005; *Casty et al.*, 2005, 2007; *Raible et al.*, 2006; *Pauling et al.*, 2006; *Fischer et al.*, 2007). Thus, they provide information about continental climate variability, and the intrinsic seasonal patterns of climate change in Europe, which are not visible at the hemispheric scale. Furthermore, they offer the possibility for an extension even further back in time. An accurate picture of European climate variability over the last millennium does not yet exist. Many proxy series across Europe explain past local climate variability (e.g *Kirchhefer*, 2001; *Proctor et al.*, 2002; *Chuine et al.*, 2004; *Mangini et al.*, 2005; *Büntgen et al.*, 2006; *Esper et al.*, 2007; *Blass et al.*, 2007), however they have not yet been comprised for a multiproxy approach attempting to explain past climate variability for the whole of Europe. Thus, the question, to what extent the "Medieval Warm Period" (MWP), the "Little Ice Age" (LIA), and the sharp temperature increase of the 20th century can be captured in large-scale European climate field reconstructions over the last millennium, has not yet been answered. Furthermore, as proxy data availability is exceptionally good in Europe compared to other continents, the European area is suitable for methodological investigations aiming at the reconstruction of millennial seasonal climate variability.

1.2.1 The Palvarex 2 project

This PhD thesis contributes to research work within the second phase of the project PALeoclimate VARiability and EXtreme events (PALVAREX 2) in the workpackage 1 of the National Center of Competence in Research (NCCR) in Climate (http://www.nccr-climate.unibe.ch), funded by the Swiss National science Foundation. PALVAREX 2 focuses on the further enhancement of seasonal reconstructions of past European climate based on the multi-proxy approach. The climate reconstructions performed in phase 1 for the last 500 years, are to be extended over the last millennium. The objectives of PALVAREX 2 are related to:

1. The collection and compilation of highly resolved proxy evidence from natural archives and documentary sources.
2. The understanding of the response of European past climate to natural and anthropogenic forcings, as well as internal variability, on the basis of gridded climate fields reconstructions.
3. The assessment of anomalous periods and extreme events of the past.
4. The development and in-depth understanding of methodologies.

The aims of PALVAREX 2 correspond to those of the Millennium project of the European Union. Both projects have the overall aim to answer the question, to what extent the magnitude and rate of recent climate change exceed the natural variability of European climate over the last millennium.

1.2.2 Aims of this PhD thesis

The main aim of this PhD thesis is the in-depth examination and evaluation of the traditionally used climate field reconstruction technique PC regression at the European scale, and of the statistical methods CPS and RegEM also considered for millennial European climate field reconstructions. As methodological considerations are in the foreground and center of this PhD, particularly temperature is considered, due to the reasons mentioned above. On the basis of the gained knowledge about the performance of the reconstruction methods, the creation and interpretation of temperature reconstructions

covering the last millennium is aimed at in a second step.
To achieve these aims three sub-aims are defined:

1. The creation of “a priori” knowledge about the skill and robustness of millennial reconstructions by testing plausible and worst-case input scenarios for climate field reconstructions.

2. Detect flaws and enhance the reliability of the reconstruction results, by the use of an “ensemble approach”.

3. Visualize the key role of the proxy data quality and availability in the reconstruction process.

To summarize: Prior to the reconstruction of past climate variability over the last millennium, a reassessment of the performance of the different statistical methods is achieved.

1.3 Outline

This thesis is structured as follows: After the introduction to the scientific background and the topic in chapter 1, the scientific contributions are presented in chapter 2 to chapter 4. In chapter 2, PC regression and RegEM are introduced. Their performance is compared temporally and spatially, assuming different pseudoproxy scenarios within a “surrogate climate” over the past millennium at the European scale. In chapter 3, PC regression, RegEM, and additionally CPS are then applied to real proxy data covering the past 500 years. Similarities and differences of reconstructed past European summer and winter temperature variations and extremes are analyzed and discussed. Chapter 4 presents an ensemble of European summer temperature reconstructions over the past millennium. The interpretation is complemented with analyses focusing on the robustness of the results. Finally, conclusions and perspectives are contained in Chapter 5. The appendix includes the skill assessment of PC regression and RegEM to reconstruct three climate variables (pressure, temperature and precipitation).

References

Allen, M., and P. Stott (2003), Estimating signal amplitudes in optimal fingerprinting, part I: theory, *Climate Dynamics*, *21*, 477–491.

Ammann, C., F. Joos, D. Schimel, B. Otto-Bliesner, and R. Tomas (2007), Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model., *Proc. Natl. Acad. Sci. USA (PNAS)*, *104*, 3713–3718.

Blass, A., C. Bigler, M. Grosjean, and M. Sturm (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.

Bengtsson, L., K.I. Hodges, E. Roeckner, and R. Brokopf (2006), On the natural variability of the pre-industrial European climate, *Climate Dynamics*, *27*, 743–760.

Böhm, R., M. Brunetti, J. Hiebl, P. Jones, D. Frank, and M. Maugeri (2008), The early instrumental warm-bias: A solution for long central European temperature series, *Climatic Change, in review*.

Bradley, R.S. (1999), Paleoclimatology: Reconstructing Climates of the Quarternary, *Academic Press, San Diego, USA*.

Brázdil, R., C. Pfister, H. Wanner, H. Storch, and J. Luterbacher (2005), Historical Climatology In Europe–The State Of The Art, *Climatic Change*, *70*, 363–430.

Briffa, K., F. Schweingruber, P. Jones, T. Osborn, S. Shiyatov, and E. Vaganov (1998), Reduced sensitivity of recent tree-growth to temperature at high northern latitudes, *Nature*, *391*, 678–682.

Briffa, K., T. Osborn, F. Schweingruber, I. Harris, P. Jones, S. Shiyatov, and E. Vaganov (2001), Low-frequency temperature variations from a northern tree ring density network, *Journal of Geophysical Research*, *106*, 2929–2942.

Brohan, P., J. Kennedy, I. Harris, S. Tett, and P. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850, *Journal of Geophysical Research*, *111*, D12106.

Bürger, G., and U. Cubasch (2005), Are multiproxy climate reconstructions robust?, *Geophysical Research Letters*, *32*, L23711.

Büntgen, U., D. Frank, D. Nievergelt, and J. Esper (2006), Summer Temperature Variations in the European Alps, ad 755–2004, *Journal of Climate*, *19*, 5606–5623.

Casty, C., D. Handorf, and M. Sempf (2005a), Combined winter climate regimes over the North Atlantic/European sector 1766–2000, *Geophysical Research Letters*, *32*, L13801.

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

Casty, C., C. Raible, T. Stocker, H. Wanner, and J. Luterbacher (2007), European climate pattern variability since 1766, *Climate Dynamics, in press.*

Chuine, I., P. Yiou, N. Viovy, B. Seguin, V. Daux, and E. Le Roy Ladurie (2004), Historical phenology: Grape ripening as a past climate indicator, *Nature*, *432*, 289–290.

Cook, E., J. Esper, and R. D'Arrigo (2004), Extra-tropical Northern Hemisphere land temperature variability over the past 1000 years, *Quaternary Science Reviews*, *23*, 2063–2074.

Crowley, T.J. (2000), Causes of Climate Change Over the Past 1000 Years, *Science*, *289*, 270–277.

Crowley, T.J., S.K. Baum, K.Y. Kim, G.C. Hegerl, and W.T. Hyde (2003), Modeling ocean heat content changes during the last millennium, *Geophysical Research Letters*, *30*, 1932, doi:10.1029/2003GL017801.

DArrigo, R., R. Wilson, and G. Jacoby (2006), On the long-term context for late twentieth century warming, *Journal of Geophysical Research*, *111*, doi:10.1029/2005JD006352.

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. Frank, R. Wilson, and K. Briffa (2005), Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, *Geophysical Research Letters*, *32*, L07711.

Esper, J., D. Frank, U. Buentgen, A. Verstege, J. Luterbacher, and E. Xoplaki (2007), Long-term drought severity variations in Morocco, *Geophysical Research Letters*, *34*, L17702.

Fischer, E., J. Luterbacher, E. Zorita, S. Tett, C. Casty, and H. Wanner (2007), European climate response to tropical volcanic eruptions over the last half millennium, *Geophysical Research Letters*, *34*, L05707.

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, *Quaternary Science Reviews*, *26*, 3298–3310.

Glaser, R. (2001), Klimageschichte Mitteleuropas. 1000 Jahre Wetter, Klima, Katastrophen, *Wiss. Buchgesellschaft, Darmstadt.*

Gleser, L. (1992), The importance of assessing measurement reliability in multivariate regression, *Journal of the American Statistical Association*, *87*, 696–707.

González-Rouco, J., H. Beltrami, E. Zorita, and H. von Storch (2006), Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling, *Geophysical Research Letters*, *33*, L01703.

González-Rouco, J., H. Beltrami, E. Zorita, and M. Evans (2008), Borehole climatology: a discussion based on contributions from climate modelling, *Climate of the Past Discussions*, *4*, 1–80.

Goosse, H., O. Arzel, J. Luterbacher, M. Mann, H. Renssen, N. Riedwyl, A. Timmermann, E. Xoplaki, and H. Wanner (2006), The origin of the European "Medieval Warm Period", *Climate of the Past*, *2*, 99–113.

Grootes, P., and M. Stuiver (1997), Oxygen 18/16 variability in Greenland snow and ice with 10 33 to 10 5-year time resolution, *Journal of Geophysical Research*, *102*, 26455–26470.

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

Jansen, E., and coauthors (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*, Cambridge University Press, Cambridge, United Kingdom and New York.

Johnsen, S., and B. Vinther (1998), Stable isotope records from the Greenland Ice Cores, *Encyclopedia of Quaternary Science*, 1250–1258.

Jones, P., and coauthers (2008), High-resolution paleoclimatology of the last millennium: a review of curent status and future prospects, *Holocene, in review.*

Jones, P., and M. Mann (2004), Climate over past millennia, *Reviews of Geophysics*, *42*, RG2002.

Jones, P., K. Briffa, T. Barnett, and S. Tett (1998), High-resolution paleoclimatic records for the last millennium: Interpretation, integration and comparison with circulation model control run temperatures, *Holocene*, *8*, 455–471.

Kirchhefer, A. (2001), Reconstruction of summer temperatures from tree-rings of Scots pine (Pinus sylvestris L.) in coastal northern Norway, *The Holocene*, *11*, 41–52.

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, *Geophysical Research Letters*, *34*, L07710.

Lee, T., F. Zwiers, and M. Tsao (2008), Evaluation of proxy-based millennial reconstruction methods, *Climate Dynamics, in press.*

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 forStudying Decadal Scale Climatic Change in Europe, *Climatic 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.

Luterbacher, J., Eds. P. Lionello, P. Malanotte-Rizzoli, and R. Boscolo (2006), Mediterranean climate variability over the last centuries: A review in: *The Mediterranean Climate: an overview of the main characteristics and issues*, Elsevier, Amsterdam, the Netherlands, 27–148.

Mangini, A., C. Spötl, and P. Verdes (2005), Reconstruction of temperature in the Central Alps during the past 2000 yr from a $\delta 18$O stalagmite record, *Earth and Planetary Science Letters*, *235*, 741–751.

Mann, M. (2007), Climate Over the Past Two Millennia, *Annual Review of Earth and Planetary Sciences*, *35*, 111–136.

Mann, M., and P. Jones (2003), Global surface temperatures over the past two millennia, *Geophysical Research Letters*, *30*, 1820.

Mann, M., R. Bradley, and M. Hughes (1998), Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, *392*, 779–787.

Mann, M., R. Bradley, and M. Hughes (1999), Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophysical Research Letters*, *26*, 759–762.

Mann, M., S. Rutherford, E. Wahl, and C. Ammann (2005), Testing the Fidelity of Methods Used in Proxy-Based Reconstructions of Past Climate, *Journal of Climate*, *18*, 4097–4107.

Mann, M., S. Rutherford, E. Wahl, and C. Ammann (2007), Robustness of proxy-based climate field reconstruction methods, *Journal of Geophysical Research*, *112*, D12109.

McIntyre, S., and R. McKitrick (2003), Corrections to the Mann et al.(1998) proxy data base and northern hemispheric average temperature series, *Energy and Environment*, *14*, 751–771.

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

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 (2007), Analysis of the Moberg et al. (2005) hemispheric temperature reconstruction , *Climate Dynamics, in press.*

NRC, National Reseach Council of the US National Academy (2006), Surface Temperature Reconstructions for the last 2000 Years, *(The National Press)*, 44–93.

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, *Climate Dynamics*, *26*, 387–405.

Pfister, C., H. Graßl, and J. Luterbacher (1999), *Wetternachhersage: 500 Jahre Klimavariationen und Naturkatastrophen (1496-1995)*, Paul Haupt, Bern.

Proctor, C., A. Baker, and W. Barnes (2002), A three thousand year record of North Atlantic climate, *Climate Dynamics*, *19*, 449–454.

Raible, C., et al. (2006), Climate Variability-Observations, Reconstructions, and Model Simulations for the Atlantic-European and Alpine Region from 1500-2100 AD, *Climatic Change*, *79*, 9–29.

Rutherford, S., M. Mann, T. Delworth, and R. Stouffer (2003), Climate Field Reconstruction under Stationary and Nonstationary Forcing, *Journal of Climate*, *16*, 462–479.

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, *Journal of Climate*, *18*, 2308–2329.

Schneider, T. (2001), Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values., *Journal of Climate*, *14*, 853–871.

Shindell, D., G. Schmidt, R. Miller, and M. Mann (2003), Volcanic and Solar Forcing of Climate Change during the Preindustrial Era, *Journal of Climate*, *16*, 4094–4107.

Shindell, D., G. Schmidt, M. Mann, and G. Faluvegi (2004), Dynamic winter climate response to large tropical volcanic eruptions since 1600, *Journal of Geophysical Research*, *109*, D05104.

Smerdon, J., and A. Kaplan (2007), Comments on Testing the Fidelity of Methods Used in Proxy-Based Reconstructions of Past Climate: The Role of the Standardization Interval, *Journal of Climate*, *20*, 5666–5670.

Thejll, P., and T. Schmith (2005), Limitations on regression analysis due to serially correlated residuals: Application to climate reconstruction from proxies, *Journal of Geophysical Research*, *110*, D18103.

Treydte, K., et al. (2007), Signal strength and climate calibration of a European tree-ring isotope network, *Geophysical Research Letters*, *34*, L24302.

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 (2007), Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation, *Journal of Earth Sciences, in press.*

Wahl, E., and C. Ammann (2007), Robustness of the Mann, Bradley, Hughes reconstruction of Northern Hemisphere surface temperatures: Examination of criticisms based on the nature and processing of proxy climate evidence, *Climatic Change*, *85*, 33–69.

Wilks, D. (1995), *Statistical methods in the atmospheric sciences*, Academic Press San Diego.

Wilson, R., et al. (2007), A matter of divergence: Tracking recent warming at hemispheric scales using tree ring data, *Journal of Geophysical Research*, *112*, D17103.

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, *Geophysical Research Letters*, *32*, L15711.

Chapter 2

Comparison of climate field reconstruction techniques: Application to Europe

Nadja Riedwyl, Marcel Küttel, Jürg Luterbacher and Heinz Wanner

Oeschger Centre for Climate Change Research (OCCR), Bern, Switzerland
Institute of Geography, Climatology and Meteorology, University of Bern, Bern, Switzerland

Climate dynamics, 2008, doi 10.1007/s00382-008-0395-5

Abstract

This paper presents a comparison of principal component (PC) regression and regularized expectation maximization (RegEM) to reconstruct European summer and winter surface air temperature over the past millennium. Reconstruction is performed within a surrogate climate using the National Center for Atmospheric Research (NCAR) Climate System Model (CSM) 1.4 and the climate model ECHO-G 4, assuming different white and red noise scenarios to define the distortion of pseudoproxy series. We show how sensitivity tests lead to valuable “a priori” information that provides a basis for im-

proving real world proxy reconstructions. Our results emphasize the need to carefully test and evaluate reconstruction techniques with respect to the temporal resolution and the spatial scale they are applied to. Furthermore, we demonstrate that uncertainties inherent to the predictand and predictor data have to be more rigorously taken into account. The comparison of the two statistical techniques, in the specific experimental setting presented here, indicates that more skilful results are achieved with RegEM as low frequency variability is better preserved. We further detect seasonal differences in reconstruction skill for the continental scale, as e.g. the target temperature average is more adequately reconstructed for summer than for winter. For the specific predictor network given in this paper, both techniques underestimate the target temperature variations to an increasing extent as more noise is added to the signal, albeit RegEM less than with PC regression. We conclude that climate field reconstruction techniques can be improved and need to be further optimized in future applications.

2.1 Introduction

Knowledge of temperature amplitudes is of utmost importance in gaining a better understanding of past temperature evolution and change. Reconstruction of past temperature variability based on paleoclimatic data can provide insights into the interpretation of the role of climatic forcings. Many existing reconstructions place the twentieth century warming at continental to global scale into a broader context (*Mann et al.*, 1998, 1999, 2005; *Esper et al.*, 2002; *Luterbacher et al.*, 2004, 2007; *Mann and Rutherford*, 2002; *Xoplaki et al.*, 2005; *Rutherford et al.*, 2005; *Casty et al.*, 2005a, 2007; *Guiot et al.*, 2005; *Moberg et al.*, 2005; *Jansen and coauthors*, 2007). However these reconstructions have various limitations, primarily related to the availability of proxy data and their quality. It is a methodological challenge to filter out the climatic signal from a range of different proxy archives, given the short instrumental period for calibration and the increasing lack of predictors back in time. Reconstruction is generally approached in two ways. One possibility is to reconstruct the average, i.e. a single time series over a specific time period, e.g. the Northern Hemisphere average over the past millennium. The average series is reconstructed by making a composite of multiple proxy series, centered and scaled according to the target, i.e. composite-plus-scaling (CPS) (see *Jones and Mann*, 2004; *Esper et al.*, 2005). The other possibility is to focus on the whole climatic field of interest. In this case climate field reconstruction (CFR) techniques provide temporal and spatial information (*Jones and Mann*, 2004, and references therein). The CFR approach provides a distinct advantage over averaged climate reconstructions, for instance, when information on the spatial response to external forcing (e.g. volcanic, solar) is sought (e.g. *Shindell et al.*, 2001, 2003, 2004; *Waple et al.*, 2002; *Fischer et al.*, 2007). The results of both approaches, CFR and CPS, have led to some controversy over temperature amplitudes, raising questions about associated uncertainties, and the robustness and skill of the various reconstructions, as well as the influence of trends, and the length and climatology of the calibration period (*von Storch et al.*, 2004, 2007; *Bürger and Cubasch*, 2005; *Thejll and Schmith*, 2005; *Wahl and Ammann*, 2007; *Moberg et al.*, 2007). Recent studies provide some answers to these questions (*Wahl and Ammann*, 2007) and introduce improved methodologies for reconstructions, e.g. the application of different parameter estimation techniques (*Schneider*, 2001; *Hegerl et al.*, 2006), the use of wavelet analysis (*Moberg et al.*, 2005) or state space models (*Lee et al.*, 2008). In this contribution we concentrate on CFR techniques.

