Testing a European winter surface temperature reconstruction in a surrogate climate

M. Küttel,[1] J. Luterbacher,[1] E. Zorita,[2] E. Xoplaki,[1] N. Riedwyl,[1] and H. Wanner[1]

Received 17 August 2006; revised 2 March 2007; accepted 13 March 2007; published 13 April 2007.

[1] We evaluate the skill of a European winter surface air temperature reconstruction over the last 500 years using pseudoproxies obtained from the ECHO-G and HadCM3 climate models. The emphasis is thereby on the effect of the reduction of available predictors back in time, an issue that has not yet been investigated in detail at continental and seasonal scale. It is found that the key factor in determining the reconstruction skill is the number of predictors and particularly their spatial distribution. However considering the usually insufficient spatial and temporal predictor availability in paleo-reconstructions, the quality of the predictors becomes more important further back in time. Not surprisingly, the lowest reconstruction skill is found in the early period when the predictor network is reduced. Important differences between ECHO-G and HadCM3-based pseudoproxy reconstructions are discussed and implications for future analyses are presented.
Citation: Küttel, M., J. Luterbacher, E. Zorita, E. Xoplaki, N. Riedwyl, and H. Wanner (2007), Testing a European winter surface temperature reconstruction in a surrogate climate, *Geophys. Res. Lett.*, *34*, L07710, doi:10.1029/2006GL027907.

1. Introduction

[2] Reconstructions of past climate have recently been subject to investigative studies centering on their ability to appropriately reproduce low-frequency temperature variability [*von Storch et al.*, 2004; *Mann et al.*, 2005, 2007; *Rutherford et al.*, 2005; *Bürger et al.*, 2006; *Wahl et al.*, 2006; *Wahl and Ammann*, 2007]. These studies were performed using AOGCMs as a numerical laboratory in which reconstruction methods as well as their potential limitations can be tested and assessed by deriving proxy records from the model climate, so-called "pseudoproxies". This approach thereby proved to be a valuable tool to determine the dependence of the reconstruction skill on the temporal availability and spatial distribution of predictors, their quality as well as the statistical model applied. This is a crucial contribution to identify and possibly quantify reconstruction uncertainties and thus to improve our knowledge of past climate variability.

[3] This study tests the winter reconstruction of European land surface air temperature over the last 500 years by *Luterbacher et al.* [2004] (hereinafter referred to as L04) using pseudoproxies (see Figure 1) obtained from the AOGCMs ECHO-G [*von Storch et al.*, 2004] and HadCM3 [*Tett et al.*, 2007]. The L04 reconstruction appears particularly qualified since it allows testing the influence of various factors on the reconstruction skill: First, the regression method used by L04 is a nested approach, i.e. separate regression models were calculated for each different proxy network available over the 500 years. Around 100 models had to be calibrated/verified and reconstructed to obtain a European winter temperature reconstruction. This method thus allows the testing of the impact of a spatially and temporally reduced predictor network within a single predictor set. Previous studies also tested the influence of an increasingly sparse network by using different models, however each one with a constant number of predictors over the reconstruction period [*Mann and Rutherford*, 2002; *Rutherford et al.*, 2003; *Zorita et al.*, 2003; *von Storch et al.*, 2004, supplementary online material; *Mann et al.*, 2007]. Secondly, the L04 reconstruction is suitable for testing the impact of the quality of the predictors on the reconstruction skill since it is based on a large number of instrumental data primarily after 1750 (Figure 1, bottom) and proxy information (temperature indices based on documentary evidence, ice core based temperature reconstructions, sea ice conditions, etc.) before. Finally, the L04 reconstruction covers European land areas at seasonal resolution, in contrast to previous pseudoproxy based studies which addressed annually resolved reconstructions on hemispheric or even global scale. It is therefore reasonable to perform a pseudoproxy based study on a smaller temporal (sub-annual) and spatial scale with a generally larger temperature amplitude. This serves as a test for scale dependencies that might possibly help explaining the loss of amplitude in low-frequency temperature variability in some regression-based reconstructions of past climate [*von Storch et al.*, 2004; *Bürger et al.*, 2006; also discussed in *Mann et al.*, 2005, 2007; *Wahl et al.*, 2006; *Wahl and Ammann*, 2007].