Principal component (PC) regression is the classical method used to reconstruct past European climate field information and has been widely applied (*Briffa et al.*, 1987; *Cook et al.*, 1994; *Luterbacher et al.*, 2004; *Casty et al.*, 2005, 2007; *Xoplaki et al.*, 2005; *Pauling et al.*, 2006). With PC regression, CFR is commonly performed under the assumption that no errors are inherent to the predictor data, and regression coefficient estimates are achieved using ordinary least squares (OLS). However, if noise is inherent to the predictor data, these estimates are negatively biased towards an underestimation that results in loss of variance (*Lee et al.*, 2008). Several authors (*Hegerl et al.*, 2006; *Mann et al.*, 2005, 2007; *Rutherford et al.*, 2005; *Brohan et al.*, 2006; *Esper et al.*, 2007; *Lee et al.*, 2008; *Li et al.*, 2007) have recently discussed the necessity of taking into account not only the uncertainties of the statistical model, i.e. the residuals, but also the errors inherent to the predictand and predictor data:

$$\mathbf{Y} + \mathbf{e}_{instr} = \mathbf{B}(\mathbf{X} + \mathbf{e}_{proxy}) + \mathbf{e} \tag{2.1}$$

where $\mathbf{e}$ are the residuals, $\mathbf{e}_{instr}$ the errors associated with the instrumental measurements, i.e. the predictand and $\mathbf{e}_{proxy}$ the errors associated with the predictors. Thus the methodological problems can be partly solved by better incorporating the different uncertainties in the statistical reconstruction models. Studies by *Schneider* (2001), *Mann et al.* (2005, 2007) and *Rutherford et al.* (2005) have proved the capability of the Regularized Expectation Maximization (RegEM) algorithm to more accurately reconstruct past temperature variations. One reason for this is that RegEM integrates $\mathbf{e}_{proxy}$ in the reconstruction technique, as ill-posed problems are regularized. *Mann et al.* (2007) found truncated total least squares (TTLS) to be a particularly successful option for undertaking the regularization. RegEM with TTLS is used here as proposed by *Mann et al.* (2007) and following the instructions therein.
Some of the studies mentioned above found differences between the results obtained by using principal component (PC) regression, on the one hand, and those achieved by means of the more sophisticated RegEM approach, on the other. However, these studies are limited to the hemispheric to global scale and, mainly, to annual resolution (*Rutherford et al.*, 2005; *Mann et al.*, 2007; *Lee et al.*, 2008). One might expect to obtain different results when applying these techniques at a smaller spatial scale, such as Europe, and considering seasonal, rather than annual, data. In this study we therefore examine the sensitivity of the reconstruction skill at the continental scale, with seasonally resolved synthetic proxy data, i.e. proxies derived from climate model data. We use data from two simulations -one generated by the National Center for Atmospheric Research (NCAR) Climate System Model (CSM) 1.4 (*Ammann et al.*, 2007), and the other generated by ECHO-G 4, which consists of the atmosphere and ocean general circulation models (GCM) ECHAM4 and HOPE-G (*González-Rouco et al.*, 2006). Both simulations are likely to provide realistic opportunities for testing CFR approaches (*Mann et al.*, 2005, 2007; *von Storch et al.*, 2004; *González-Rouco et al.*, 2006; *Lee et al.*, 2008). Utilizing climate model data in a systematic experiment setup to at-

tempt to reconstruct simulated past temperatures helps to understand the two techniques better. This would be less easily undertaken with real world multiproxy data as input, due to their heterogeneous nature and limited availability. The evaluation of CFR techniques is an important step in the process of identifying methodological deficits and limitations, providing "a priori" knowledge about the performance of the methodologies. Testing the techniques is therefore a good preparation for the next step: the improvement of reconstruction using real world proxy data. Apart from the choice of the reconstruction technique, there are several other factors limiting the skill of reconstructions of past climate variability, e.g. the varying number and spatial distribution of proxies over time (*Pauling et al.*, 2003; *Küttel et al.*, 2007; *Mann et al.*, 2007). However, here we focus on three things: on the dependence of reconstruction skill on a specific predictor network, comparable in size and spatial distribution to a millennial European real world network, on the two techniques applied, and on the quality of the predictor data. We evaluate RegEM (*Schneider*, 2001; *Rutherford et al.*, 2005; *Mann et al.*, 2007) for European summer and winter temperatures over the past millennium. In this study, RegEM is for the first time applied to spatial scales smaller than the hemispheric. Furthermore, we compare RegEM to PC regression, the basic multivariate regression model applied at the European scale, e.g. in *Luterbacher et al.* (2004, 2007), *Casty et al.* (2005, 2007) and *Xoplaki et al.* (2005). In section 2 we describe the NCAR CSM 1.4 and ECHO-G 4 climate model data and the experimental setting. Then we introduce the two CFR techniques and the criteria for comparison. In section 3 we present the results. We begin by looking at the European average temperatures and diagnosing the skill. Then, we evaluate the spatial skill. The results are compared and discussed in section 4, followed by a summary of our principal conclusions and a glance at future research in section 5.

2.2 Data and Methods

We test the performance of PC regression and RegEM in the surrogate climate of the two global coupled models NCAR CSM 1.4 and the ECHO-G 4. The use of climate model data permits an evaluation of the skill of the European reconstructions over a time period of 1000 years and not only during the twentieth century verification period, as would be the case in reality. The brevity of the real world instrumental period for calibration makes it very difficult to compare techniques and assess reliability of their performance (e.g. *Lee et al.*, 2008). Moreover, different virtual scenarios can be created by altering the input data of the statistical models, in order to better understand their performance and their sensitivities.

2.2.1 Simulated European surface air temperature data

NCAR CSM 1.4 (*Ammann et al.*, 2007) and ECHO-G 4 (*González-Rouco et al.*, 2006) are both global coupled models. NCAR CSM 1.4 has a grid resolution of $3.75° \times 3.75°$ and is forced over the period 850 to 1999 AD. ECHO-G 4 has a grid resolution of $3.75° \times 3.75°$ for the atmospheric component and $2.8° \times 2.8°$ at low latitudes for the ocean, and is forced over 1000 to 1990 AD. NCAR CSM 1.4 forcings included are observation-based time histories of solar irradiance, aerosol loadings from explosive volcanism, greenhouse gases and anthropogenic sulfate aerosols (*Ammann et al.*, 2007). Orbital parameters and land use changes are not included as forcings in NCAR CSM 1.4. Any potential long-term drift is removed by subtracting a millennial-scale spline fit for individual months of the annual cycle, obtained from the control integration, at each gridpoint (*Ammann et al.*, 2007). ECHO-G 4 forcing includes natural (solar irradiance, radiative effects of stratospheric volcanic aerosols) and anthropogenic (greenhouse gas concentrations) estimates (*González-Rouco et al.*, 2006) of past millennial external forcings. A flux adjustment constant in time and zero spatial average are used to inhibit climate drift (*González-Rouco et al.*, 2006). The NCAR CSM 1.4 simulation used here is the one with 'medium' solar irradiance scaling (0.25% Maunder Minimum reduction) in the terminology of *Ammann et al.* (2007). The ECHO-G 4 simulation (using 0.3% Maunder Minimum reduction) is the one sometimes known as 'Erik 2' (*González-Rouco et al.*, 2006), which has cooler initial conditions than the older 'Erik 1' simulation used in several previous pseudoproxy studies.

The predictand in the reconstruction experiments is the simulated gridded surface air temperature field, generated by the NCAR CSM 1.4 and the ECHO-G 4 simulations respectively. To represent Europe we selected the area 52.5° W to 71.25° E and 28.125° N to 76.875° N of the global model run, which gives a rather coarse picture of the European area, namely 476 gridboxes (land and sea). Gridded model surface temperature information with a higher spatial resolution is not available for the past millennium. Nevertheless, testing and comparing CFR techniques in this experimental setting is reasonable. The original NCAR CSM 1.4 and ECHO-G 4 simulation temperature data are monthly resolved. We have calculated seasonal mean temperatures for summer (JJA) and winter (DJF) starting in December 1000 AD and ending in August 1990 AD. Some analyses are made on a gridpoint basis, while others are made for the (latitude weighted) European average temperature.

2.2.2 The Pseudoproxy data

The predictor data used for this study corresponds to NCAR CSM 1.4 and ECHO-G 4 model gridpoints closest to real world proxy locations in Europe. As the proxies are derived from the model data, we call them synthetic proxies or "pseudoproxies" (*Mann and Rutherford*, 2002; *Rutherford et al.*, 2005; *von Storch et al.*, 2004, 2006). The pseudoproxy locations are chosen according to published data (*Mann et al.*, 1999; *Briffa et al.*, 2001; *Klimenko et al.*, 2001; *Proctor et al.*, 2002; *Shabalova and van Engelen*, 2003; *Luterbacher et al.*, 2004, 2007; *Casty et al.*, 2005; *Rutherford et al.*, 2005; *Guiot et al.*, 2005; *Mangini et al.*, 2005) and some other data that will be potentially available from current research projects (NCCR Climate and MILLENNIUM). The real world proxy data referred to consists of 1000 year long series and some series covering several centuries. Additionally, a few gridpoints refer to shorter real world series, which are primarily used to optimize the spatial distribution of the network towards Eastern and Southern Europe. We argue that if the techniques already fail using input data covering the full length of 1000 years, they certainly can be expected to do so, if the number and spatial distribution are reduced and change through time. Thus the pseudoproxies derived and used in this paper are idealized, as we assume them

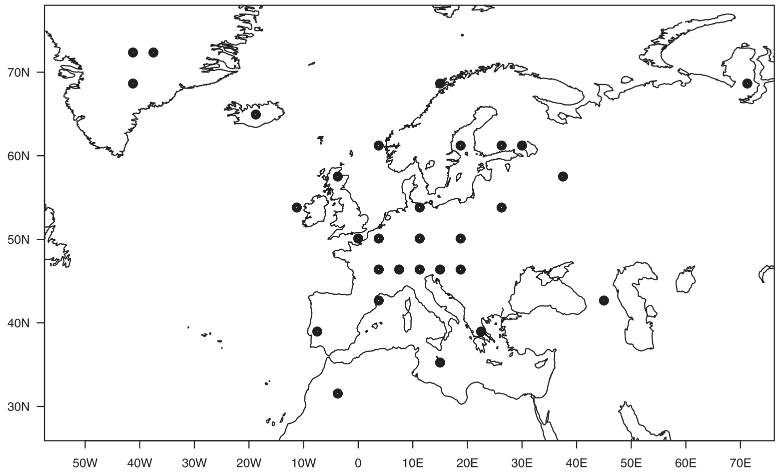

Figure 2.1: *The distribution of the 30 pseudoproxies used in this study. Each dot corresponds to the north-western corner of one* $3.75° \times 3.75°$ *gridbox of the NCAR CSM 1.4 and ECHO-G 4 model.*

all to be constantly available over the full time period of 1000 years (e.g. as been done in *Mann et al.* (2007)). Keeping the spatial distribution and the number of proxies constant over time allows us to focus on the actual variable of interest, the performance of the two different CFR techniques. We limit ourselves to considering mainly one network, equal for both seasons and without changing availability over time. The predictor network (Fig. 2.1) consists of 30 gridpoints and is seen as a reasonable selection of a predictor network for a 1000 year European temperature reconstruction. Additional testing has been made with a smaller pseudoproxy network, which consists of 12 gridpoints (not shown). These pseudoproxies refer to real world proxy series available to reconstruct the late Maunder Minimum (*Küttel et al.*, 2007). The conclusions drawn are conditional upon the specific network configuration considered. Accordingly, this study can not apply in complete generality. Moreover, we restrict our analysis based on the assumption that our pseudoproxies have seasonal resolution and do not combine temporally low and high resolved climate proxies such as those for instance in *Moberg et al.* (2005). Generally, the quantity and, even more, the spatial distribution of the proxy information plays a crucial role in determining the reconstruction skill. Even a single point, if optimally situated, has an impact on the reconstruction result, and thus improves the skill (*Küttel et al.*, 2007). However, the focus of this study lies more on the performance of the two reconstruction techniques as such.

We use different scenarios for errors in the local pseudoproxy series, i.e. the predictors are characterized by the addition of red or white noise with varying signal to noise ratios (SNR) to the simulated temperature signal. Noise is added as indicated e.g. in *Mann and Rutherford* (2002) and *von Storch et al.* (2004), with the difference that the pseudoproxies here are constructed based on seasonal means, correlations etc., i.e. separately for summer and winter, taking into account different responses of real-world proxy data to warm and cold season conditions.

The predictand is regarded as "perfect", i.e. no noise is added. The noise is intended to mimic errors inherent to the predictor data (Equation 2.1). White noise is added to be consistent with the premises given by the regression model used, i.e. the residuals are independent and identically distributed (i.i.d.). We have selected the five SNR 0.25, 0.4, 0.5, 1 and ∞ (no added noise) according to *Mann et al.* (2007). With $r = SNR/(1 + SNR^2)^{1/2}$ the SNR is related to the associated root-mean-square correlation between the predictor data and their associated local climate signal (*Mann et al.*, 2007). We obtain r = 0.24, 0.37, 0.45, 0.71 and 1.0 for the five SNR values under consideration, respectively (*Mann et al.*, 2007). As it is plausible that errors in proxy series are serially autocorrelated, we use red noise to make the uncertainties more realistic. The red noise is modeled as a first-order autoregressive AR(1) process (*Mann et al.*, 2007) and represented by $X_t = \phi X_{t-1} + Z_t$, where $Z_t \sim WN(o, \sigma^2)$ and $\phi \neq 0$. For AR(1) processes the autoregressive parameter ϕ is equal to the sample lag-1 autocorrelation coefficient ρ, here $\rho = 0.32$, 0.71. The sample lag-1 autocorrelation coefficients for red noise as well as the five SNR for white noise are the same as those evaluated in *Mann et al.* (2007). This allows for direct comparison, making it possible to determine whether RegEM performs better than PC regression at the continental scale as well, and how the increase in temperature variability due to the downscaling affects the reconstruction results.

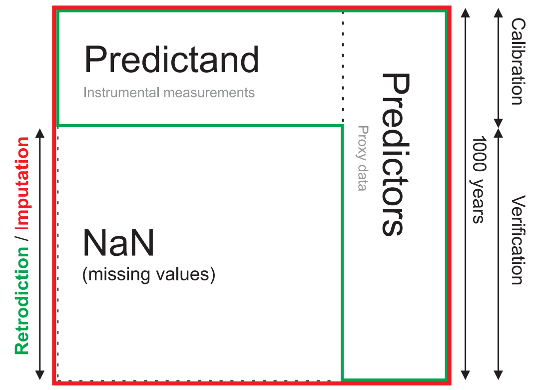

Figure 2.2: *Scheme of the analogousness / differences between PC regression (green) and RegEM (red). PC regression corresponds to "retrodiction" and RegEM to the imputation of past temperature values. The input matrix for both techniques is indicated in colors.*

2.2.3 PC regression versus RegEM

RegEM was first described by *Schneider* (2001). It has only recently been further developed and implemented by *Rutherford et al.* (2005) and *Mann et al.* (2007), and is compared to PC regression in the present paper. The two reconstruction techniques each take a different approach to the reconstruction "problem" (Fig. 2.2). With PC regression, past temperature values are "retrodicted", i.e. predicted into the past, whereas with RegEM missing values are imputed, i.e. missing values are replaced by plausible ones. While for RegEM the input is the whole data matrix including the missing and available values, as indicated in red (Fig. 2.2), for PC regression only the available predictand and predictor values are part of the input, as shown in green (Fig. 2.2).

2.2.4 Multivariate principal component regression

Multivariate PC regression seeks to reconstruct the past temperature field using the principal components of both the predictand and the predictors:

$$\mathbf{y}_{pc} = \mathbf{x}_{pc}\mathbf{B} + \mathbf{e} \tag{2.2}$$

where $\mathbf{B}$ are the regression coefficients relating the explanatory variables $\mathbf{x}_{pc}$, i.e. the predictor information, and the target $\mathbf{y}_{pc}$, i.e. the predictand. The relationship is assumed to be a linear function of parameters stationary over time. The regression coefficients of the calibration period, here $\mathbf{B}$, are estimated by OLS and then used to "retrodict" past temperature values. Predictand and predictors are transformed to their principal components to obtain orthogonal series and make it possible to reduce the dimensionality of the data while still retaining most of the variability contained in the full dataset (*Wilks*, 1995). This allows for climatic interpretation of temperature fields, as first few principal components typically capture large-scale modes. Here the calculation of the principal components is based on the correlation matrix as for instance in *Luterbacher et al.* (2004). Furthermore, they are truncated as in that study, i.e. most of the variance is captured by considering only the most important directions of the joint variations, thus avoiding redundancy (*Wilks*, 1995).

2.2.5 Regularized Expectation Maximization

RegEM is a covariance-based iterative CFR technique based on the idea of gradual linear modeling of the relationship between missing values and available values, also taking into account ill-posed or under-determined settings (*Mann et al.*, 2007). The input data matrix combines both predictand and predictor data over the full reconstruction period:

$$\mathbf{x}_m = \mu_m + (\mathbf{x}_a - \mu_a)\mathbf{B} + \mathbf{e} \tag{2.3}$$

where $\mathbf{B}$ refers to the regression coefficients relating available values $\mathbf{x}_a$ and missing values $\mathbf{x}_m$ within the multivariate data set. $\mathbf{e}$ is the random vector representing the error with mean zero and the according covariance matrix $\mathbf{C}$ to be determined (*Schneider*, 2001; *Mann et al.*, 2007). The conventional iterative Expectation Maximization algorithm (EM) estimates the mean and the covariance matrix of an incomplete data matrix and imputes values for the missing ones (*Schneider*, 2001). The EM algorithm is used under the assumption that the predictand and predictor data are Gaussian. With each iteration step, estimates of the mean μ and the covariance-variance matrix

Σ of the input matrix are calculated, followed by the computation of estimates of the coefficient matrix $\mathbf{B}$ and the residual covariance matrix $\mathbf{C}$. The iteration is repeated step by step until the convergence criterion is fulfilled (*Schneider*, 2001; *Mann et al.*, 2007).
In cases where the number of variables exceeds sample size the EM algorithm has to be regularized, as ill-posed problems lead to singularity of the covariance-variance matrix Σ (*Schneider*, 2001). Instead of estimating the coefficients $\mathbf{B}$ by the conditional maximum likelihood method given the estimates of μ and Σ, the parameters are estimated by truncated total least squares (TTLS). Thus, in order to regularize the covariance matrix Σ its principal components are truncated, i.e. only a specific number of principal components is considered, according to the truncation parameter. For further information and a more detailed description of RegEM see *Schneider* (2001), *Rutherford et al.* (2005) and *Mann et al.* (2007). In our study the non-hybrid, revised version of RegEM is used (*Mann et al.*, 2007). We standardized the available values with regard to the calibration period, 1900 to 1990 AD, to ensure that the testing of climate reconstruction methods relies on the appropriate application of real world constraints (e.g. *Smerdon and Kaplan*, 2007). The truncation parameters for TTLS are chosen in two ways. The first is as explained in *Mann et al.* (2007). *Mann et al.* (2007) identify optimal truncation parameters based on the estimate of the noise continuum to the log-eigenvalue spectrum (*Wilks*, 1995). This procedure serves to determine leading eigenvalues that lie above the estimated noise continuum. The second way is by evaluating a range of possible other truncation parameters and then selecting the parameters leading to reconstruction results with smallest differences in mean and standard deviation to the target over the verification period. As stated in *Mann et al.* (2007), the choice of the truncation parameters is not unique. This is illustrated here: validation scores of reconstruction results obtained with the log-eigenvalue spectrum criteria are shown together with those of reconstruction results (see supplementary online material) using alternative truncation parameters.
Furthermore, the reconstructions were performed both with and without the principal components of the predictand. However, analyses indicated that results using or not using PC analysis do not differ much (not shown), and

therefore, in this paper, we restrict our results to the case of not using the principal components of the predictand. In this way another ambiguous choice is avoided and the whole range of variability is retained for reconstruction.

2.2.6 The comparison criteria

PC regression and RegEM are compared to each other in the same experimental setting. As mentioned above, the reconstructions are performed within the surrogate climate of the NCAR CSM 1.4 and ECHO-G 4 climate models using 30 pseudoproxies with different SNR, all constant over time. We investigate how and to what extent the quality of the predictor data affects the reconstruction skill. Furthermore, we evaluate the results of the two techniques. On the one hand, the skill of the reconstructions is analyzed focusing on the European average only. For this reason, figures display the target, the European average temperature from 1001 to 1990 AD, in comparison to the reconstruction results, accompanied by a quantitative summary of the skill. The commonly used reduction of error (RE) and coefficient of efficiency (CE) skill scores are calculated. Tables 2.1 and 2.2 indicate the RE and CE skill scores over the verification period 1001 to 1899 AD, both for NCAR CSM 1.4 and ECHO-G 4. On the other hand, we concentrate on the climate field information, i.e. the spatial patterns. Our focus here lies on the averaged reconstruction bias, and RE calculated for the 30-year filtered reconstruction results at each gridpoint, both over the verification period 1001 to 1899 AD. In first comparing the target with the reconstruction results for each technique separately, and subsequently comparing the results of PC regression with those of RegEM, we determine how well the techniques perform, depending on the influence of the errors inherent to the predictor information.

2.3 Results

2.3.1 Impact of the quality of the predictor data

The subsequent figures all refer to results obtained using NCAR CSM 1.4, whereas the results produced with ECHO-G 4 are provided in the supplementary online material, with the exception of the skill scores tables (Tables 2.1 and 2.2), which are shown for both climate models.

Figures 2.3 and 2.4 show the methodological comparison for averaged European summer (Fig. 2.3, suppl. Fig. 2.3) and winter (Fig. 2.4, suppl. Fig. 2.4) temperature reconstructions from 1001 to 1990 AD (land and sea). The figures display temperature anomalies with regard to the calibration period 1900 to 1990 AD. The target, i.e. the average of the simulated European surface air temperature over the past millennium, is shown in black, while the average of the reconstructed summer and winter temperature fields are given in color. All curves are smoothed with a 30-year running mean. The results differ according to the five white noise scenarios used in the reconstructions. The NCAR CSM 1.4 target exhibits variability with quite large quasi-periodic amplitude variations over the past millennium, both in summer and in winter. The variability for summer and winter average temperatures is similar to that exhibited by the ECHO-G 4 run (*von Storch et al.*, 2004; *González-Rouco et al.*, 2006). The reconstructions realized with PC regression (Fig. 2.3, suppl. Fig. 2.3, top) and the perfect pseudoproxy set, i.e. no white noise added (yellow line), capture the target very well. However, the more white noise is added to the signal, the more this technique fails to properly reconstruct, and underestimates the amplitude of the target temperature variations. Thus, the difference between negative temperature anomalies of the reconstruction results and the calibration period mean is not as large as that of the target and the calibration period mean, i.e. the reconstruction being to warm. There is a shift from the scenarios with higher SNR, SNR ∞ and SNR 1 (yellow and red lines) to those with lower SNR, SNR 0.5, 0.4, 0.25 (blue and green lines), and a decrease in skill indicated by the RE and CE scores in Tables 2.1 and 2.2. The RegEM reconstruction result of the SNR ∞ scenario captures the target well, too. In comparison to PC regression,

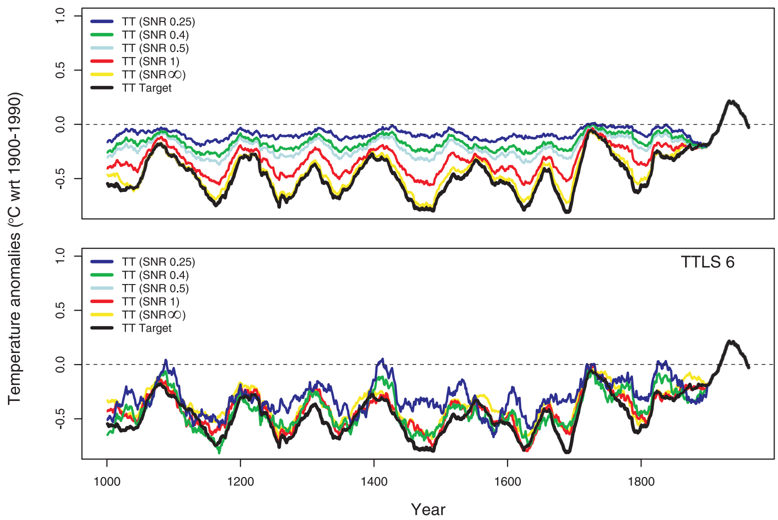

Figure 2.3: *European summer average temperature anomalies (30-year running mean) wrt 1900 to 1990 AD, for PC regression (top) and RegEM (bottom), using 30 pseudoproxies (see Fig. 2.1) with varying white noise added to the signal. The target (black line) is compared to the reconstruction results (colored lines). TTLS indicates which truncation parameter is used to reconstruct.*

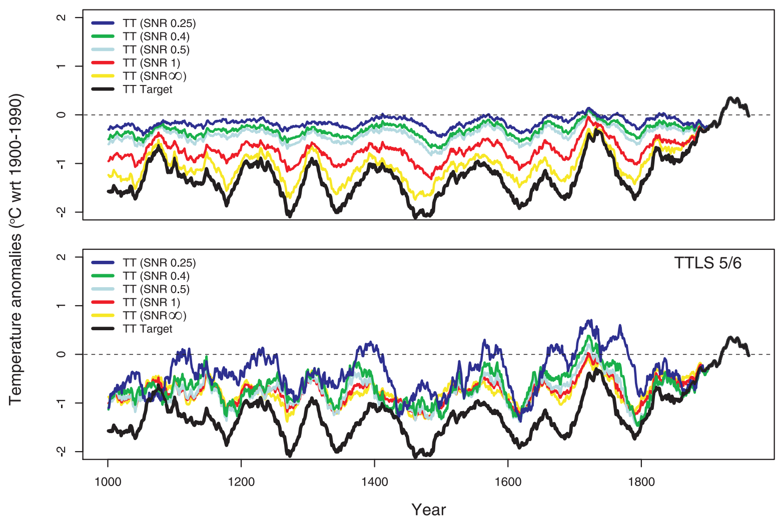

Figure 2.4: *as Figure 2.3, but for winter.*

Table 2.1: *RE as well as CE skill scores for the NCAR CSM 1.4 results shown in Figure 2.3, 2.4, 2.5 and 2.6, and the non-filtered reconstruction results (not shown). The calibration period is from 1900 to 1990 AD, the verification period from 1001 to 1899 AD. For RegEM RE and CE are shown for two different TTLS parameters (left, TTLS parameters chosen as in Mann et al. (2007), right, as additionally proposed in this paper).*

NCAR CSM 1.4
Reduction of error (RE)

	PC reg				**RegEM**							
	30 year filtered		**non-filtered**		**30 year filtered**				**non-filtered**			
	su	**wi**	**su**	**wi**	**su**		**wi**		**su**		**wi**	
perfect:	0.99	0.97	0.96	0.95	0.96	0.98	0.874	0.91	0.84	0.91	0.77	0.88
SNR 1:	0.92	0.86	0.81	0.73	0.98	0.96	0.873	0.92	0.65	0.78	0.73	0.75
SNR 0.5:	0.74	0.66	0.56	0.42	0.97	0.95	0.86	0.84	0.58	0.51	0.4	0.32
SNR 0.4:	0.65	0.59	0.45	0.32	0.96	0.91	0.8	0.77	0.34	0.31	0.14	0.02
SNR 0.25:	0.45	0.45	0.22	0.14	0.839	0.82	0.6	0.54	-0.12	-0.24	-0.58	-0.56
SNR 1,phi=0.32:	0.86	0.74	0.74	0.6	0.93	0.9	0.86	0.95	0.65	0.63	0.62	0.76
SNR 1,phi=0.71:	0.8	0.68	0.67	0.49	0.843	0.91	0.75	0.9	0.66	0.71	0.39	0.6

Coefficient of efficiency (CE)

	PC reg				**RegEM**							
	30 year filtered		**non-filtered**		**30 year filtered**				**non-filtered**			
	su	**wi**	**su**	**wi**	**su**		**wi**		**su**		**wi**	
perfect:	0.98	0.92	0.85	0.86	0.86	0.94	0.633	0.74	0.45	0.68	0.33	0.66
SNR 1:	0.73	0.59	0.36	0.2	0.92	0.85	0.630	0.77	-0.19	0.24	0.22	0.28
SNR 0.5:	0.13	0.003	-0.47	-0.7	0.88	0.81	0.60	0.53	-0.43	-0.66	-0.76	-0.98
SNR 0.4:	-0.19	-0.21	-0.86	-0.996	0.59	0.7	0.43	0.34	-1.23	-1.35	-1.53	-1.86
SNR 0.25:	-0.85	-0.62	-1.63	-1.52	0.45	0.4	-0.16	-0.35	-2.8	-3.19	-3.62	-3.55
SNR 1,phi=0.32:	0.54	0.24	0.13	-0.18	0.77	0.67	0.6	0.85	-0.18	-0.24	-0.11	0.31
SNR 1,phi=0.71:	0.31	0.05	-0.13	-0.5	0.47	0.70	0.263	0.71	-0.14	0.02	-0.77	-0.16

RegEM captures the target summer average temperature (Fig. 2.3, suppl. Fig. 2.3, bottom) more adequately for all white noise levels. After focusing on the performance of the techniques for summer reconstructions, we now turn to the reconstruction results for European winter average temperatures (Fig. 2.4, suppl. Fig. 2.4). Figure 2.4 shows that both techniques capture the target average temperature less accurately for winter than for summer (Fig. 2.3, suppl. Fig. 2.3), a finding which is more pronounced for NCAR CSM 1.4 than for ECHO-G 4. In principle, we obtain the same picture for PC regression as described above for the European summer average temperature reconstruction results. However, the RE and CE skill scores are higher for summer than for winter (Tables 2.1 and 2.2). Overall, RegEM seems to be more robust and less sensitive to the amount of white noise added to the signal than PC regression, although, as seen for winter (Figure 2.4, and even more so suppl. Figure 2.4), it appears that RegEM can 'invent' undesirable,

Table 2.2: *As Table 2.1, but for ECHO-G 4 (results see supplementary online material).*

ECHO-G 4

Reduction of error (RE)

	PC reg				RegEM							
	30 year filtered		non-filtered		30 year filtered				non-filtered			
	su	wi	su	wi	su		wi		su		wi	
perfect:	0.99	0.996	0.96	0.93	0.947	0.98	0.986	0.99	0.84	0.92	0.75	0.78
SNR 1:	0.93	0.89	0.85	0.7	0.92	0.96	0.95	0.97	0.79	0.81	0.36	0.04
SNR 0.5:	0.79	0.66	0.65	0.35	0.89	0.87	0.82	0.85	0.63	0.46	-1.73	-1.45
SNR 0.4:	0.7	0.6	0.54	0.25	0.88	0.82	0.77	0.77	0.53	0.27	-2.72	-2.31
SNR 0.25:	0.62	0.49	0.42	0.07	0.91	0.92	0.81	0.65	0.12	-0.11	-2.3	-2.36
SNR 1,phi=0.32:	0.94	0.952	0.84	0.78	0.948	0.95	0.95	0.94	0.82	0.72	0.39	0.4
SNR 1,phi=0.71:	0.81	0.953	0.68	0.62	0.97	0.84	0.86	0.85	0.79	0.32	0.18	0.17

Coefficient of efficiency (CE)

	PC reg				RegEM							
	30 year filtered		non-filtered		30 year filtered				non-filtered			
	su	wi	su	wi	su		wi		su		wi	
perfect:	0.95	0.991	0.8	0.86	0.760	0.92	0.97	0.98	0.28	0.61	0.47	0.53
SNR 1:	0.67	0.76	0.29	0.35	0.64	0.79	0.9	0.93	0.02	0.13	-0.37	-1.04
SNR 0.5:	0.01	0.28	-0.62	-0.37	0.51	0.39	0.62	0.68	-0.7	-1.48	-4.81	-4.22
SNR 0.4:	-0.38	0.15	-1.11	-0.59	0.45	0.18	0.51	0.52	-1.15	-2.37	-6.91	-6.05
SNR 0.25:	-0.75	-0.09	-1.67	-0.97	0.6	0.62	0.59	0.27	-3.07	-4.11	-6.01	-6.15
SNR 1,phi=0.32:	0.72	0.898	0.25	0.53	0.762	0.78	0.88	0.87	0.15	-0.3	-0.29	-0.27
SNR 1,phi=0.71:	0.12	0.899	-0.47	0.2	0.87	0.26	0.71	0.69	0.01	-2.14	-0.74	-0.77

temporal features, such as various spurious quasi-periodic variations, which do not exist in the target data. Nevertheless, the range of the variability of the 30-year filtered results corresponds better to that of the target for RegEM (RE and CE 30-year filtered in Tables 2.1 and 2.2). While both techniques reconstruct the target average temperature less accurately with increasing noise level (Tables 2.1 and 2.2), RegEM does so to a considerably lesser degree than PC regression. Figures 2.5 and 2.6 (suppl. Figs. 2.5 and 2.6) show a second comparison of reconstruction results of the summer and winter average temperature anomalies with regard to the 1900 to 1990 AD calibration period, now with red noise applied in comparison to the corresponding white noise scenario. The middle white noise scenario SNR 1 is displayed together with the two red noise scenarios with the same SNR, but different sample lag-1 autocorrelation coefficients $\rho = 0.32$, 0.71 (as mentioned above, chosen according to *Mann et al.* (2007)). For PC regression (Figs. 5 and 6, suppl. Figs. 2.5 and 2.6, top) the addition of red noise (orange and magenta line) affects the skill of the reconstruction slightly more than the addition of white noise with SNR 1 (red line), both for summer and winter according to RE

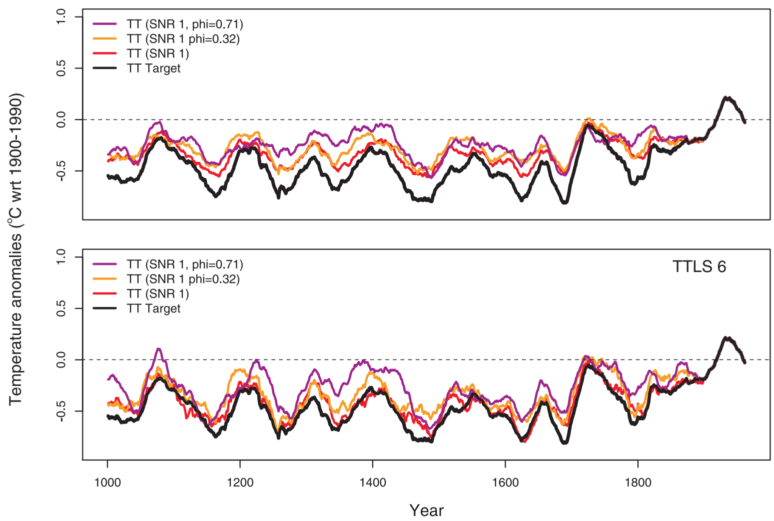

Figure 2.5: *European summer average temperatures anomalies (30-year running mean) for PC regression (top) and RegEM (bottom). The white noise scenario SNR 1 (red line) is compared with two different red noise scenarios (orange and magenta lines); the target is shown in black. TTLS indicates which truncation parameter is used.*

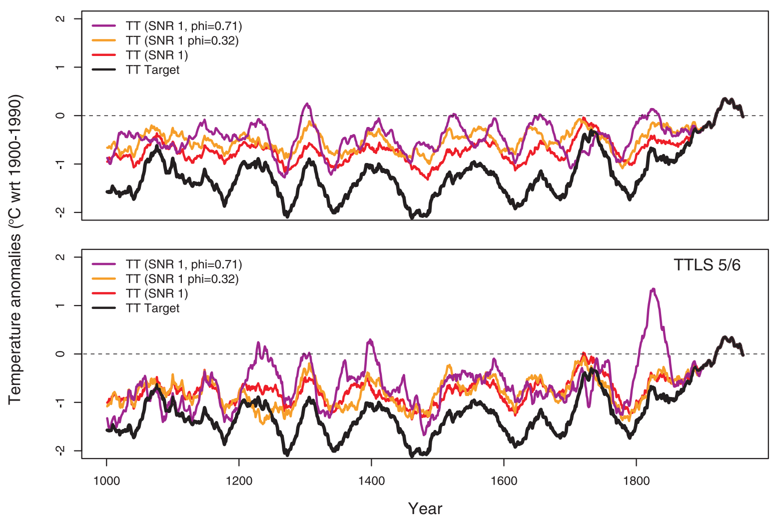

Figure 2.6: *As Figure 2.5, but for winter.*

and CE (Tables 2.1 and 2.2). For RegEM, the target temperature variations also remain appropriately reconstructed for summer when red noise is added instead of white noise according to RE and CE (Tables 2.1 and 2.2), although adding red noise with an autocorrelation coefficient $\rho = 0.71$ (magenta line) clearly increases the variability of the reconstruction result in winter.

The RE and CE scores for the 30-year filtered data (Tables 2.1 and 2.2) quantitatively describe the reconstruction results (Figs. 2.3, 2.4, 2.5 and 2.6, likewise for ECHO-G 4 in the supplementary online material) and confirm that RegEM performs better than PC regression focusing on the evaluation of the low frequency variations (RE and CE 30-year filtered in Table 2.1 and 2.2). Nevertheless, a glance at the RE and CE scores calculated for non-filtered results (figures not shown) reveals differences in the performance of the reconstructions seen in Figures 2.3, 2.4, 2.5 and 2.6. Summer average temperature reconstructions using RegEM also produce lower RE scores than those using PC regression under the different white and red noise scenarios (RE non-filtered in Tables 2.1 and 2.2). Winter temperature reconstructions based on RegEM and PC regression RE and CE scores are comparable (Table 2.1), and slightly lower in a few cases for RegEM (Table 2.2). The SNR 0.25 scenario, in particular, leads to lower skill score values, and the result is generally unsatisfactory. Using rednoise scenarios (Fig. 2.6, suppl. Fig. 2.6, bottom), the range of the variability of the SNR 1 scenario with an autocorrelation coefficient of $\rho = 0.71$ (magenta line) is rather somewhat too large compared to the target (Fig. 2.5, suppl. Fig. 2.5) for RegEM. Finally, several scenarios for both summer and winter even return negative annual RE and CE scores with RegEM, indicating that these reconstruction results have no skill. With the alternative way of determining the TTLS parameters for RegEM (supplementary online material), equally skilful, and in some cases even more skilful reconstructions can be achieved. For PC regression, RE scores indicate that all reconstruction results have skill; however, this is contradicted (for SNR 0.5, SNR 0.4 and SNR 0.25) by the corresponding CE scores (Figs. 2.3 and 2.4, suppl. Figs. 2.3 and 2.4).

To summarize: Figures 2.3 and 2.4 (Suppl. Figs. 2.3 and 2.4) as well as Tables 2.1 and 2.2 indicate that both techniques reconstruct European temperature variability more adequately for summer than for winter. RegEM seems to be more robust than PC regression with regard to the effect of noise added to the signal. Figures 2.5 and 2.6 (Suppl. Figs. 2.5 and 2.6), as well as Tables 2.1 and 2.2 display that reconstructions using red noise instead of white noise still retain skill. Nevertheless, the increase in variability in the results affects the reconstruction skill, more so in winter than in summer. Finally, there is a difference in reconstruction skill depending on variability frequency.

2.3.2 The spatial skill patterns of the reconstructions

Figures 2.7 and 2.8 (Suppl. Figs. 2.7 and 2.8) show the spatial skill patterns of the summer and winter reconstruction results from Figures 2.3 and 2.4 (Suppl. Figs. 2.3 and 2.4) under the three different white noise scenarios, i.e. SNR ∞, SNR 1 and SNR 0.5. Since examining RE, the relation between the squared reconstruction error and the squared anomalies from the calibration average, is somewhat controversial (*Bürger and Cubasch*, 2007), we have chosen to add a more intuitive skill measure, and also to look at the spatial differences of the two techniques, thus making it possible to directly determine the origins of the underestimation of the target temperature variations in the reconstruction results. Accordingly, the spatial skill is defined here as the differences between reconstructed and target temperature anomalies, i.e. the bias, averaged over the verification period, 1001 to 1899 AD, and the RE skill scores for the 30-year filtered results calculated for each gridpoint. This corresponds to a validation of the whole summer and winter temperature field. Positive bias values indicate that the difference between the average of reconstructed temperature anomalies over the verification period and the calibration period mean is smaller than that between target and calibration mean. Thus the target temperature anomalies are underestimated by the reconstructed anomalies, and overestimated for negative bias values. A lack of predictor model gridpoints (see Fig. 2.1) in the Atlantic leads to considerable uncertainties over that area both for summer and winter reconstructions (Figs. 2.7 and 2.8, suppl. Figs. 2.7 and 2.8). This effect is to be expected. However, the smaller the SNR, the larger the area with underestimation of target temperature anomalies becomes for summer and winter. Again, this is less pronounced for RegEM than for PC regression. Thus RegEM seems to be less dependent on the SNR than PC regression. The spatial skill patterns of RegEM are quite similar to those of PC regression. Nevertheless, for PC regression the underestimation of the target temperature variations during the verification period in the field is more clearly indicated. The spatial validation of the two techniques discloses the underestimation of amplitude seen for the European average temperatures in Figures 2.3 and 2.4 (Suppl. Figs. 2.3 and 2.4) for PC regression. Focusing on the spatial RE scores for the 30-year filtered reconstruction results, we conclude that no large differences can be seen, despite the fact that RE skill scores are again higher for summer results than for winter.

2.4 Discussion

The results presented in this comparison of PC regression and RegEM reveal a seasonal dependence of reconstruction skill. Both techniques seem to perform more accurately (Figs. 2.3 and 2.5 compared to Figs. 2.4 and 2.6, likewise for ECHO-G 4 in the supplementary online material) for reconstructing summer average temperatures than winter when focusing on how well the low frequency variability of the target is captured. Testing the techniques with a less dense predictor network (12 gridpoints representing real world proxy series used to reconstruct the late Maunder Minimum (*Küttel et al.*, 2007), not shown) confirms these findings, although with an additional decrease in reconstruction skill. The more skilful performance in reconstructing European summer temperatures over the last millennium might be explained by the fact that the range of temperature variability is smaller in summer than in winter. Consequently, the impact of adding noise to the signal with smaller standard deviations in summer than in winter is less remarkable. Thus the reconstruction skill is less affected for summer than for winter. Furthermore, this is also potentially related to the spatial distribution of the predictor network used here. Predictor networks which may be optimal for reconstructing summer temperatures are not necessarily optimal for reconstructing winter temperatures (*Pauling et al.*, 2003; *Luterbacher et al.*, 2006; *Küttel et al.*, 2007).