[4] The main focus of this study is on the effect of the reduction in the predictors' number back in time. Additional emphasis is put on the impact of a spatially less uniformly distributed predictor network in early centuries, following earlier studies [*Bradley*, 1996; *Mann and Rutherford*, 2002; *Rutherford et al.*, 2003; *Zorita et al.*, 2003; *von Storch et al.*, 2004]. Since the quality of the predictors also generally decreases backwards in time, the combined effect of a spatio-temporally reduced predictor network with additionally increased uncertainties in the predictors themselves can be highlighted. To determine the model-dependence of the results, this study is simultaneously performed with two AOGCMs (ECHO-G and HadCM3). By using the same code of the reconstruction algorithm (nested PCA-multiple regression) as L04 discussions about

[1]National Centre of Competence in Research on Climate (NCCR) Climate and Institute of Geography, University of Bern, Bern, Switzerland.
[2]Institute for Coastal Research, GKSS-Research Centre, Geesthacht, Germany.

Copyright 2007 by the American Geophysical Union.
0094-8276/07/2006GL027907

source: https://doi.org/10.7892/boris.25487 | downloaded: 12.10.2024

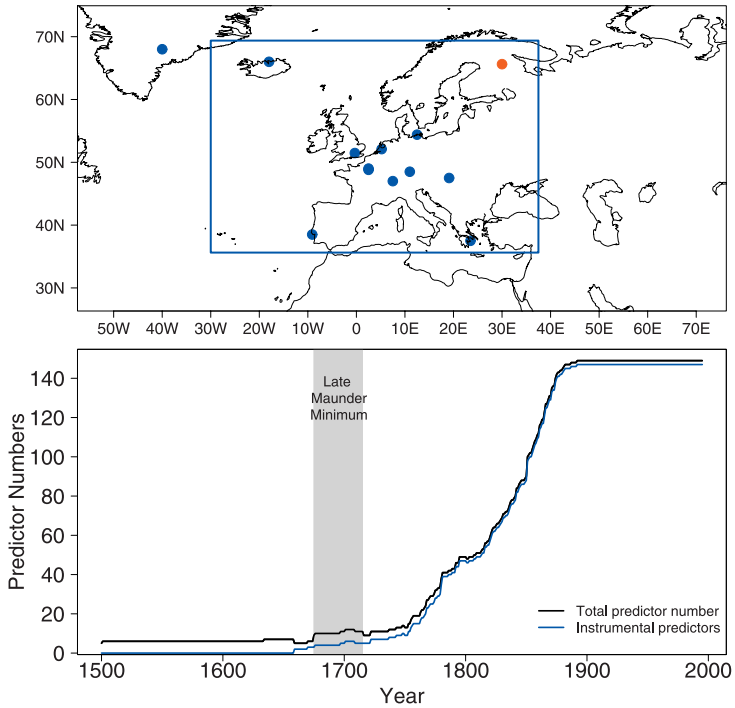

Figure 1. (top) Locations of the proxies used by L04 during the late Maunder Minimum (1675–1715, blue points) and of one additional pseudoproxy (red point) added in order to improve the spatial coverage. The blue frame covers the area of temperature reconstruction. (bottom) Total number of available predictors over time in the L04 reconstruction and number of instrumental predictors.

the proper methodological replication of the reconstruction can be excluded [*von Storch et al.*, 2006; *Wahl et al.*, 2006; *Mann et al.*, 2007]. It is important to state that this study solely tests the skill of the L04 reconstruction and does not discuss possible methodological limitations of the applied regression model as, e.g., did *Bürger et al.* [2006] and *Mann et al.* [2007].

2. Data and Methods

[5] The pseudoproxies were produced following the approach by, e.g., *Mann and Rutherford* [2002] and *von Storch et al.* [2004]. The pseudoproxy $P = T_g + \epsilon$ with T_g being the simulated surface air temperature at the grid box collocated to the L04 proxy network and ϵ being the added realization of white noise. To represent uncertainty in the proxy records and to test its influence, three uniform levels of white noise were added to T_g with the resulting pseudoproxies describing the locally simulated gridded temperature variability by 25, 50 and 75% (corresponding to correlation coefficients of 0.5, 0.7 and 0.87), respectively. Additionally, one pseudoproxy set was constructed where the noise in each individual series was scaled to values encountered in the real world proxies of L04 from the local correlation of each of the 166 proxies used (see Tables S1 and S2 in the supplementary online material of L04) with the gridded instrumental data set by *New et al.* [2000] in the overlapping period 1901–2000. This latter predictor set is designed to represent best the L04 proxy set in terms of their quality. Synthetic examples in the AOGCMs have revealed that the correlation coefficients obtained over the entire 500 year period do not significantly differ from those derived within the calibration period (not shown).