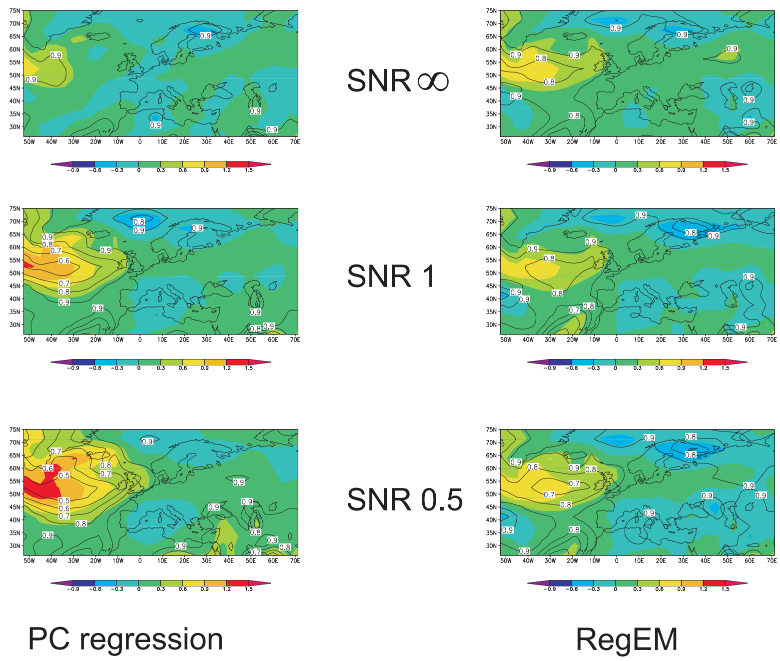

Figure 2.7: *Spatial skill patterns of the European summer temperature reconstructions using PC regression (left) and RegEM (right) with white noise scenarios SNR ∞, SNR 1, and SNR 0.5. The skill is defined by the average of the bias (reconstructed values - target values) [shaded] and RE [contours] calculated for each gridpoint over the verification period from 1001 to 1899 AD. The scale refers to the bias, i.e. differences in temperature anomalies for summer. Colors indicate reconstructed values that are about (greenish blue and green), higher (light green, yellow to red) or lower (light blue to violet) than the target values.*

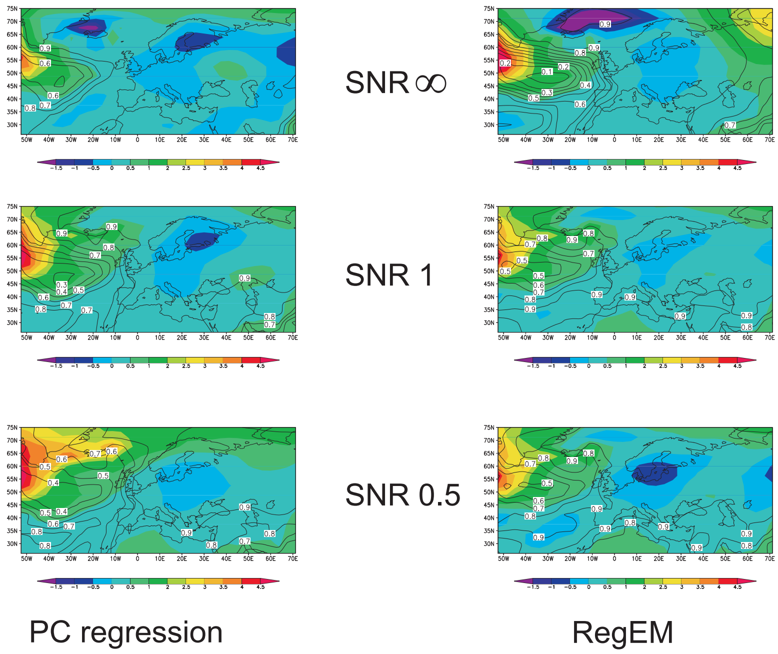

Figure 2.8: *As Figure 2.7, but for winter. Colors indicate reconstructed values that are about (light blue and greenish blue), higher (light green to green, yellow, red) or lower (dark blue to violet) than the target values.*

The performance of reconstructions seems to depend less on the red structure of noise for the SNR 1 scenario with an autocorrelation coefficient $\rho = 0.32$ than with $\rho = 0.71$ (RE and CE 30-year filtered in Tables 2.1 and 2.2). For $\rho = 0.71$ the variability of the reconstruction results, especially for winter, is considerably increased using RegEM (Figure 2.6, suppl. Figure 2.6). Furthermore the skill of the reconstruction is generally more affected for the SNR 1 scenario with $\rho = 0.71$ than with $\rho = 0.32$ or for white noise only. However, analyses of typical red noise characteristics of realworld data (in *Luterbacher et al.*, 2004) reveal, that $\rho = 0.71$ is not seen in the data and $\rho = 0.32$ is presumably more indicative of real world proxies. Still it is useful to study a range of autocorrelation coefficients to obtain an understanding of how reconstruction results depend on different types of noises. Nevertheless, the noise scenarios in this paper certainly do not mimic the full range of characteristics of noisy real world predictor series, once again indicating that there is a need to model predictor data and inherent uncertainties more realistically (*Moberg et al.*, 2007). Tables 2.1 and 2.2 indicate that both techniques lose skill to an increasing degree as more noise is added to the signal. RegEM is less sensitive to and less affected by the noise addition than PC regression, but applying RegEM instead of PC regression in reconstructing, one of the fundamental statistical problems remains. Furthermore, while for the 30-year filtered data (Tables 2.1 and 2.2) RE and CE skill scores confirm that RegEM performs more accurately than PC regression, the skill scores for the non-filtered reconstruction results are nevertheless lower for RegEM than for PC regression, especially in winter. One explanation might be, that using RegEM, mean and covariance of the whole input data matrix are iteratively estimated. The fact that the statistical characteristics of the whole input matrix are addressed together over the calibration and verification periods might be a reason for the less accurate inter-seasonal performance of RegEM, as exhibited by the validation of the non-filtered results. Furthermore, considering the reconstruction results in Figures 2.3 and 2.4 (top, likewise ECHO-G 4 in the supplementary online material) and the RE scores in Tables 2.1 and 2.2, it is an alarming sign that the PC regression results still achieve such high RE scores; moreover, the RE scores are put into the right perspective by the negative CE scores. The implication for reconstructions with real world proxy data is that verification has to be conducted very carefully by applying different means of validation. The interpretation of reconstruction skill and the reasonable verification of reconstruction results are delicate and not free from contradictions. Therefore, the development of alternative and more intuitive tools, as well as more thorough validation must be attempted (e.g. *Wahl and Ammann*, 2007).

Why should it be the case that RegEM captures the target average 30-year filtered temperature variations more adequately than PC regression? When applying PC regression, we use OLS to estimate the regression coefficients for the calibration period. By contrast, when applying RegEM we use either the conditional Maximum likelihood method (if no regularization is needed) or TTLS (if the problem is ill-posed). These different estimation techniques, especially TTLS, which takes into account errors in the explanatory variables (Equation 2.1, $\mathbf{e}_{proxy}$), have a crucial impact on the reconstruction skill. Another important difference is the nature of RegEM as an iterative process which is non-linear in general. Finally, RegEM not only provides estimates of the mean with each iteration step, but of the variance as well. We expected RegEM to be better than PC regression prior to this study, but we also expected it to be better than our results now indicate. One expectation for the less pronounced difference is that the reconstruction performance depends not only on the statistical technique chosen, but also on the choice and quality of the predictor network. Therefore, these other factors should be optimized, as well.

However, the use of RegEM also leaves room for future methodological improvements. *Mann et al.* (2007) recently addressed the problem of choosing truncation parameters. This was also investigated prior to applying RegEM here. The validation of a range of parameters, close to the one proposed by *Mann et al.* (2007), demonstrated that comparable results can be obtained by using alternative parameters (supplementary online material). We therefore urge the evaluation of several truncation parameters over the verification period.

Despite all this, we prefer RegEM to PC regression in this case, as it captures the multi-decadal variations of the target summer and winter European average temperatures more accurately (Figs. 2.3, 2.4, 2.5 and 2.6, likewise for

ECHO-G 4 in the supplementary online material) than PC regression when focusing on lower frequency variability (RE and CE 30-year filtered in Tables 2.1 and 2.2).

2.5 Conclusions and perspectives

The outcomes regarding the performance of the two reconstruction techniques are restricted to the specific experimental setting used in this paper. As mentioned above the tests are based on NCAR CSM 1.4 and ECHO-G 4 climate model data, a predictand which consists of 476 gridpoints (land and sea), a pseudoproxy network with 30 gridpoints (Fig. 2.1), and scenarios based on different SNR constant over time. By comparing the two CFR techniques, -PC regression and RegEM,- at a continental and seasonal scale, we have demonstrated that the reconstruction skill differs according to the spatial and temporal scales the techniques are applied to. The fact that RegEM achieves different results for continental and hemispheric reconstructions (*Mann et al.*, 2007) emphasizes the necessity of downscaling to smaller spatial and subannual temporal scales, in order to achieve a better understanding of the robustness and skill of the reconstruction techniques on higher temporal and spatial scales. Furthermore, hemispheric annual temperature reconstructions do not provide information about regional-scale variations, such as the intrinsic seasonal patterns of climate change as they have occurred, for instance, in Europe during past centuries (*Mann et al.*, 2000; *Luterbacher et al.*, 2004, 2007; *Xoplaki et al.*, 2005). We found seasonal differences in the performance of RegEM and PC regression, and we demonstrated that predictor data quality has a crucial impact on reconstruction skill. RegEM has proved that more adequate results can be obtained by better incorporating the errors in the predictor data to reconstruct surface air temperature fields. However, the choice of the right TTLS parameters turned out to be ambiguous, and the procedure for selecting the most accurate ones needs further investigation. If no noise, or noise with a high SNR, is added to the signal, PC regression performs just as well as RegEM for winter and for summer. If noise with a smaller SNR is added to the climatic signal, the performance of RegEM proves to be more robust compared to PC regression.

If the variability range is too large, as is the case e.g. for SNR 0.25 and SNR 1 with $\rho = 0.71$, both RegEM and PC regression exhibit deficits: the amplitude of target temperature variations tends to be underestimated by PC regression and overestimated by RegEM. However, overestimation might be adjusted by the choice of more suitable TTLS parameters.
The next step will be to quantify the differences between PC regression and RegEM by applying the two techniques to real world data, given a varying number of predictors and SNR over time. There is still a need and potential for further optimizations of CFR techniques, such as RegEM, that take better account of errors. PC regression can still be optimized as well, e.g. by restriction to land areas only (*Luterbacher et al.*, 2004; *Xoplaki et al.*, 2005), optimization of PC truncation, or the implementation of different regression coefficient estimation procedures. Certainly other settings, and more realistic real world conditions have to be considered in future. On the one hand CFR techniques need to be better adapted to the specific character of the predictor data, and on the other, the quality of the predictor data has to be better understood, quantified and modeled. Exclusive use of classical multivariate statistics should be expanded to include solutions already developed in other research areas, e.g. econometrics. Time series analysis offers still further solutions, such as state space models and the use of Kalman filters (*Lee et al.*, 2008), that are also worth exploring with regard to climate field reconstructions.

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. Publication cost contributions are kindly provided by the Foundation Marchese Francesco Medici del Vascello. We thank Caspar Ammann and Fortunat Joos for making available the NCAR CSM 1.4 results, as well as Fidel Gonzalez Rouco and Eduardo Zorita for providing the ECHO-G 4 simulation results. We also thank Scott Rutherford, Eugene Wahl and Michael E. Mann for the RegEM code and for their helpful comments and inputs. Finally, we wish to thank the reviewers for their constructive criticism and suggestions, which helped to improve the quality of this study.

2.6 Supplementary Material

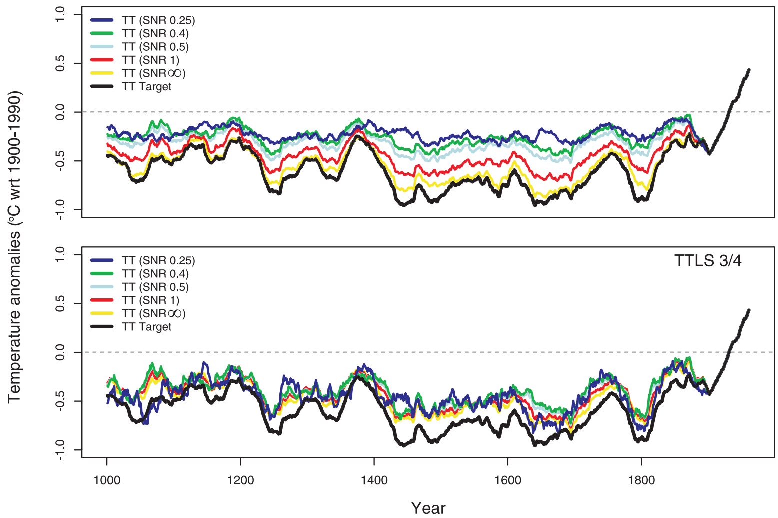

Figure 2.9: *Supplementary Figure 2.3 for ECHO-G 4. European summer average temperature anomalies (30-year running mean) wrt 1900 to 1990 AD, for PC regression (top) and RegEM (bottom), using 30 pseudoproxies (see Fig. 2.1) with varying white noise added to the signal. The target (black line) is compared to the reconstruction results (colored lines).*

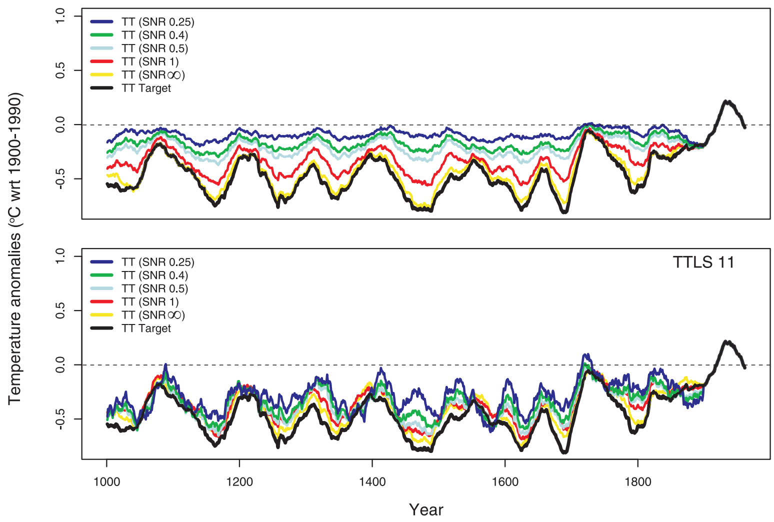

Figure 2.10: *Supplementary Figure 2.3 for NCAR CSM 1.4. As Figure 2.3, but with different TTLS parameter.*

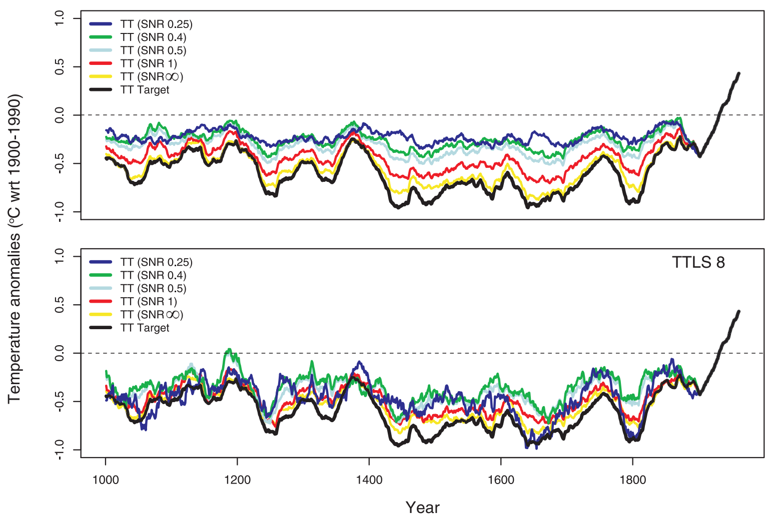

Figure 2.11: *As supplementary Figure 2.3 for ECHO-G 4, but with different TTLS parameter.*

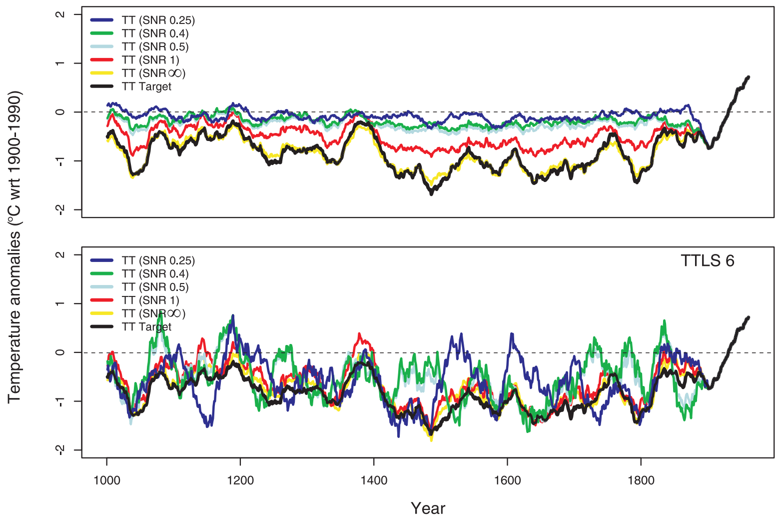

Figure 2.12: *Supplementary Figure 2.4 for ECHO-G 4. As supplementary Figure 2.3, but for winter.*

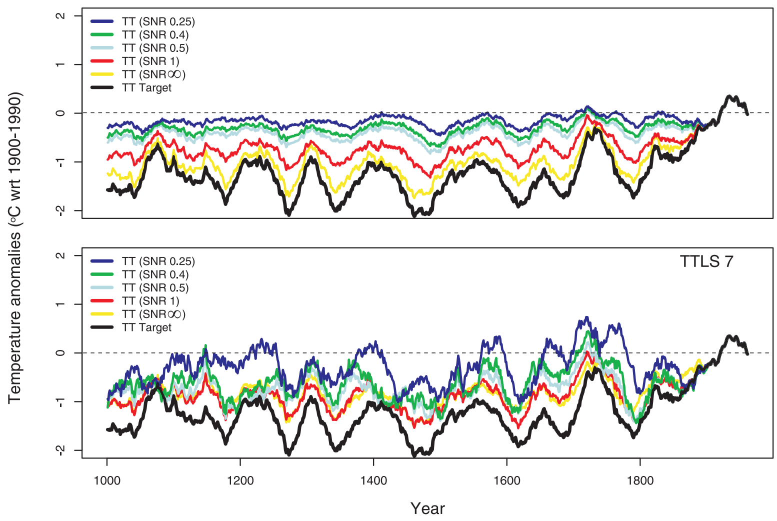

Figure 2.13: *Supplementary Figure 2.4 for NCAR CSM 1.4. As Figure 2.4, but with different TTLS parameter.*

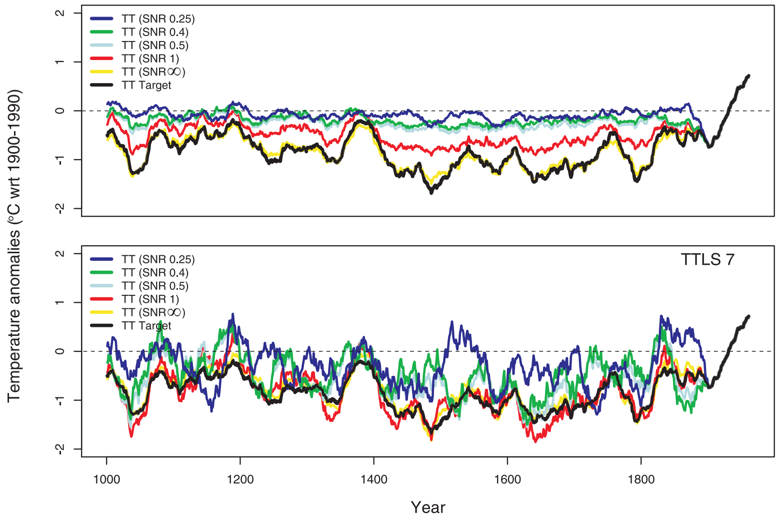

Figure 2.14: *As supplementary Figure 2.4 for ECHO-G 4, but with different TTLS parameter.*

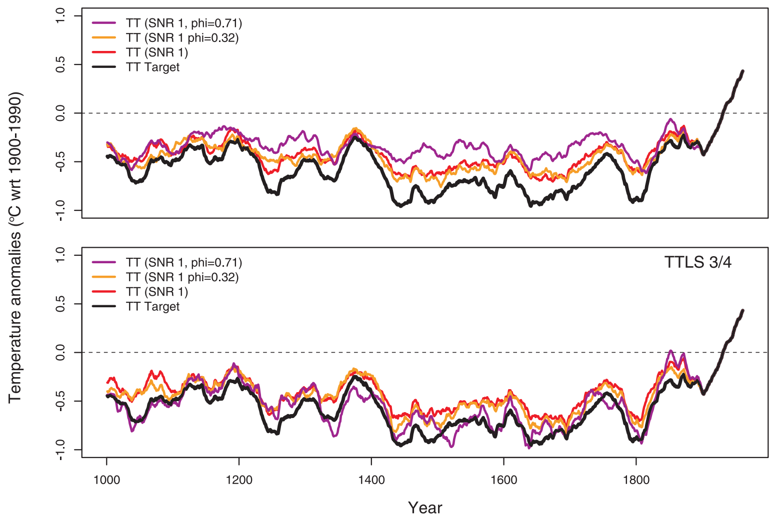

Figure 2.15: *Supplementary Figure 2.5 for ECHO-G 4. European summer average temperatures anomalies (30-year running mean) for PC regression (top) and RegEM (bottom). The white noise scenario SNR 1 (red line) is compared with two different red noise scenarios (orange and magenta lines); the target is shown in black.*

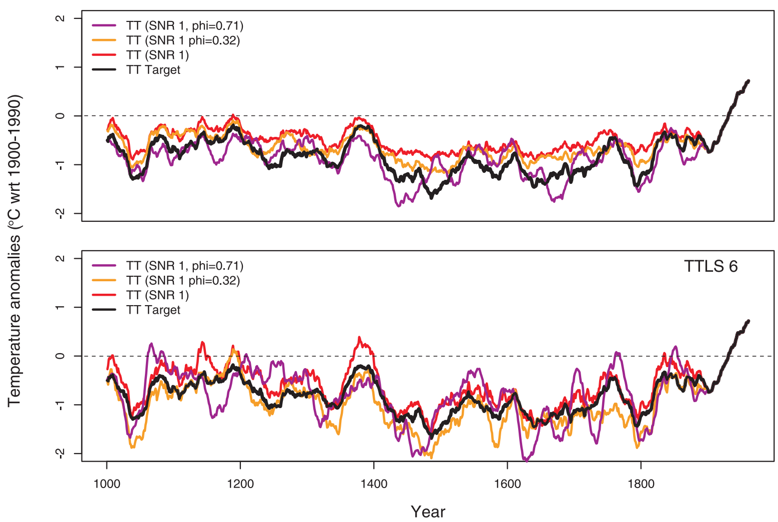

Figure 2.16: *Supplementary Figure 2.6 for ECHO-G 4. As supplementary Figure 2.5, but for winter.*

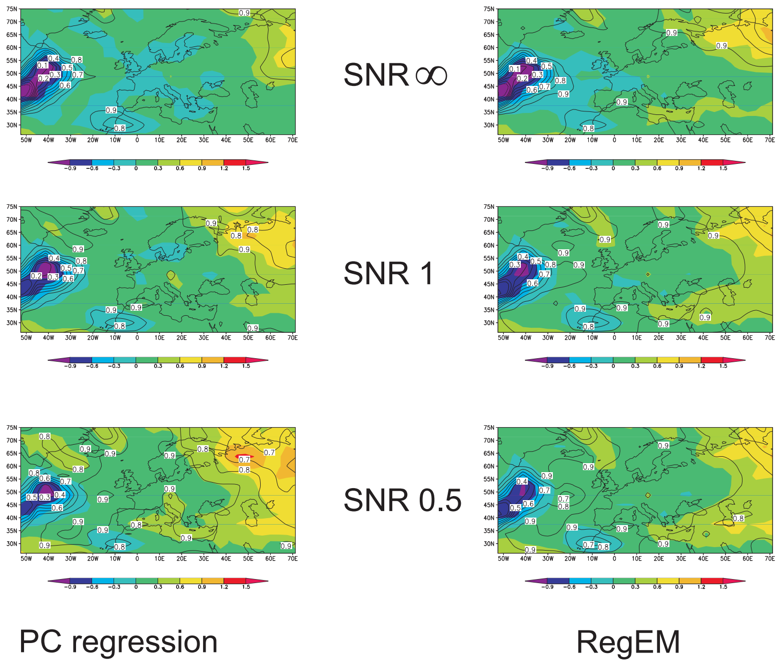

Figure 2.17: *Supplementary Figure 2.7 for ECHO-G 4. Spatial skill patterns of the European summer temperature reconstructions using PC regression (left) and RegEM (right) with white noise scenarios SNR ∞, SNR 1, and SNR 0.5. The skill is defined by the average of the bias (reconstructed values - target values) [shaded] and RE [contours] calculated for each gridpoint over the verification period from 1001 to 1899 AD. The scale refers to the bias, i.e. differences in temperature anomalies for winter and summer separately. Colors indicate reconstructed values that are about (greenish blue and green), higher (light green, yellow to red) or lower (light blue to violet) than the target values.*

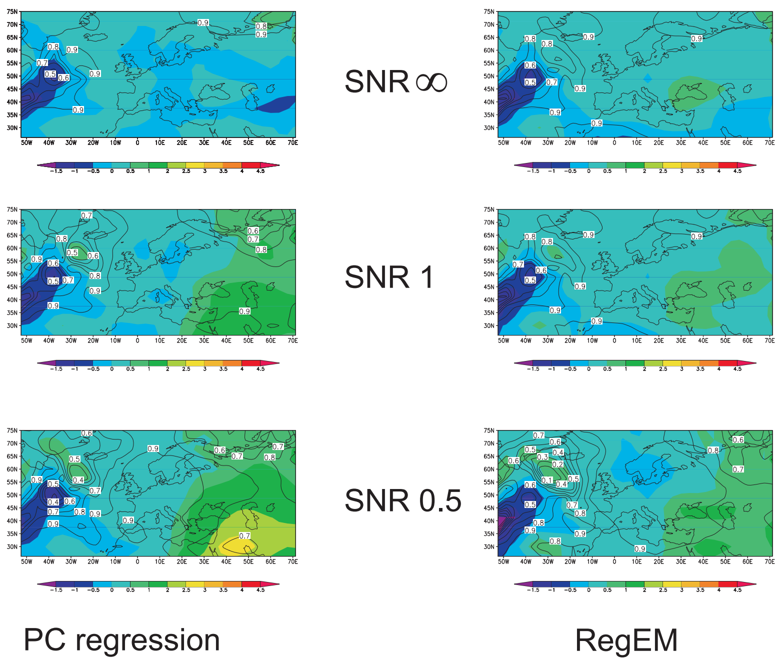

Figure 2.18: *Supplementary Figure 2.8 for ECHO-G 4. As supplementary Figure 2.7, but for winter. Colors indicate reconstructed values that are about (light blue and greenish blue), higher (light green to green, yellow, red) or lower (dark blue to violet) than the target values.*

References

Ammann, C., F. Joos, D. Schimel, B. Otto-Bliesner, and R. Tomas (2007), Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model., *Proc. Natl. Acad. Sci. USA (PNAS)*, *104*, 3713–3718.

Briffa, K., T. Wigley, P. Jones, J. Pilcher, and M. Hughes (1987), Patterns of tree-growth and related pressure variability in Europe, *Dendrochronologia*, *5*, 35–59.

Briffa, K., T. Osborn, F. Schweingruber, I. Harris, P. Jones, S. Shiyatov, and E. Vaganov (2001), Low-frequency temperature variations from a northern tree ring density network, *Journal of Geophysical Research*, *106*, 2929–2942.

Brohan, P., J. Kennedy, I. Harris, S. Tett, and P. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850, *Journal of Geophysical Research*, *111*, D12106.

Bürger, G., and U. Cubasch (2005), Are multiproxy climate reconstructions robust?, *Geophysical Research Letters*, *32*, L23711.

Bürger, G., and U. Cubasch (2007), On the verification of climate reconstructions, *Climate of the past*, *3*, 397–409.

Casty, C., D. Handorf, and M. Sempf (2005a), Combined winter climate regimes over the North Atlantic/European sector 1766–2000, *Geophysical Research Letters*, *32*, L13801.

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

Casty, C., C. Raible, T. Stocker, H. Wanner, and J. Luterbacher (2007), European climate pattern variability since 1766, *Climate Dynamics, in press.*

Cook, E., K. Briffa, and P. Jones (1994), Spatial regression methods in dendroclimatology: a review and comparison of two techniques, *International Journal of Climatology*, *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., R. Wilson, D. Frank, A. Moberg, H. Wanner, and J. Luterbacher (2005), Climate: past ranges and future changes, *Quaternary Science Reviews*, *24*, 2164–2166.

Esper, J., D. Frank, U. Buentgen, A. Verstege, J. Luterbacher, and E. Xoplaki (2007), Long-term drought severity variations in Morocco, *Geophysical Research Letters*, *34*, L17702.

Fischer, E., J. Luterbacher, E. Zorita, S. Tett, C. Casty, and H. Wanner (2007), European climate response to tropical volcanic eruptions over the last half millennium, *Geophysical Research Letters*, *34*, L05707.

González-Rouco, J., H. Beltrami, E. Zorita, and H. von Storch (2006), Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling, *Geophysical Research Letters*, *33*, L01703.

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

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

Jansen, E., and coauthors (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*, Cambridge University Press, Cambridge, United Kingdom and New York.

Jones, P., and M. Mann (2004), Climate over past millennia, *Reviews of Geophysics*, *42*, RG2002.

Klimenko, V., V. Klimanov, A. Sirin, and A. Sleptsov (2001), Climate changes in Western European Russia in the late Holocene, *Doklady Earth Sciences*, *377*, 190–194.

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, *Geophysical Research Letters*, *34*, L07710.

Lee, T., F. Zwiers, and M. Tsao (2008), Evaluation of proxy-based millennial reconstruction methods, *Climate Dynamics, in press*.

Li, B., D. Nychka, and C. Ammann (2007), The 'hockey stick' and the 1990s: a statistical perspective on reconstructing hemispheric temperatures, *Tellus A (Online Early Articles)*, *doi:10.1111/j.1600-0870.2007.00270.x*.

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.

Luterbacher, J., Eds. P. Lionello, P. Malanotte-Rizzoli, and R. Boscolo (2006), Mediterranean climate variability over the last centuries: A review in: *The Mediterranean Climate: an overview of the main characteristics and issues*, Elsevier, Amsterdam, the Netherlands, 27–148.

Luterbacher, L., M. Liniger, A. Menzel, N. Estrella, P. Della-Marta, C. Pfister, T. Rutishauser, and E. Xoplaki (2007), The exceptional European warmth of Autumn 2006 and Winter 2007: Historical context, the underlying dynamics and its phenological impacts, *Geophysical Research Letters*, *34*, L12704.

Mangini, A., C. Spötl, and P. Verdes (2005), Reconstruction of temperature in the Central Alps during the past 2000 yr from a δ18O stalagmite record, *Earth and Planetary Science Letters*, *235*, 741–751.

Mann, M., and S. Rutherford (2002), Climate reconstruction using'Pseudoproxies', *Geophysical Research Letters*, *29*, 1501.

Mann, M., R. Bradley, and M. Hughes (1998), Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, *392*, 779–787.

Mann, M., R. Bradley, and M. Hughes (1999), Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophysical Research Letters*, *26*, 759–762.

Mann, M., E. Gille, R. Bradley, M. Hughes, J. Overpeck, F. Keimig, and W. Gross (2000), Global Temperature Patterns in Past Centuries: An Interactive Presentation, *Earth Interactions*, *4* (1234), 1–29.

Mann, M., S. Rutherford, E. Wahl, and C. Ammann (2005), Testing the Fidelity of Methods Used in Proxy-Based Reconstructions of Past Climate, *Journal of Climate*, *18*, 4097–4107.

Mann, M., S. Rutherford, E. Wahl, and C. Ammann (2007), Robustness of proxy-based climate field reconstruction methods, *Journal of Geophysical Research*, *112*, D12109.

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 (2007), Analysis of the Moberg et al. (2005) hemispheric temperature reconstruction , *Climate Dynamics, in press.*

Pauling, A., J. Luterbacher, and H. Wanner (2003), Evaluation of proxies for European and North Atlantic temperature field reconstructions, *Geophysical Research Letters*, *30*, 1787.

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, *Climate Dynamics*, *26*(4), 387–405.

Proctor, C., A. Baker, and W. Barnes (2002), A three thousand year record of North Atlantic climate, *Climate Dynamics*, *19*, 449–454.

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, *Journal of Climate*, *18*, 2308–2329.

Schneider, T. (2001), Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values., *Journal of Climate*, *14*, 853–871.

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

Shindell, D., G. Schmidt, M. Mann, D. Rind, and A. Waple (2001), Solar Forcing of Regional Climate Change During the Maunder Minimum.

Shindell, D., G. Schmidt, R. Miller, and M. Mann (2003), Volcanic and Solar Forcing of Climate Change during the Preindustrial Era, *Journal of Climate*, *16*, 4094–4107.

Shindell, D., G. Schmidt, M. Mann, and G. Faluvegi (2004), Dynamic winter climate response to large tropical volcanic eruptions since 1600, *Journal of Geophysical Research*, *109*, D05104.

Smerdon, J., and A. Kaplan (2007), Comments on Testing the Fidelity of Methods Used in Proxy-Based Reconstructions of Past Climate: The Role of the Standardization Interval, *Journal of Climate*, *20*(22), 5666–5670.

Thejll, P., and T. Schmith (2005), Limitations on regression analysis due to serially correlated residuals: Application to climate reconstruction from proxies, *Journal of Geophysical Research*, *110*, D18103.

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, J. Jones, F. Gonzalez-Rouco, and S. Tett (2006), Response to Comment on Reconstructing Past Climate from Noisy Data.

von Storch, H., E. Zorita, and F. Gonzalez-Rouco (2007), Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation, *Journal of Earth Sciences, in press.*

Wahl, E., and C. Ammann (2007), Robustness of the Mann, Bradley, Hughes reconstruction of the Northern hemisphere surface temperatures: Examination of criticisms based on the nature and processing of proxy climate evidence, *Climatic Change (in press).*

Waple, A., M. Mann, and R. Bradley (2002), Long-term patterns of solar irradiance forcing in model experiments and proxy based surface temperature reconstructions, *Climate Dynamics*, *18*, 563–578.

Wilks, D. (1995), *Statistical methods in the atmospheric sciences*, Academic Press San Diego.

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, *Geophysical Research Letters*, *32*, L15711.

Chapter 3

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

Nadja Riedwyl, Jürg Luterbacher and Heinz Wanner

Oeschger Centre for Climate Change Research (OCCR), Bern, Switzerland
Institute of Geography, Climatology and Meteorology, University of Bern, Bern, Switzerland

Geophysical Research Letters, submitted

Abstract

An ensemble of reconstruction results for past European temperature variability back to 1500 is presented. We apply principal component (PC) regression, regularized expectation maximization (RegEM) and composite-plus-scaling (CPS). The reconstruction results of the three techniques for summer and winter European temperature averages, and spatial fields related to warmest and coldest decades are analyzed and discussed. We show that PC regression and RegEM perform more similarly compared to CPS, and that

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 also very likely to be due to limited spatial and temporal availability of the proxy data. The comparison of PC regression, RegEM and CPS reveals that past temperature variability is likely more variable than indicated by earlier European seasonal temperature reconstructions, still indicating 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.

3.1 Introduction

Reconstruction of past temperature variability is of high importance in the discussion on current climate change (*Jansen and coauthors*, 2007). The question, if caveats of reconstruction techniques can lead to false 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, 2007; *Rutherford et al.*, 2005; *Küttel et al.*, 2007; *Lee et al.*, 2008; *Mann et al.*, 2007; *Moberg et al.*, 2007; *Riedwyl et al.*, 2008a) 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 to further examine lessons learnt from "a surrogate climate", and to better take into account the impact of real world conditions, i.e. the much higher resolved real gridded target surface air temperature field in contrast to the lower resolved hemispheric climate model fields. Furthermore, the heterogeneity and limited spatial and temporal availability of the real proxy data, which is idealized often using pseudoproxies (*Riedwyl et al.*, 2008a).

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), composite-plus-scaling (CPS) (e.g. *Jones and Mann*, 2004; *Esper et al.*, 2005), and regularized expectation maximization (RegEM) (*Schneider*, 2001; *Rutherford et al.*, 2005; *Mann et al.*, 2007). Thus, 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 reconstruction results leads to a more in-depth understanding of the reliability and robustness of existing European temperature reconstructions back to 1500.

3.2 Data and Methods

The predictand is the European surface air temperature field taken from *Mitchell and Jones* (2005) at 0.5°x0.5° 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). 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, though with some additional information that has become available since (tree ring series and grape harvest dates). In order to use maximal predictor information, 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 series, of the available proxy data is calculated first, and then used for reconstruction. PC regression is applied as outlined in *Luterbacher et al.* (2004) and *Riedwyl et al.* (2008a). 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 detail by *Schneider* (2001); *Rutherford et al.* (2005); *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 the truncation levels tend to be overestimated by the above mentioned criteria. We therefore set the truncation levels in these cases as low as possible, i.e. to 1.

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.

We mainly focus on lower frequency variations, thus only the 30-year Gaussian filtered reconstruction results with associated uncertainties (filtered 2 standard errors SE) are shown for the European averages. 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. The SE for the 30-yr filtered reconstruction results are calculated as in *Xoplaki et al.* (2005).

Furthermore, we compare the spatial temperature anomaly averages of the warmest summer and coldest winter decades as well as extreme single summer and winter years, using PC regression and RegEM. The warmest summer and coldest winter decades are defined as the consecutive 10 warmest and coldest years within the reconstruction period from 1500 to 1900.

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 reconstructions.

3.3 Results and Discussion

3.3.1 Analysis of the temperature average reconstructions

Figure 3.1 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 at the end of the 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.*, 2008a), RegEM with TTLS is less affected, and CPS, by definition, fully retains the variability of the calibration period. The 2 SE bounds of the PC regression and CPS results do not fully overlap. However, the PC regression and CPS results lie within the filtered 2 SE bounds of the RegEM result. Furthermore, the RegEM result is not fully included in the 2 SE bounds of PC regression and CPS at the end of the 18th century. For winter (Figure 3.1, bottom) the three results show distinctly positive temperature anomalies for the mid 18th century and negative anomalies at the end of the 17th century. The results of PC regression and RegEM mostly agree on the magnitude of temperature amplitudes, the ones of RegEM being slightly higher than those of PC regression. 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 uncertainty bounds of the CPS result do not overlap scarcely with the results of PC regression and RegEM. Table 3.1 presents RE, CE and RRMSE scores for the results shown in Figure 3.1. PC regression reveals highest skill both for summer and winter, followed by RegEM. CPS has lowest skill. The average temperature anomalies, both for summer and winter, are coldest for CPS (Figure 3.1) compared to PC regression and RegEM. This seems to deteriorate the reconstruction skill of CPS. 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 reconstruction period seems not to be penalized by the validation measures (*Riedwyl et al.*, 2008a). There is a decrease in skill (Table 3.1), 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. Seasonal

Table 3.1: *RE, CE and RRMSE skill scores for the results shown in Figure 3.1. 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.*

Reduction of error (RE)

	PC reg		RegEM		CPS	
	su	wi	su	wi	su	wi
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)

	PC reg		RegEM		CPS	
	su	wi	su	wi	su	wi
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)

	PC reg		RegEM		CPS	
	su	wi	su	wi	su	wi
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

differences in the performance are evident for RegEM and CPS: 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 winter predictor net-

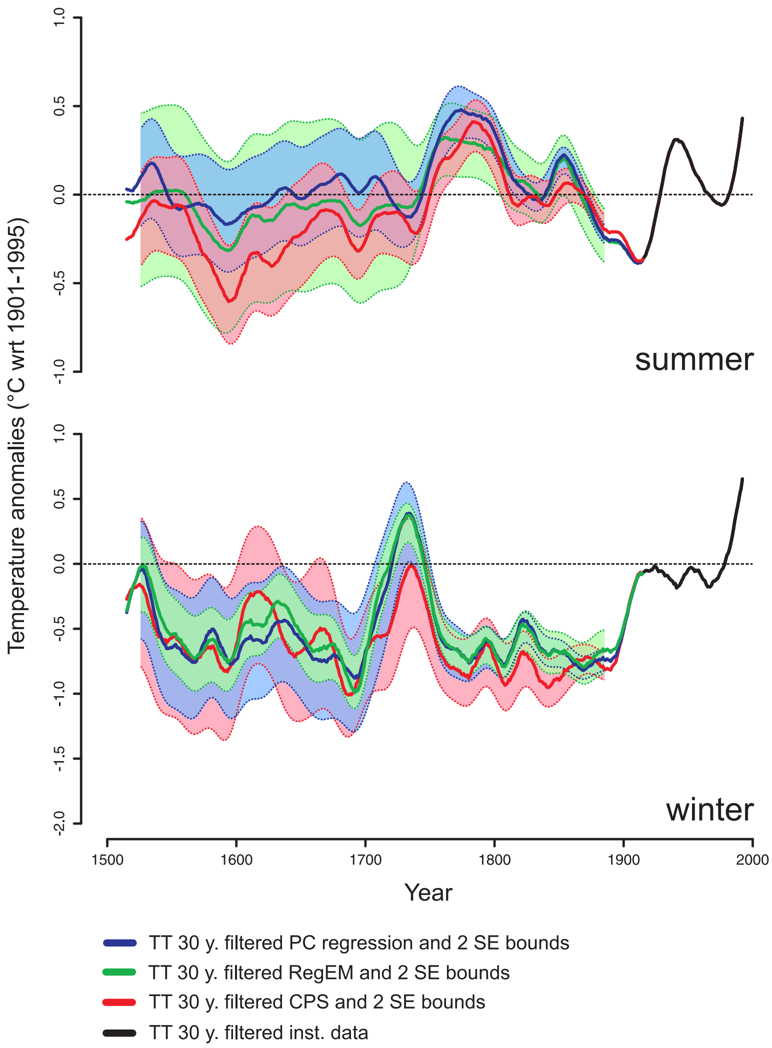

Figure 3.1: *30-year Gaussian filtered European summer and winter average temperature anomalies (wrt the 1901 to 1995 calibration period) back to 1500, PC regression (blue line, corresponding 2 standard errors (SE) blue shaded), RegEM (green line, corresponding 2 SE green shaded) and CPS (red line, corresponding 2 SE red shaded). Instrumental surface air temperature data (Mitchell and Jones (2005), 1901 to 2002, and Hansen et al. (2001), 2003 to 2007) in black.*

works are continuous over longer time periods. In consequence, less predictor networks are used for winter than for summer temperature reconstruction, and the imputation errors are smaller. The uncertainty bounds of CPS very likely differ from those of PC regression and RegEM, e.g. 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 3.1), which is not the case for PC regression (Table 3.1, selected subset summer compared to winter). Nevertheless, Figure 3.1 shows that there is better accordance of the three results for winter with regard to 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. While in winter early instrumental series, ice core data and documentary evidence only are used, the types of proxy data, and the temperature signal inherent vary more for summer.