[6] Figure 1 summarizes the data used in L04. The bottom panel identifies the contribution of instrumental records to the full predictor set. In order to clearly contrast the influence from the instrumental series on the reconstruction performance, we regard them here as "perfect", i.e. unperturbed samples directly from the model grid. In reality, this might be overly optimistic [see *Brohan et al.*, 2006] despite the fact that they have been homogenized or at least quality checked (L04 supplementary online material). To determine the impact of the reduction in the number of available predictors back in time, a common feature in paleo-reconstructions, continuous pseudoproxy sets over the entire reconstruction period (1500–1900) and sets with a reduction back in time (Figure 1, bottom) according to L04 were designed. The limited network prior to 1750 (Figure 1) cannot properly resolve some of the high-amplitude variations at the European periphery, in particular Scandinavia. Therefore, an artificially augmented proxy network is tested, where a single predictor over eastern Scandinavia is added during the pre-1750 period. The goal is to evaluate how strongly a single point can affect European average reconstruction skill. Because only the last 500 years of the ECHO-G run were used, the critical influence from the climate drift visible in the first few centuries of the 1000 year simulation has mostly vanished [*Osborn et al.*, 2006]. The HadCM3 simulation does not appear to suffer from such issues. Nevertheless, the two climate models used in this study differ considerably in their climate sensitivity, the forcings included and their historical changes and amplitudes. Most importantly, ECHO-G does

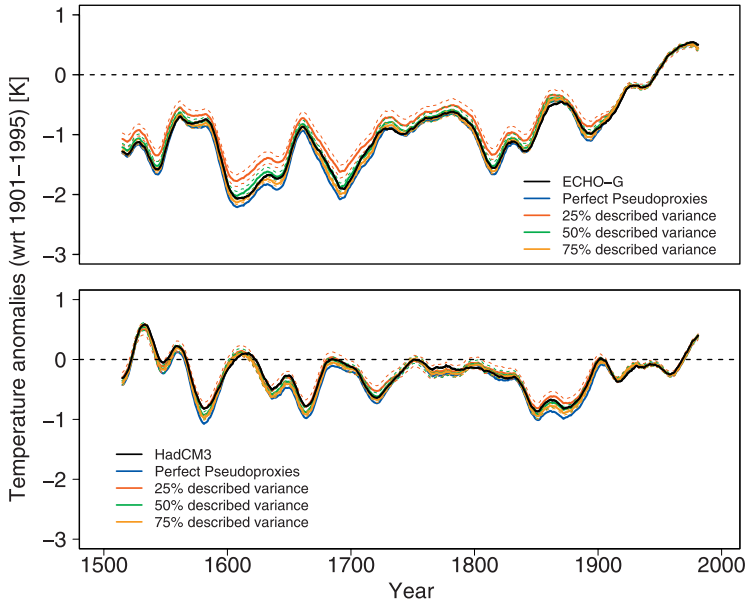

Figure 2. Average European winter surface air temperature anomalies (wrt 1901–1995) over the last 500 years based on (top) ECHO-G [*von Storch et al.*, 2004] and (bottom) HadCM3 [*Tett et al.*, 2007], smoothed with a 30-year gaussian low-pass filter. The black line represents the simulated mean temperature while the colored lines are reconstructions based on pseudoproxies with different levels of described temperature variance. All predictors are continuous over the entire reconstruction period and all are degraded, independent from the proxy they represent. One hundred Monte Carlo realizations of noise were used to estimate the median and the 5% and 95% range (dashed lines).

not contain any representation of the anthropogenic tropospheric aerosol as well as land-use change forcing, both implemented in HadCM3. Along with influences from different climate sensitivities, this may cause the noticeably larger European temperature variability and the 20th century trend simulated by ECHO-G (Figure 2, top). The significantly different temporal temperature evolution in the two models points to a large internal variability on the regional scale, as E. G. found by *Wagner and Zorita* [2005].

[7] For the reconstruction, we used the same code as described in L04. It is a multivariate principal component regression designed to reconstruct climate fields (see *Luterbacher et al.* [2002, 2004] for a detailed description). Unlike some recent studies [*von Storch et al.*, 2004, and partially *Bürger et al.*, 2006] no detrending of the data was applied prior to the calibration. The reconstruction produced in this study represents the winter season (December–February average) covering the European land areas 30°W–37.5°E and 35.625°N–69.375°N with a 3.75° × 3.75° resolution, according to the climate models' resolution.

3. Results and Discussion

[8] Figure 2 presents the averaged European winter surface air temperature anomalies (with respect to the 1901–1995 average) over the last 500 years, smoothed with a 30-year gaussian low-pass filter. The upper panel is based on ECHO-G and the lower one on HadCM3. The black curve shows the simulated temperature while the colored lines are the pseudoproxy-based reconstructions with different qualities of the predictors. Here, all predictors are continuous throughout the entire reconstruction period and all of the 166 predictors were degraded, i.e. independent of the quality of the proxy used by L04 that they represent. For the degraded series the median of 100 Monte Carlo iterations is shown along with the 5% and 95% quantiles.

[9] The reconstructions capture the shape of the simulated temperature history generally very well, largely independent of the noise level. This is in agreement with recent studies investigating hemispheric data [e.g., *Mann et al.*, 2005, 2007; *Rutherford et al.*, 2005; *Wahl et al.*, 2006; *Wahl and Ammann*, 2007]. The reconstructions based on perfect pseudoproxies apparently overestimate the cold spells throughout the pre-calibration period in both models. Tests revealed (not shown) that this might be related to an overfitting due to a large number of predictors relative to the number of predictands during the calibration period (caused by the coarse model resolution) as well as the length of the calibration period itself. This important issue should however be investigated in more detail.

[10] Figure 3 shows reconstructions that more closely mimic the real world conditions of L04 with their quickly deteriorating number of predictors before 1750 (Figure 1, bottom). Additionally, it is important to note that only non-instrumental predictors were degraded, leading to a maximum impact prior to 1750 when the number of instrumental predictors decreases to zero before 1659 (Figure 1, bottom). The generally good visual skill of the pseudoproxy-based reconstructions during the calibration (1901–1960) and verification period (1961–1995) confirms that the L04 reconstruction is properly implemented in this study and has climatological meaning [*Wahl and Ammann*, 2007]. Prior to the twentieth century the reconstructions present lower skill and show an increased underestimation towards earlier centuries. The underestimation is however strongly model-dependent. The ECHO-G shows a larger amplitude than the HadCM3 and also a stronger dependence on the

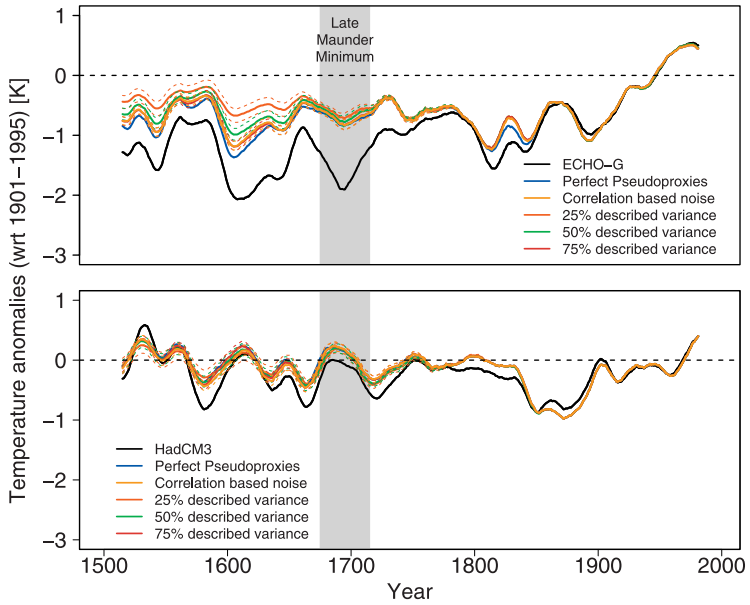

Figure 3. Reconstructions based on predictors with the same availability over time as in L04 (Figure 1, bottom). Only non-instrumental predictors are degraded.