In "a surrogate climate" RegEM proved that more skilful results can be obtained by better incorporating errors inherent to proxy data to capture low-frequency variability (*Riedwyl et al.*, 2008a). 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 to perform 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.

The strong positive temperature anomalies at the end of the 18th century of the summer results (Figure 3.1, top) are likely to be an artefact of too warm early instrumental measurements, and support the findings of *Moberg et al.* (2003), *Frank et al.* (2007) and *Böhm et al.* (2008). The pronounced negative winter temperature anomalies at the end of the 17th century (Figure 3.1, 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.

3.3.2 Analysis of the temperature field reconstructions

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 3.2 shows the comparison of reconstructed temperature anomaly fields averaged over the warmest summer decade (1789 to 1798) and coldest winter decade (1689 to 1698), using PC regression (left) and RegEM (right). For the warmest summer decade (Figure 3.2, top), the maximal temperature anomalies are more pronounced using PC regression than using RegEM. There are some simi-

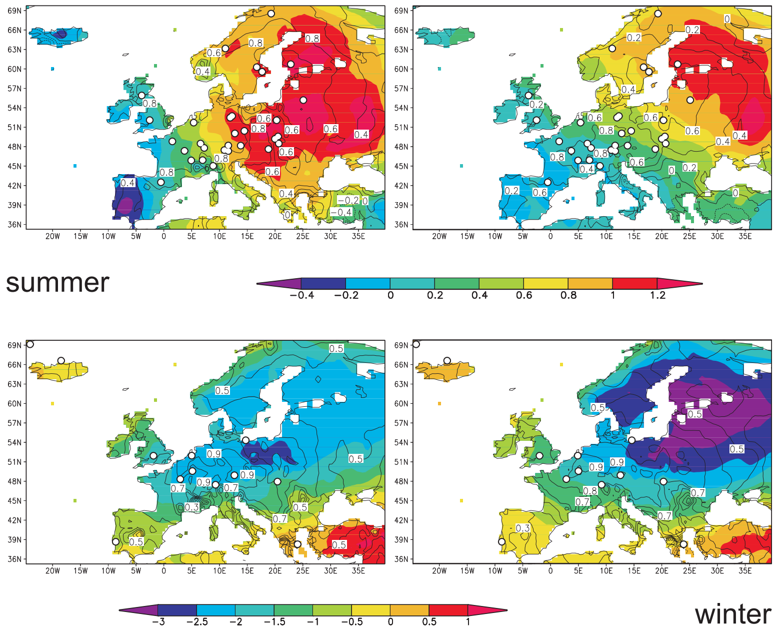

Figure 3.2: *European surface air temperature anomaly (wrt to the 1901 to 1995 calibration period) fields, averaged over the warmest summer decade (1789 to 1798, top) and coldest winter decade (1689 to 1698, bottom) using PC regression (left) and RegEM (right). Contours represent RE skill patterns. RE is calculated for each gridpoint over the verification period from 1961 to 1995. White dots indicate the locations of the proxies used for reconstruction.*

larities 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 3.2, bottom) the negative temperature anomalies in the North East are more pronounced using RegEM than using PC regression, in general though the two anomaly patterns are remarkably similar. Overall, the results agree more for the coldest winter decade while there are discordances in summer (see also the supplementary online material); the positive temperature anomalies are generally more pronounced for summer using PC regression. The reason 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 seem to have more weight using PC regression, and dominate periods, where they explain most variance, which is not the case for RegEM. The spatial field of reconstructed single cold and warm winters and summers are provided in the supplementary online material. PC regression and RegEM agree more for winter (warmest year: 1724; coldest year: 1695) than for summer (warmest year: 1798 PC regression, 1826 RegEM; coldest year: 1821). Again, the RE scores for PC regression (Figure 3.2, left) indicate slightly higher skill than those for RegEM (Figure 3.2, right). However, the performance of PC regression and RegEM to reconstruct warmest summer and coldest winter decades (Figure 3.2), as well as extreme years (supplementary online material) are comparable.

3.4 Conclusions

Related to European summer and winter average and field 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).

Testing RegEM with TTLS shows that this technique is 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 easier applied.

Highest skill both for summer and winter is achieved for PC regression. 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.*, 2008a). We conclude that temperature reconstructions can not be improved significantly by only replacing the reconstruction technique. Discordances are very likely to be due to the spatial distribution and uncertainties inherent to the proxy data, as well as their limited availability. Thus, here more robust results are achieved 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.

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.

3.5 Supplementary Material

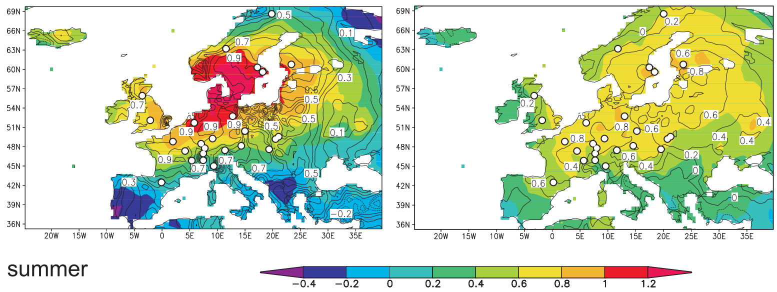

Figure 3.3: *As Figure 3.2 in the main article, but for warmest summer decade for RegEM (1774 to 1783).*

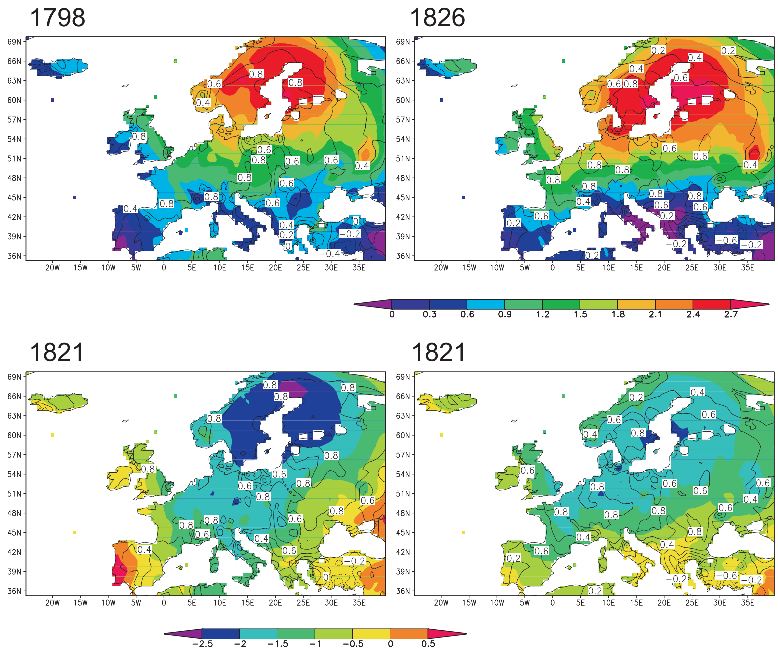

Figure 3.4: *Warmest (top) and coldest (bottom) summer year for PC regression (right) and RegEM (left) in the reconstruction period 1500 to 1900.*

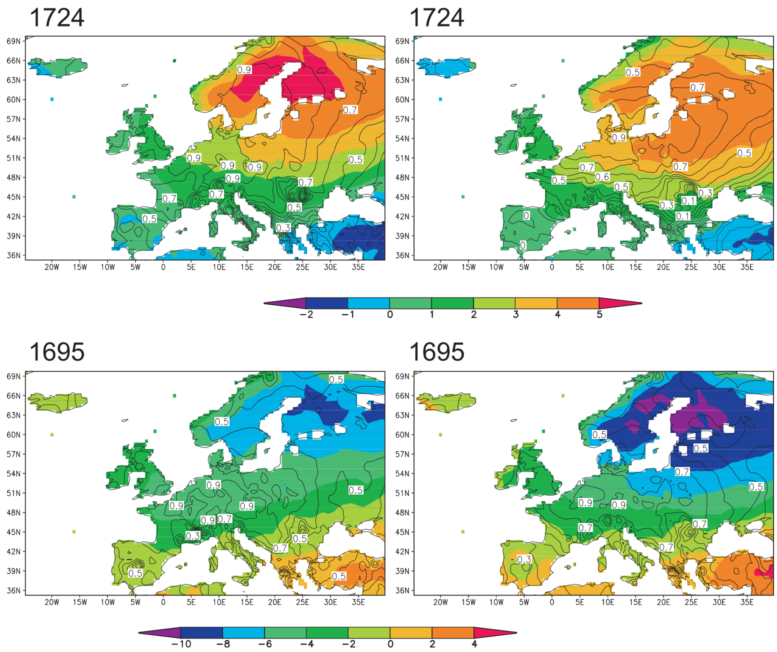

Figure 3.5: *Warmest (top) and coldest (bottom) winter year for PC regression (right) and RegEM (left) in the reconstruction period 1500 to 1900.*

References

Böhm, R., M. Brunetti, J. Hiebl, P. Jones, D. Frank, and M. Maugeri (2008), The early instrumental warm-bias: A solution for long central European temperature series, *Climatic Change, in review.*

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

Cook, E., K. Briffa, and P. Jones (1994), Spatial regression methods in dendroclimatology: a review and comparison of two techniques, *International Journal of Climatology*, *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. Frank, R. Wilson, and K. Briffa (2005), Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, *Geophysical Research Letters*, *32*, L07711.

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, *Quaternary Science Reviews*, *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, *Journal of Geophysical Research*, *106*, 23947–23963.

Jansen, E., and coauthors (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*, Cambridge University Press, Cambridge, United Kingdom and New York.

Jones, P., and M. Mann (2004), Climate over past millennia, *Reviews of Geophysics*, *42*, RG2002.

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, *Geophysical Research Letters*, *34*, L07710.

Lee, T., F. Zwiers, and M. Tsao (2008), Evaluation of proxy-based millennial reconstruction methods, *Climate Dynamics, in press.*

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 forStudying Decadal Scale Climatic Change in Europe, *Climatic 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., and S. Rutherford (2002), Climate reconstruction using'Pseudoproxies', *Geophysical Research Letters*, *29*, 1501.

Mann, M., S. Rutherford, E. Wahl, and C. Ammann (2007), Robustness of proxy-based climate field reconstruction methods, *Journal of Geophysical Research*, *112*, D12109.

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

Moberg, A., H. Alexandersson, H. Bergstroem, and P. Jones (2003), Were southern Swedish summer temperatures before 1860 as warm as measured?, *International Journal of Climatology*, *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 (2007), Analysis of the Moberg et al. (2005) hemispheric temperature reconstruction , *Climate Dynamics, 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, *Climate Dynamics*, *26*, 387–405.

Riedwyl, N., M. Kuettel, J. Luterbacher, and H. Wanner (2008), Comparison of climate field reconstruction techniques: Application to Europe, *Climate Dynamics, 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, *Journal of Climate*, *18*, 2308–2329.

Schneider, T. (2001), Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values., *Journal of Climate*, *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 (2007), Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation, *Journal of Earth Sciences, in press.*

Wilks, D. (1995), *Statistical methods in the atmospheric sciences*, Academic Press San Diego.

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, *Geophysical Research Letters*, *32*, L15711.

Chapter 4

European summer temperature variability over the last millennium

Nadja Riedwyl, Jürg Luterbacher and Heinz Wanner

Oeschger Centre for Climate Change Research (OCCR), Bern, Switzerland
Institute of Geography, Climatology and Meteorology, University of Bern, Bern, Switzerland

to be submitted

Abstract

We present reconstructions of European summer temperature variability over the last millennium. Reconstruction is performed using PC regression, RegEM, and additionally CPS. The combination of three reconstruction techniques and compilation of long and continuous proxy series provide the basis for these new results, and for a detailed analysis of European millennial summer temperature amplitudes. Their robustness is tested by cross-validation with the leave-one-out algorithm. Furthermore, the performances of PC regression and RegEM are compared focusing on reconstructed temperature

fields averaged over periods of accordance and discordance. We show that a rather diverse picture of summer temperature variability over the last millennium is very likely caused by a lack of coherence in the temperature signals inherent to the proxy data. We conclude that the "Medieval Warm Period" is not noticeable in the results, whereas some evidence is provided for the "Little Ice Age".

4.1 Introduction

There is still little evidence on European past temperature evolution and change back to the year 1000 AD (*Brázdil et al.*, 2005). Many reconstructions at the European scale put the current warming in the context of climate variability over the past 500 years (*Luterbacher et al.*, 2004, 2006; *Xoplaki et al.*, 2005; *Casty et al.*, 2005, 2007; *Raible et al.*, 2006; *Pauling et al.*, 2006; *Fischer et al.*, 2007), whereas only few multiproxy reconstructions cover the last millennium (*Guiot et al.*, 2005). An extension back to the year 1000 AD can help to further establish whether or not the recent 20th century temperature increase is unusual in a longer-term context (*Goosse et al.*, 2006).

While past temperature variability over the last millennium has been thoroughly analyzed and discussed at the northern hemispheric scale (*Jansen and coauthors*, 2007), this study is a contribution for the regional scale. On the basis of long, continuous proxy series, it is examined to what extent the "Medieval Warm Period" (MWP) (*Lamb*, 1965; *Hughes and Diaz*, 1994; *Bradley et al.*, 2003b), the "Little Ice Age" (LIA) (*Grove*, 1988, 2004; *Pfister et al.*, 1999; *Wanner et al.*, 2000; *Luterbacher et al.*, 2001), and the temperature increase of the 20th century (*Jansen and coauthors*, 2007) are captured by summer temperature reconstructions.

The number of millennial European proxy records is still moderate, and their spatial distribution sparse. A "complete" picture of past temperature field variability over the last millennium can thus not be provided. In consequence, the focus lies mainly on the reconstruction of average summer temperatures, and moreover on their robustness. Reconstructed summer temperature fields for specific periods are presented in order to compare the performance of the reconstruction techniques spatially. Reconstruction is performed, applying

an ensemble of three reconstruction techniques. Principal component (PC) regression, the technique traditionally used for European climate field reconstruction (*Luterbacher et al.*, 2004; *Casty et al.*, 2005; *Pauling et al.*, 2006) is applied. In addition, regularized expectation maximization (RegEM) introduced by *Schneider* (2001); *Rutherford et al.* (2005), and *Mann et al.* (2007) which proved to be suitable and promising for seasonal European climate field reconstructions (*Riedwyl et al.*, 2008b), as well as composite-plus-scaling (CPS) (e.g. *Jones and Mann*, 2004; *Esper et al.*, 2005) are used.
The robustness of the reconstructions is tested by cross-validation with the leave-one-out algorithm. This allows for an understanding to what extent individual proxy series influence the reconstruction results.
In section 2 we describe the proxy data used, and the three reconstruction techniques applied. In section 3, first the summer average temperature reconstruction results and the robustness exercise are presented. Second, PC regression and RegEM are compared spatially over specific periods back to 1000 AD. The results are discussed in section 4, followed by conclusions in section 5.

4.2 Data and Methods

Past summer temperatures are estimated for the period 1000 to 1900 AD by establishing a statistical relationship between the predictand and the predictors during the calibration period 1901 to 1970 AD. A longer calibration period is preferable, though the limitation in length here is due to some shorter proxy series. The predictand is the European surface summer air temperature field taken from *Mitchell and Jones* (2005) at 0.5°x0.5° resolution. To represent summer temperatures, the average of June, July and August is calculated. Europe is defined by the land area 24.5° W to 39.75°E and 35.25° N to 69.75° N. The number of predictors is maximally 18 (by the end of the 20th century) and minimally 6 (in the year 1000 AD), varying in temporal and spatial availability (Table 4.1, Figure 4.1). Thus, 12 different predictor networks, decreasing in number back in time, are defined for reconstruction back to the year 1000 AD. The data comprise mostly long tree ring chronologies (*Briffa et al.*, 2001; *Kirchhefer*, 2001; *Hantemirov and Shiyatov*,

Table 4.1: *List and description of summer proxies used for reconstruction.*

No.	Proxy type definition	Period (AD)	Reference
1	Tree ring widths, Tornetrask, 68°2 N, 19° E	5500 BC-1997	Briffa et al. 2001
2	Tree rings (MXD), Alps, 46.26 N, 7.49 E	755-2004	Buentgen et al. 2006
3	Oak indexed ring widths, Bourgogne, 46°15 N, 4° E	681-1991	Lambert et al. 1992
4	Larch index ring widths, living trees, Vallée des Merveilles, 44°02 N, 7°27 E	1187-1974	Serre et al. 1978
5	Larch index ring widths, all trees, Vallée des Merveilles, 44°02 N, 7°27 E	988-1974	Serre et al. 1978
6	Tree rings, Jaemtland, 63.3 N, 13.3 E	1632BC-2002	Linderholm and Gunnarson, 2005
7	Larch trees, Yamal, 67.25 N, 69.5 E	2000 BC-1996	Hantemirov et al. 2002
8	Oak tree ring widths, Belfast, Northern Ireland	1001-1970	Baillie et al. 1977
9	Oak tree ring widths, Southern Scotland	946-1975	Pilcher and Baillie 1980
10	Larch index ring widths, all trees, Haut Verdon, 44°06 N, 6°39 E	1160-1994	Bélingard and Tessier, 1993
11	Tree rings, Pyrenees, 42.4 N, 0.3 E	1260-2005	Buentgen et al., in press
12	Scots pine tree rings, Coastal northern Norway,	1354-1989	Kirchhefer et al. 2001
13	Scots pine tree rings, Coastal northern Norway	1355-1989	Kirchhefer et al. 2001
14	Grape harvest dates, Burgundy, 47.14 N, 4.95 E	1370-2003	Chuine et al. 2004
15	Documentary evidence, Low Countries	761-2005	Shabalova et al. 2003
16	Documentary evidence, Germany, 48 N, 11.4 E	1500-1995	Luterbacher et al. 2004
17	Documentary evidence, England, 51.47 N, 0.32 W	1500-1995	Luterbacher et al. 2004
18	Documentary evidence, Debilt, 52.1 N, 5.18 E	1500-1995	Luterbacher et al. 2004

2002; *Guiot et al.*, 2005; *Büntgen et al.*, 2006, 2008; *Linderholm and Gunnarson*, 2005). Furthermore, the reconstructions are based on grape harvest dates (*Chuine et al.*, 2004) and historical documentary series (*Shabalova and van Engelen*, 2003; *Luterbacher et al.*, 2004). The selection comprises long, continuous and mostly natural proxy records spanning over the last millennium.

PC regression is used as traditionally applied and described in (*Luterbacher*

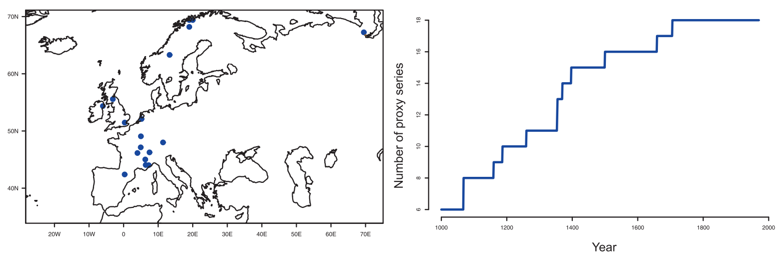

Figure 4.1: *In blue the proxy locations used for reconstruction (left), and the varying number of proxy series over time (right) are indicated.*

et al., 2004; *Riedwyl et al.*, 2008a), and RegEM with truncated total least squares (TTLS) as regularization scheme according to *Mann et al.* (2007). With PC regression, the proxy series that explain most variance are calibrated against the most dominant patterns of spatial variability of the instrumental measurement series. RegEM with TTLS is a covariance-based iterative algorithm, replacing missing values (given available ones) with plausible values, taking into account under-determined settings (*Schneider*, 2001; *Rutherford et al.*, 2005; *Mann et al.*, 2007). Furthermore, the predictand is usually represented by its leading PC (*Mann et al.*, 2007). Here, reconstruction with RegEM is not only performed by reconstructing the PC of the underlying spatial field, but by reconstructing the past European summer average temperature series directly as well. Thus, we apply a multiple variant of RegEM with TTLS, given the European summer temperature average series as predictand. As proposed in *Huybers* (2005), ambiguity is avoided using the simple average rather than PC. For CPS too, the predictand is the average series, and the summer temperatures are reconstructed by building a composite of the predictors, which is then scaled by the amplitude of the predictand (e.g. *Esper et al.*, 2002, 2005; *Jones and Mann*, 2004; *Moberg et al.*, 2005).