noise level (Figure 3). Both models indicate significant loss in skill primarily prior to the late eighteenth century with the cold spells being significantly underestimated. Interestingly, ECHO-G indicates a significant underestimation of the cold late Maunder Minimum (1675–1715, e.g., *Luterbacher et al.* [2001]) largely insensitive to the quality of the predictors, while the preceding cold spell also shows an underestimation, however with a strong dependence on the noise level: the higher the signal-to-noise ratio the smaller is the underestimation. This difference in performance between late Maunder Minimum and earlier, equally severe cold spells can be explained by the presence of "perfect" instrumental predictors during the late Maunder Minimum and their absence before. These results suggest that under real world conditions L04 should be capable of capturing the true temperature variations over Europe after ~1750 as the spatial coverage, the total number, and in particular the number of instrumental series increases. However, from our model based exercises one has to conclude that loss of amplitude and significant underestimation of cold periods might exist.

[11] The underestimation of the larger temperature anomalies in the early period shown in Figure 3 may partly be explained by a spatially insufficient predictor network. The spatial plots of the reconstructions produced in this study (not shown) have indicated that the underestimation of particularly the cold spells is largest over Scandinavia. This is not surprising since no proxy data is available in this region during the early centuries. As a test to evaluate the impact of an artificially improved spatial network, we have added one predictor in this region (Figure 1, red dot). For the results shown in Figure 4 this predictor is made available in the period 1500–1750. Its data was degraded with white noise to mimic documentary data [*Pauling et al.*, 2003; *Xoplaki et al.*, 2005]. The red line in Figure 4 demonstrates the significant improvement of the reconstruction if the additional predictor is present over Scandinavia compared to the L04 network (blue line). While the improved spatial network leads to an almost perfect overlap of the reconstruction and the model mean in HadCM3 during some periods, ECHO-G still shows some general underestimations, however clearly smaller than with the original proxy network. Thus the addition of a single predictor leads in this case to a significant decrease in the reconstruction uncertainties.

4. Conclusions

[12] The test of the European winter surface air temperature reconstruction of L04 in the surrogate climate of the two AOGCMs ECHO-G and HadCM3 has indicated that the real world reconstruction skill over Europe could be influenced by the quality of the predictors as well as their availability over time and space. The results appear to be partly dependent on the amplitudes of the simulated temperature variability, thereby emphasizing the need to perform such studies with more than one climate model.

[13] The reconstruction performs well when a predictor set with a continuous availability over time and space is assumed. In this specific case, the quality of the predictors is of rather lower importance. This is in agreement with recent evidence at larger spatial scales [e.g., *Mann et al.*, 2005, 2007; *Wahl et al.*, 2006; *Wahl and Ammann*, 2007]. In reality, paleo-reconstructions however have to deal with predictor networks that decrease significantly backwards in time. Our surrogate climate exercises point to a danger that this can lead to spatially insufficient coverage and non-reliable reconstructions. In this context significant underestimations of the true grid mean appear, with the influence from the quality of the predictors to become much more important. Artificially improving the spatial coverage by an additional predictor clearly improves the reconstruction skill.

[14] The availability of predictors over time and mainly space has thus proven to be the key factor in determining the reconstruction skill. It is the factor significantly