Separate reconstructions are performed based on the maximal predictor network (18 proxy series), using the calibration period 1901 to 1940 AD, and the verification period 1941 to 1970 AD for validation. The reduction of error (RE), coefficient of efficiency (CE), and the squared correlation coefficient (r^2) are calculated.

To access the robustness and the uncertainties inherent to the reconstructions, cross-validation with the leave-one-out algorithm is performed. For $k = 1, ..., 18$, we temporarily remove the k^{th} proxy series from the predictor network, and calibrate and reconstruct with $k - 1$ proxy series to finally compare this result to the one including the k^{th} series.

To more easily compare the results of the three techniques, the lower-frequency variations, and thus the filtered reconstruction results with associated filtered uncertainty bounds are shown. As in *Xoplaki et al.* (2005) the standard errors (SE) for the 30-yr Gaussian filtered reconstruction results are calculated using the predictor verification residuals after making the residuals consistent with Gaussian white noise (*Mann et al.*, 1998; *Briffa et al.*, 2002).

4.3 Results

4.3.1 European summer average temperatures and their robustness

Figure 4.2 shows European summer average temperature anomalies over the last millennium, using PC regression (blue), CPS (red) and RegEM (green). The three results display temperature anomalies being on a lower level over the reconstruction period (1000 to 1900 AD) than during the calibration period (1901 to 1970 AD); moreover including the 2 SE bounds, the reconstructed values do not exceed the 20th century warmth (indicated by instrumental measurements in black). The result for PC regression is rather balanced. Noticeable are maximal positive temperature anomalies around the 14th, 15th and 16th century. The CPS result displays generally lower and more variable temperature anomalies than the one of PC regression. They agree in a raise from negative to positive anomalies in the 12th century, both being more pronounced for CPS than for PC regression. Thus, the maximal positive anomaly peak of the CPS result is at the end of the 12th century. Furthermore, a negative anomaly peak is displayed at the beginning of the 19th century, which is much more pronounced than in the PC regression and RegEM results. The result of RegEM shows a noticeable period of pronounced negative anomalies in the 11th and 12th century, followed

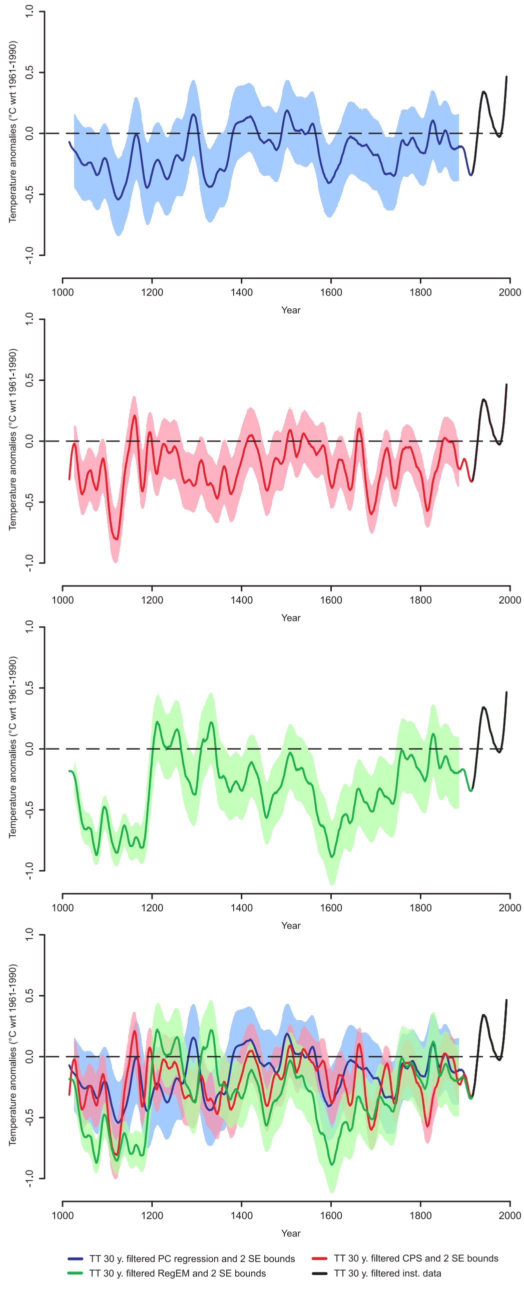

Figure 4.2: *30-year Gaussian filtered European summer average temperature anomalies (wrt 1961 to 1990) over the last millennium. PC regression (blue line, corresponding 2 standard errors (SE) blue shaded, CPS (red line, corresponding 2 SE red shaded) and RegEM (green line, corresponding 2 SE green shaded). Instrumental surface air temperature data (Mitchell and Jones (2005), 1901 to 2002, and Hansen et al. (2001), 2003 to 2007) in black.*

Table 4.2: *RE, CE and r^2 scores for the reconstructions using the maximal predictor network.*

Reduction of error (RE)

	PC reg su	RegEM su	CPS su
maximal set:	0.436	0.29	0.019

Coefficient of efficiency (CE)

	PC reg su	RegEM su	CPS su
maximal set:	0.233	0.035	-0.333

Correlation coefficient r squared (r2)

	PC reg su	RegEM su	CPS su
maximal set:	0.441	0.466	0.28

by a steep increase to maximal positive anomalies around the 13th to 14th century. Furthermore, pronounced negative anomalies are also displayed at the beginning of the 17th century. The increase thereafter, resulting in a positive anomaly peak at the beginning of the 19th century is in agreement with the PC regression result. The variability of the RegEM result is largest and the temperature anomalies lowest, compared to PC regression and CPS. Displaying the three results together (Fig. 4.2, bottom), reveals that their behavior, besides few exceptions, is rather diverse. There is some agreement between the CPS and PC regression result in the 12th century, and between the three of them by the end of the reconstruction period. Table 4.2 shows that the skill of the three reconstructions is modest, the reconstruction of CPS having lowest skill. Furthermore, the RE score of PC regression is highest, followed by RegEM, and r^2 is slightly higher for RegEM than for PC regression.

Figure 4.3 displays the results of the robustness exercise. The 18 reconstructions based on $k-1$ proxy series (grey), together with their average (colors) are shown. Thus, instead of 2 SE, here the uncertainty bounds are indicated by the deviations from the original reconstructions leaving out each proxy series once. The deviations are smallest for PC regression, followed by CPS and finally RegEM. The main meanderings and outliers result from having left out the same proxy series. Thus, e.g. the maximum late wood density tree ring series from the Alps (Table 4.1, No. 2) seems to have con-

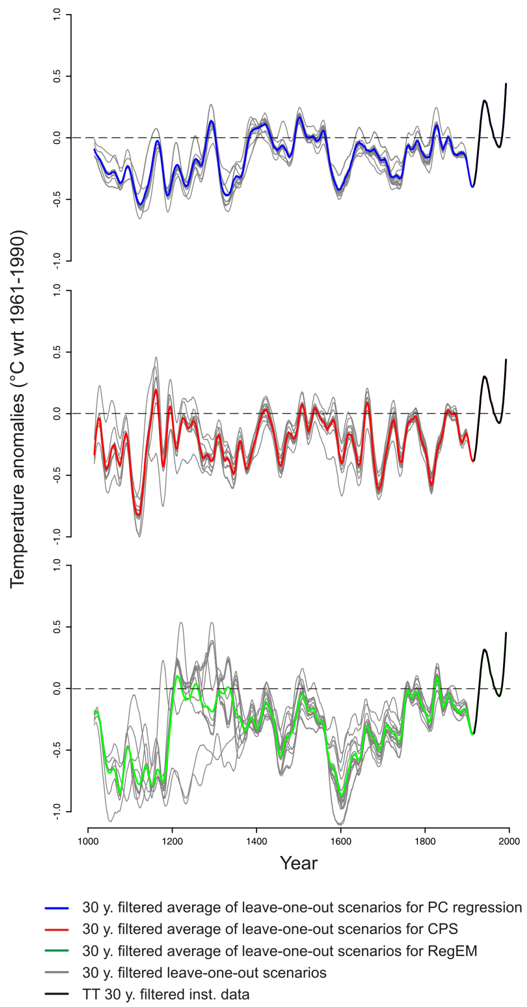

Figure 4.3: *The results using the leave-one-out algorithm (grey lines), and their average (colors as in Fig. 4.2), are shown.*

siderable influence on the result, as well as the two tree ring series from the "Valle des Merveilles" (Table 4.1, No. 4 and 5), and the tree ring series from Belfast (Table 4.1, No. 8). The RegEM result significantly deviates in the 13th to 14th century due to leaving out the tree ring series from "Valle des Merveilles" (Table 4.1, No. 4), and those from Northern Norway (Table 4.1, No. 12 and 13). In consequence, for RegEM the average of the leave-one-out scenarios does not agree with the original result in Figure 4.2, which seems not to be the case for PC regression and CPS.

4.3.2 European summer temperatures fields

Figure 4.4 shows eight reconstructed European summer temperature fields (a-h) using PC regression (left), and RegEM (right), averaged over periods of accordance and discordance back to the year 1000 AD. The spatial fields are chronologically ordered from past (top) to present (bottom). Analyzing the reconstruction results in Figure 4.2, we define periods of discordance,

- *1040 to 1105 AD*, *1141 to 1200 AD* and *1275 to 1305 AD* (Fig. 4.4a, c, and d)

and periods of accordance

- *1106 to 1140 AD*, *1480 to 1525 AD*, *1590 to 1635 AD*, *1710 to 1740 AD* and *1800 to 1835 AD* (Fig. 4.4b, e, f, g, h).

Obviously, accordance of the average summer temperature reconstruction results (Fig. 4.2) does not necessarily include the same for the spatial fields (e.g. as seen in Fig. 4.4b and Fig. 4.4f). During the 11th and 12 century (Fig. 4.4a-c), the RegEM temperature anomaly fields clearly display cooler conditions than PC regression (see also Fig. 4.2). The maximal negative temperature anomalies are strongly pronounced in the North East for RegEM, whereas for PC regression they are only moderately indicated in the South West. Furthermore, Figure 4.4d and Figure 4.4e show the pronounced positive temperature anomaly peaks of the PC regression result in the 14th and 16th century, including large areas of the European East being warmer, whereas the RegEM result is first still indicating cooler conditions in disagreement with PC regression, however warmer conditions (Fig. 4.4e) thereafter. Clearly, the exceptional cold of the RegEM result (Fig. 4.2) at the beginning of the 17th century is shown (Fig. 4.4f). The PC regression field is revealing cooler conditions as well, however not as pronounced, and locally different. Finally, the spatial fields more and more agree, both showing rather similar patterns (Fig. 4.4g), particularly closest to present (Fig. 4.4h).

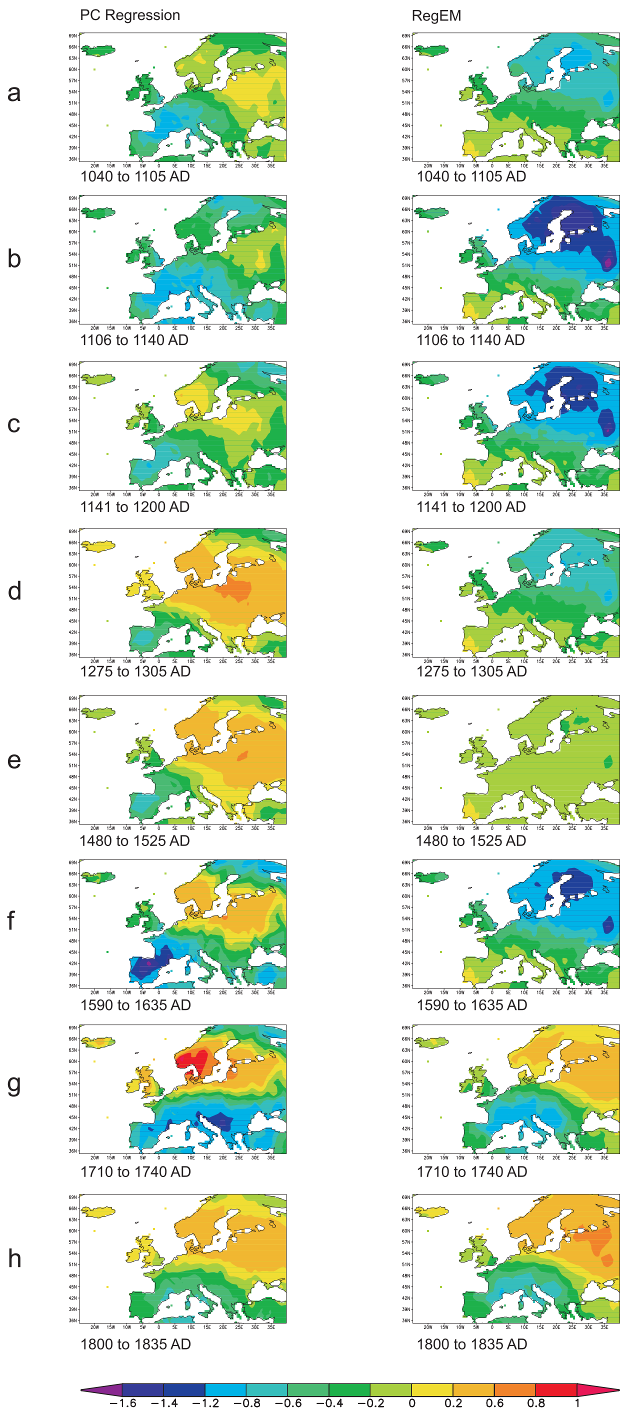

Figure 4.4: *European summer temperature anomaly fields (wrt to 1961 to 1990), averaged over selected periods of accordance and discordance between PC regression (left) and RegEM (right), over the last millennium. Chronologically ordered from past (top) to present (bottom).*

4.4 Discussion

4.4.1 Methodological considerations

The variability of the reconstruction is lowest for PC regression, and seems to be rather damped over the reconstruction period compared to CPS and RegEM. Testing the performance of PC regression in "a surrogate climate" showed a similar effect (*Küttel et al.*, 2007; *Riedwyl et al.*, 2008a). In contrast, the RegEM result retains more variability. One reason is likely to be that the estimation of the regression coefficients of the two techniques differ: PC regression is applied estimating the coefficients using ordinary least squares (OLS), whereas for RegEM the coefficient are estimated based on TTLS. RegEM has proved that adequate results can be achieved, incorporating errors inherent to proxy data using TTLS as regularization scheme (*Mann et al.*, 2007). However, the choice of the truncation parameter can still be improved, by a more rigorous adaptation, which is a field of ongoing research (*Sima and Van Huffel*, 2007). Using CPS, by definition the variability of the result corresponds to the one of the predictand average series during the calibration period.

The main cause of the diversity of the reconstruction results in Figure 4.2 seems to be the heterogeneity of the proxy data, and the incoherence in their summer temperature signal over the last millennium. An earlier study showed that for summer the temperature signals inherent to proxy data seems to be less coherent already over the last 500 years, which leads to discrepancies in the results of the three techniques (*Riedwyl et al.*, 2008b). There are differences in the performance of the techniques in handling the diversity of responses in the proxy data. PC regression seems to be less affected than CPS and RegEM. The truncation not only of the predictand, but of the predictors as well, seems to reduce deviations.

Figure 4.3 indicates that leaving out single series leads to largest deviations between the RegEM results. This is very likely due to the fact that the variability of the RegEM reconstruction is largest. Thus, with regard to its variability and robustness the CPS result lies between PC regression and RegEM. It is problematic that the influence of certain proxy series turnes

out to be considerable. Certainly, a larger number of coherent proxy series can make a difference in the reconstruction performance. Nevertheless, it is a future challenge to enhance the robustness of the results, specially of RegEM with TTLS. Furthermore, the results of this robustness exercise put into perspective the impressions of confidence by only focusing on the 2 SE bounds. It is certainly important, not only to rely on SE, but other measures of errors to address uncertainties and enhance the reliability of the reconstruction (*Jones and coauthers*, 2008).
Finally, Figure 4.4 displays the discordances further back in time, and the increasing similarity by the end of the reconstruction period related to the increasing availability of proxy series. Only few proxies are located in the Southwest and the East, which very likely is the reason of disagreement between PC regression and RegEM over these areas. Thus, not much confidence can be attributed to the results for Eastern and Southern Europe. We hypothesize that if the proxy data quality is high and the spatial and temporal distribution sufficient, less discrepancies between the results can be expected.

4.4.2 Millennial temperature history

The MWP is not explicitly seen in the three European summer reconstruction results, despite for some single series this is the case (*Büntgen et al.*, 2006). Although the multiproxy results display some peaks of warmth (Fig. 4.2, PC regression in the 14th, 15th and 16th century, and CPS at the end of the 12th, and RegEM at the beginning of the 13th century), they are not distinct enough to assume the MWP. Thus, more coherent summer temperature signals are needed to verify its existence. More evidence can be found for the LIA, which is also well marked in the reconstruction of *Guiot et al.* (2005). The reconstruction using RegEM indicates a period of cooling from the mid 14th to the mid 19th century, in correspondence with the LIA (*Wanner et al.*, 2000; *Luterbacher et al.*, 2001). There might be a relation of the cold period in the RegEM result of the 11th to 12 century with an assumed major glacier advance in 1100 AD (*Holzhauser et al.*, 2005). However, the climax of the Late Maunder Minimum (1675-1715 AD), seen in reconstructions covering the past 500 years, which include early instrumental measurement series as proxies (*Luterbacher et al.*, 2004; *Riedwyl et al.*, 2008b), is not evident here using mostly natural proxy archives. Furthermore, for both the PC regression and the CPS result, a cooling during the LIA period is less obvious than for RegEM. Likely the significant peak at the beginning of the 19th century in the CPS results, which is e.g. also seen in the reconstruction of *Büntgen et al.* (2006), is related to the LIA.

4.5 Conclusions and perspectives

The combination of three reconstruction techniques and compilation of appropriate long and continuous proxy series have provided the basis for the detailed reconstruction of European summer temperature variability over the last millennium. The picture of past summer temperature variability displayed by the results is rather diverse, still the recent warming can be put in a longer-term context. There is no noticeable evidence for the MWP in the reconstruction results. Mainly the RegEM result provides evidence for the LIA, however the cooling from the 14th to 19th century is not as pronounced in the results of PC regression and CPS. The reconstruction results with associated 2 SE bounds do not reach the level seen in the instrumental measurement records for the 20th and 21th century.
The incoherence between the reconstructions using PC regression, RegEM and CPS indicate that the sample size of long and continuous proxy series needs to be further enlarged, and methodological adjustments to be made in order to achieve more reliable results. To exclude, if possible, the arbitrariness of the PC selection fully (CPS), and to some extent (RegEM) is strongly suggested. The errors-in-variables approach TTLS is a promising regularization scheme (*Mann et al.*, 2007; *Sima and Van Huffel*, 2007), however more rigorous adaptations for the unambiguous selection of the truncation parameters need to be developed in future.
Finally, the robustness exercise shows the considerable influence of some proxy series on the reconstruction result. They could be identified with the leave-one-out algorithm. Thus, cross-validation can help to interpret reconstructions with associated uncertainties more adequately. More robust reconstructions of summer temperature variability over the last millennium

can likely be achieved on the basis on more coherent multi proxy data. They require a more detailed understanding of the climate signal and careful interpretation of its spatial significance for the individual proxy series.

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.

References

Baillie, M.G.L. (1977), The Belfast, Northern Ireland, UK, oak chronology to AD 1001, *Tree-ring Bulletin*, *37*, 1–12.

Bélingard, C., and L. Téssier (1993), Etude dendrochrécologique comparée de vieux peuplements de Larix decidua Mill. dans les Alpes Françaises du Sud, *Dendrochronologia*, *11*, 69–78.

Bradley, R., M. Hughes, and H. Diaz (2003b), CLIMATE CHANGE: Climate in Medieval Time, *Science*, *302*, 404–405.

Brázdil, R., C. Pfister, H. Wanner, H. Storch, and J. Luterbacher (2005), Historical Climatology In Europe–The State Of The Art, *Climatic Change*, *70*, 363–430.

Briffa, K., T. Osborn, F. Schweingruber, I. Harris, P. Jones, S. Shiyatov, and E. Vaganov (2001), Low-frequency temperature variations from a northern tree ring density network, *Journal of Geophysical Research*, *106*, 2929–2942.

Briffa, K., T. Osborn, F. Schweingruber, P. Jones, S. Shiyatov, and E. Vaganov (2002), Tree-ring width and density data around the Northern Hemisphere: Part 1, local and regional climate signals, *The Holocene*, *12*, 759–789.

Büntgen, U., D. Frank, D. Nievergelt, and J. Esper (2006), Summer Temperature Variations in the European Alps, ad 755–2004, *Journal of Climate*, *19*, 5606–5623.

Büntgen, U., D. Frank, H. Grudd, and J. Esper (2008), Long-term summer temperature variations in the Pyrenees, *Climate Dynamics, in press.*

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

Casty, C., C. Raible, T. Stocker, H. Wanner, and J. Luterbacher (2007), European climate pattern variability since 1766, *Climate Dynamics, in press.*

Chuine, I., P. Yiou, N. Viovy, B. Seguin, V. Daux, and E. Le Roy Ladurie (2004), Historical phenology: Grape ripening as a past climate indicator, *Nature*, *432*, 289–290.

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. Frank, R. Wilson, and K. Briffa (2005), Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, *Geophysical Research Letters*, *32*, L07711.

Fischer, E., J. Luterbacher, E. Zorita, S. Tett, C. Casty, and H. Wanner (2007), European climate response to tropical volcanic eruptions over the last half millennium, *Geophysical Research Letters*, *34*, L05707.

Goosse, H., O. Arzel, J. Luterbacher, M. Mann, H. Renssen, N. Riedwyl, A. Timmermann, E. Xoplaki, and H. Wanner (2006), The origin of the European "Medieval Warm Period", *Climate of the Past*, *2*, 99–113.

Grove, J. (1988), *The Little Ice Age*, Routledge, London and New York.

Grove, J. (2004), *Little Ice Ages: Ancient and Modern*, Routledge, London and New York.

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

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, *Journal of Geophysical Research*, *106*, 23947–23963.

Hantemirov, R., and S. Shiyatov (2002), A continuous multimillennial ring-width chronology in Yamal, northwestern Siberia, *The Holocene*, *12*, 717–726.

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

Hughes, M., and H. Diaz (1994), Was there a medieval warm period, and if so, where and when?, *Climatic Change*, *26*, 109–142.

Huybers, P. (2005), Comment on Hockey sticks, principal components, and spurious significance by S. McIntyre and R. McKitrick, *Geophysical Research Letters*, *32*, L20705.

Jansen, E., and coauthors (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*, Cambridge University Press, Cambridge, United Kingdom and New York.

Jones, P., and coauthers (2008), High-resolution paleoclimatology of the last millennium: a review of curent status and future prospects, *Holocene, in review.*

Jones, P., and M. Mann (2004), Climate over past millennia, *Reviews of Geophysics*, *42*, RG2002.

Kirchhefer, A. (2001), Reconstruction of summer temperatures from tree-rings of Scots pine (Pinus sylvestris L.) in coastal northern Norway, *The Holocene*, *11*, 41–52.

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, *Geophysical Research Letters*, *34*, L07710.

Lamb, H. (1965), The early medieval warm epoch and its sequel, *Palaeogeography, Palaeoclimatology, Palaeoecology*, *1*, 13–37.

Lambert, G., et al. (1992), Létalon de datation dendrochronologique Bourgogne 29, *In Maurice, G.L.B.,editor, Les Veines du Temps*, Autun, France: Musée Rolin, 123–56.

Linderholm, H., and B. Gunnarson (2005), Summer tempaerature variability in central Scandinavia during the last 3600 years, *Geografiska Annaler*, *87A*, 231–241.

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 forStudying Decadal Scale Climatic Change in Europe, *Climatic 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.

Luterbacher, J., Eds. P. Lionello, P. Malanotte-Rizzoli, and R. Boscolo (2006), Mediterranean climate variability over the last centuries: A review in: *The Mediterranean Climate: an overview of the main characteristics and issues*, Elsevier, Amsterdam, the Netherlands, 27–148.

Mann, M., R. Bradley, and M. Hughes (1998), Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, *392*, 779–787.

Mann, M., S. Rutherford, E. Wahl, and C. Ammann (2007), Robustness of proxy-based climate field reconstruction methods, *Journal of Geophysical Research*, *112*, D12109.

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

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.

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, *Climate Dynamics*, *26*, 387–405.

Pfister, C., H. Graßl, and J. Luterbacher (1999), *Wetternachhersage: 500 Jahre Klimavariationen und Naturkatastrophen (1496-1995)*, Paul Haupt, Bern.

Pilcher J.R and M.G.L. Baillie (1980), Six modern oak chronologies from Ireland, *Tree-Ring Bulletin*, *40*, 23–34.

Raible, C., et al. (2006), Climate Variability-Observations, Reconstructions, and Model Simulations for the Atlantic-European and Alpine Region from 1500-2100 AD, *Climatic Change*, *79*, 9–29.

Riedwyl, N., M. Kuettel, J. Luterbacher, and H. Wanner (2008a), Comparison of climate field reconstruction techniques: Application to Europe, *Climate Dynamics, in press.*

Riedwyl, N., J. Luterbacher, and H. Wanner (2008b), An ensemble of European summer and winter temperature reconstructions back to 1500, *Geophysical Research Letters, submitted.*

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, *Journal of Climate*, *18*, 2308–2329.

Schneider, T. (2001), Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values., *Journal of Climate*, *14*, 853–871.

Serre, F. (1978), The dendroclimatic value of the European larch (Larix decidua Mill.) in the the French Alps, *Tree-ring Bulletin*, *38*, 25–34.

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

Sima, D., and S. Van Huffel (2007), Level choice in truncated total least squares, *Computational Statistics and Data Analysis*, *52*, 1103–1118.

Wanner, H., H. Holzhauser, C. Pfister, and H. Zumbühl (2000), Interannual to century scale climate variability in the European Alps, *Erdkunde*, *54*, 62–69.

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, *Geophysical Research Letters*, *32*, L15711.

Chapter 5

Conclusions and perspectives

The equilibrium of atmosphere, ocean, land and humans is disturbed, and unpredictable, unprecedented climatic events might occur in the future. In this context, the question arises to what extent the past is still the key to the future.
However, future scenarios depend on “a priori” knowledge, about past and present climate conditions. The analysis of past European climate variability thus provides valuable knowledge, that helps to propagate future projections, and places recent climate changes in a longer-term context. Therefore, the description of past temperature variability is essential, and methodological considerations involved need to be addressed with due care. Intense discussions about the use of reconstruction methods at the NH scale have shown that clarifications are needed. This thesis contributes to better understanding the strengths and weaknesses in performances of methods using PC regression, RegEM and CPS at the European seasonal scale over the last millennium. Our results emphasize the necessity to test, validate and adapt the reconstruction methods with respect to the specific spatial and temporal resolution they are applied to.

5.1 “A priori” knowledge

The comparison of PC regression and RegEM in the “surrogate climates” of the NCAR CSM 1.4 and ECHO-G 4 climate models over the last millennium reveals seasonal differences in the performance of the two CFR methods: European temperature variability is more adequately reconstructed for summer than for winter. Furthermore, it is demonstrated that the proxy data quality has a crucial impact on the reconstruction skill. More skilful results are achieved with RegEM with regard to lower frequency variability, whereas, using PC regression, the past temperature amplitudes are underestimated. The spatial validation of the two methods discloses an underestimation of the target temperature variations by PC regression, however indicating similar RE scores for RegEM. RegEM with TTLS proves that more adequate reconstruction results are achieved by taking into account errors inherent to the proxy data. If the SNR in the pseudoproxies is high, PC regression and RegEM perform alike. If the SNR is low, the performance of RegEM is more robust than that of PC regression. Both methods exhibit deficits if the noise inherent to the pseudoproxies is red or the SNR is very low. PC regression tends to underestimate the amplitudes of the target temperature series, while they are overestimated by RegEM. We conclude that CFR methods can be improved, however need to be further optimized in future.

5.2 Evaluation of reconstructions over the last 500 years

Analyzing an ensemble of summer and winter temperature reconstructions for the last 500 years improves the understanding of how reliably past temperature amplitudes are reconstructed using PC regression compared to CPS and RegEM. We conclude that PC regression and RegEM show a more similar performance compared to CPS. However, differences between the results of the three methods are larger for summer than for winter reconstructions; indeed the difference in similarity of the results between summer and winter is striking. This is likely due to more strength and coherence in the temperature signal inherent to winter proxy data. We found that the highest skill, both for summer and for winter, is achieved using PC regression, followed by RegEM and, finally, CPS. Furthermore, we conclude that reconstructions can not be significantly improved by the choice of a different method only:

further improvement can be achieved by the enhancement of the temporal and spatial availability of the proxy data. Applying an ensemble of several methods to reconstruct the same target based on identical input reveals that past temperature variability over the last 500 years is likely larger than indicated by earlier PC regression reconstructions.

5.3 Reconstructions and their robustness over the last millennium

Focusing on the results of European summer average temperature reconstructions over the last millennium, the MWP is not apparent, whereas evidence is found for the LIA. However, the cooling during the LIA is mainly seen in the RegEM result, and is not as pronounced in the results of PC regression and CPS. Generally, the ensemble of millennial summer temperature reconstructions displays a rather diverse picture of past summer temperature amplitudes. We use the leave-one-out algorithm as a further tool for statistical error characterization of reconstructions besides the associated 2 SE. The robustness exercise discloses that the influence of few proxy series is considerable. New long and continuous proxy records enlarging the sample size of the predictors are likely to further augment the confidence and reliability of the summer reconstruction results. Nevertheless, the robustness using RegEM with TTLS in this context needs to be further optimized. A first step towards this aim proves to be taking the European average series as predictand and not the PC of the European spatial field. Furthermore, a more rigorous adaptation for the selection of the TTLS parameters needs to be developed in future.

5.4 Recommendations

From a theoretical point of view RegEM is an improvement compared to PC regression. However, no method can be said to perform best in absolute terms. Thus, RegEM too, has its deficits, e.g. with regard to robustness for a small and incoherent sample size, and requires as well as PC regression future adjustments and optimizations. In this respect, specific analyses to assess the skill and robustness of CFR methods for particular purposes are crucial. “Surrogate climates” are a valuable tool to obtain “a priori” knowledge of the performance of reconstruction methods, and offer the possibility to test hypotheses about the nature of the proxy data. Nevertheless, as seen in our results, the use of a “virtual reality” can also lead to discrepancies. While for “surrogate climates” over the last millennium, temperature reconstructions have higher skill for summer than for winter, it is the opposite for real-world climate. On the one hand, this might be explained by the choice of the calibration period or the quality and spatial distribution of the proxy data. On the other hand, the fact that under ideal conditions, i.e. equal strength of the temperature signal inherent to pseudoproxy data, the range of European temperature variability is larger for winter than for summer, possibly plays a more dominant role. Furthermore, as Europe is represented by a very coarse grid in “a surrogate climate”, spatially higher resolved regional climate model simulations are needed for more consistent comparisons between physical climate model simulations (forced signals) and statistical reconstructions (estimated variability).

- We recommend consequent use of several methods and climate variables for reconstruction. An ensemble of reconstructions can help to either disclose probable biases and inconsistencies or enhance the reliability of estimated past climate amplitudes.

- We recommend to explore Bayesian approaches for obtaining robust calibrations, integrating the high- and low frequency signals inherent to proxy data by estimating Bayesian priors.

- We recommend to consider not only temperature, but other climate variables for the robustness and skill tests of methods in “surrogate climates”, and for real-world climate reconstructions.

Many conversations with colleagues about the SNR inherent to proxy data and its quantification have accompanied this work. Much progress is being made in filtering out climatic signals of proxies, and an impressive wealth of knowledge exists about what individual paleoclimate proxy series respond to and what the associated uncertainties are. More conversations and specific efforts are now needed to "translate" this expert knowledge of proxy data into statistical terms and parameters. One of the most limiting factors, however, is the finite length of instrumental records. All efforts to improve the target, and each additional year of instrumental measurements, will in future help to reduce uncertainties and strengthen the basis for calibration, and thus for reconstruction.

Appendix A

Skill assessment of two reconstruction techniques for three climate variables

Skill assessment of two reconstruction techniques for three climate variables

Nadja Riedwyl, Jürg Luterbacher and Heinz Wanner
Oeschger Centre for Climate Change Research and National Centre of Competence in Research on Climate (NCCR), University of Bern, Switzerland

Introduction

Reconstruction of past climate variability is of high importance in the discussion on climate change. The question if caveats of climate field reconstruction (CFR) techniques lead to false interpretations and conclusions is crucial and has to be addressed. In consequence, sensitivity tests and thorough evaluations are performed (1,2) and further needed. We present the skill assessment of principal component (PC) regression and regularized expectation maximization (RegEM) for past temperature, precipitation and 500 hPa geopotential height reconstructions.

Data and Methods

We independently reconstructed European land surface temperature (LST), land surface precipitation (LSP) and 500hPa geopotential height (Z500) fields using PC regression and RegEM back to 1766 AD, as in the reconstruction period 1766 to 1900 AD data availabilty is well assured (3). The spatial resolution of the LST and LSP gridded fields is 0.5° x 0.5°, and 2.5° x 2.5° of the Z500 grid. Figure 1 shows schematically how the two CFR techniques are applied to predictor and predictand data. Skill is assessed with the commonly used reduction of error (RE), and by the analysis of the reconstructed anomaly patterns over the verification period. RE values in the interval]0,1] indicate reconstructive skill.

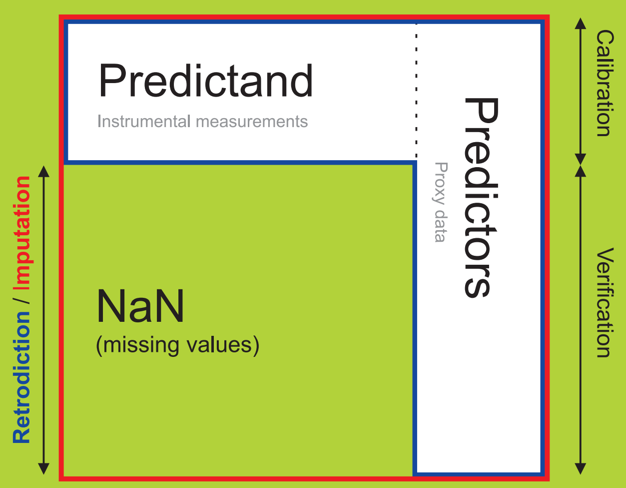

Figure 1

Scheme of the analogousness / differences between PC regression (blue) and RegEM (red). PC regression corresponds to "retrodiction" and RegEM to the imputation of past temperature values. The input matrix for both techniques is indicated in colors.

Results

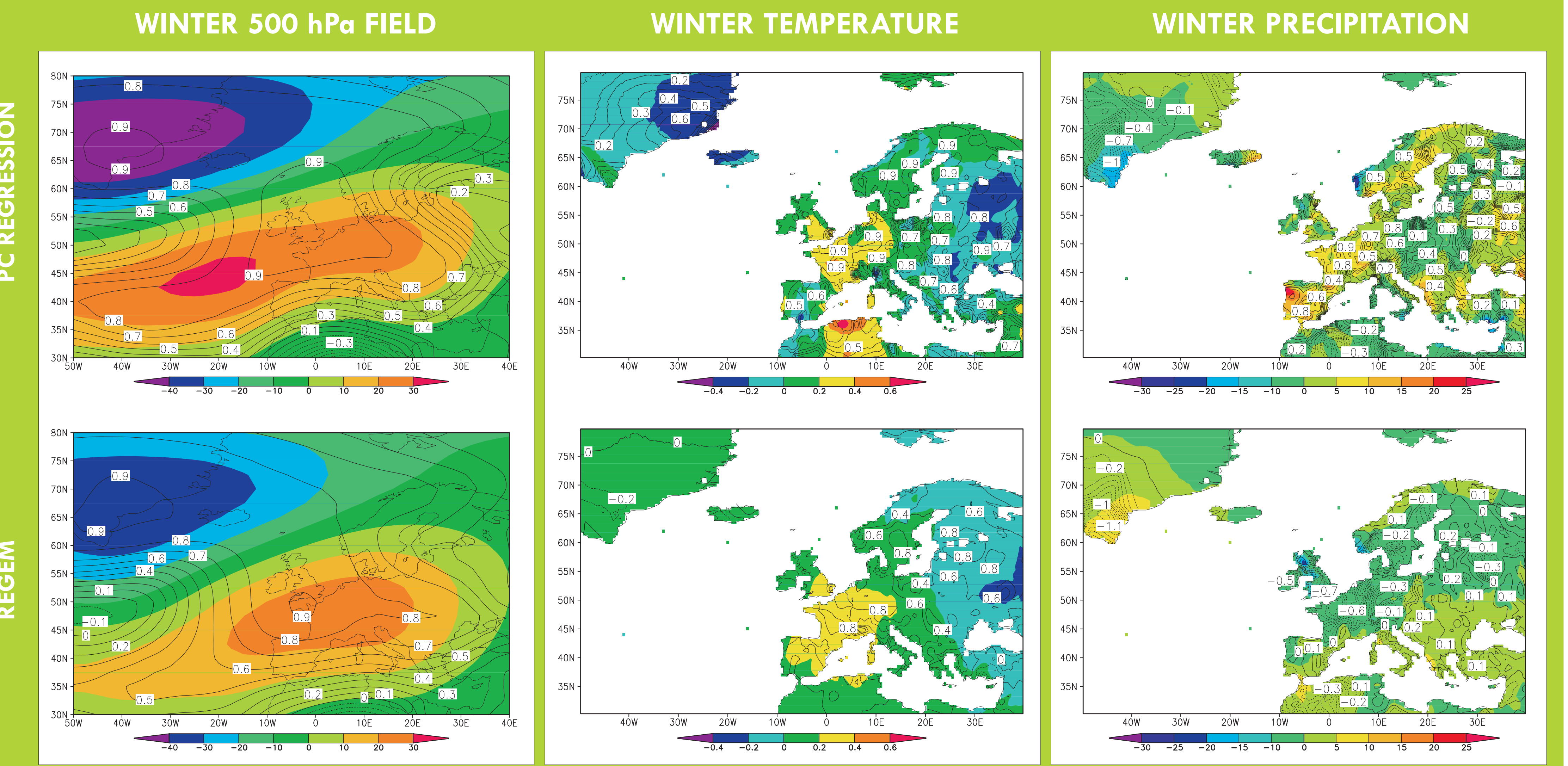

Figure 2

Average of reconstructed anomalies [shaded] and RE [contour] over the verification period using PC regression (top) and RegEM (bottom). For reconstructed LST and LSP the anomalies are with regard to (wrt) the 1901 to 1960 AD calibration period, and with verification over 1961 to 1995 AD. The anomalies of Z500 are wrt 1948 to 1978 AD for calibration, and with verification from 1979 to 1995 AD. Shown are the results for winter only. The order from left to right corresponds to the decrease in skill of the results.

Conclusion

Reconstructions with highest RE values are achieved for Z500, followed by LST and finally LSP. Furthermore, the RE values are rather higher for PC regression than for RegEM, especially for LSP. However, additional tests show that RE does not seem to fully capture the performance of CFR techniques (4). We conclude that there are differences in skill between reconstructions of PC regression and RegEM, and for LST, LSP and Z500. Thus more reliable measures to assess skill have to be developed in the future.

References

1. M. E. Mann **et al.**, ***J. of Geophys. Res.* 112**, D12109 (2007).
2. T. Lee **et al.**, ***Clim. Dynam.*** in press (2007).
3. C. Casty **et al.**, ***Clim. Dynam.* 29**, 791-805 (2007).
4. N. Riedwyl **et al.**, ***Clim. Dynam.*** in press (2008).

Acknowledgments

This work has been supported by the Swiss National Science Foundation (SNSF) through its National Center of Competence in Research on Climate (NCCR) project PALVAREX 2. Many thanks go to Carlo Casty for kindly providing the data.

Acknowledgments

My thanks go to all the people who have supported, accompanied and guided me throughout my PhD.
To Heinz Wanner, who has given me the opportunity to work in the KLIMET group, and to Jürg Luterbacher for inspiring discussions. I thank Scott Rutherford for the second referee of this thesis and his scientific input. Christoph Kull for taking me to the PAGES workshop in Wengen. Many thanks go to Eugene Wahl, Caspar Ammann and Nick Graham for valuable discussions, which found their starting point there. Their inputs, and those of Rob Wilson, Dave Frank and Christoph Raible let me learn and rethink. Basilio Ferrante and Pavel Michna saved me from tedious computer problems, my special thanks go to them.
To my friend Marlène Thibault for her commitment to cross-read my articles.
Monika, thank you for taking me to the right direction.
Henne, thank you for believing in me.

Curriculum Vitae

Nadja Riedwyl, born 28 February 1976 in Bern, Switzerland

Education

2005-2008	PhD student in climate sciences
2000-2005	Diploma in Mathematics, Thesis on ‘Das Agglomerieren von INAR Zeitreihen’
2003-2004	Erasmus student, Universidad Autonoma de Madrid, Spain
1997-1999	Physics, ETH, Zürich
1991-1997	Matura Typus B
1993-1994	AFS Exchange student, Amtsgymnasiet Roskilde, Denmark
1989-1991	Untergymnasium, Bern
1987-1989	Sekundarschule, Bern
1983-1987	Primarschule, Bern

Professional experience

2005-2005	Assistant, Interkantonaler Rückversicherungsverband, Bern
2001-2002/ 2003-2005	Research assistant, Mathematical institute, University of Bern
1999-2000	Marketing assistant, UBS Private Banking, Zürich