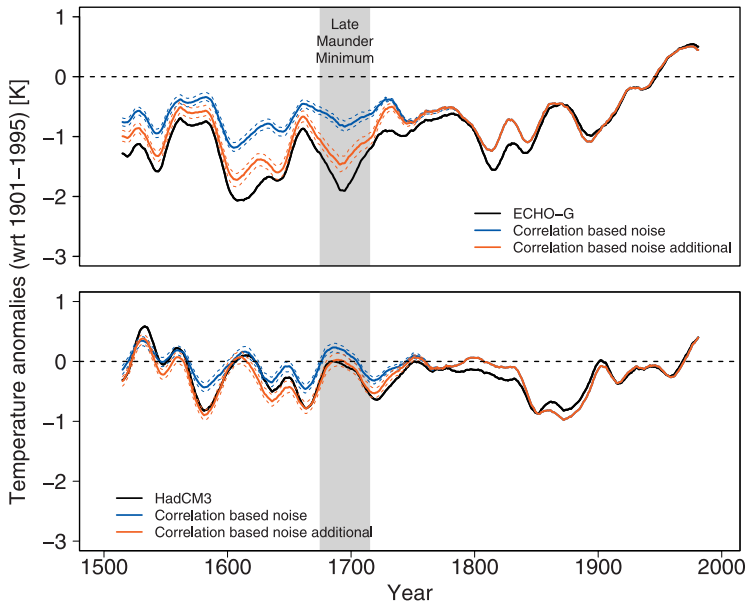

Figure 4. Same as for Figure 3 but showing only the reconstructions based on degraded pseudoproxies according to the correlation of the proxies used by L04 and the instrumental data by *New et al.* [2000] in the 20th century. The red line is based on the same predictor set as the blue line but with one additional predictor over the period 1500–1750 in Scandinavia (region with the largest uncertainties during the early centuries).

controlling the importance of the quality of the predictors. However as 'real world' paleo-reconstructions are over most periods based on spatially and temporally insufficient networks, the predictor quality is the key factor for improvements of reconstruction skill.

[15] It is recommended that systematic pseudoproxy-based testing should become part of every reconstruction, being an important contribution to methodological improvements and understanding of causes of past climate variability in the 'real world.'

[16] **Acknowledgments.** The authors thank the Met Office's Hadley Centre for kindly providing the HadCM3 data and Fidel González-Rouco for his helpful comments. C. Ammann and another reviewer made useful comments and suggestions and helped to improve the quality of the paper. MK, JL, NR, and EX are supported by the Swiss National Science Foundation (NCCR Climate); EX, EZ, and JL were also supported by the EU project SOAP. This publication was supported by the Foundation Marchese Francesco Medici del Vascello.

References

Bradley, R. S. (1996), Are there optimum sites for global paleotemperature reconstruction?, in *Climatic Variations and Forcing Mechanisms of the Last 2000 Years*, *NATO ASI Ser., Ser. 1*, vol. 41, edited by P. D. Jones, R. S. Bradley, and J. Jouzel, pp. 603–624, Springer, New York.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.

Bürger, G., I. Fast, and U. Cubasch (2006), Climate reconstruction by regression: 32 variations on a theme, *Tellus, Ser. A*, *58*(2), 227–235.

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

Luterbacher, J., E. Xoplaki, D. Dietrich, R. Rickli, J. Jacobeit, C. Beck, D. Gyalistras, C. Schmutz, and H. Wanner (2002), Reconstruction of sea level pressure fields over the eastern North Atlantic and Europe back to 1500, *Clim. Dyn.*, *18*(7), 545–561.

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*(5663), 1499–1503.

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

Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann (2005), Testing the fidelity of methods used in proxy-based reconstructions of past climate, *J. Clim.*, *18*, 4097–4107.

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

New, M., M. Hulme, and P. Jones (2000), Representing twentieth-century space-time climate variability. part II: Development of 1901–96 monthly grids of terrestrial surface climate, *J. Clim.*, *13*, 2217–2238.

Osborn, T., S. Raper, and K. Briffa (2006), Simulated climate change during the last 1,000 years: Comparing the ECHO-G general circulation model with the MAGICC simple climate model, *Clim. Dyn.*, *27*(2-3), 185–197.

Pauling, A., J. Luterbacher, and H. Wanner (2003), Evaluation of proxies for European and North Atlantic temperature field reconstructions, *Geophys. Res. Lett.*, *30*(15), 1787, doi:10.1029/2003GL017589.

Rutherford, S., M. E. Mann, T. L. Delworth, and R. J. Stouffer (2003), Climate field reconstruction under stationary and nonstationary forcing, *J. Clim.*, *16*, 462–479.

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

Tett, S. F. B., R. Betts, T. J. Crowley, J. Gregory, T. C. Johns, A. Jones, T. J. Osborn, E. Öström, D. L. Roberts, and M. J. Woodage (2007), The impact of natural and anthropogenic forcings on climate and hydrology since 1550, *Clim. Dyn.*, *28*(1), 3–34.

von Storch, H., E. Zorita, J. M. Jones, Y. Dimitriev, F. González-Rouco, and S. F. B. Tett (2004), Reconstructing past climate from noisy data, *Science*, *306*(5296), 679–682.

von Storch, H., E. Zorita, J. M. Jones, F. González-Rouco, and S. F. B. Tett (2006), Response to comment on "Reconstructing past climate from noisy data", *Science*, *312*(5773), 529.

Wagner, S., and E. Zorita (2005), The influence of volcanic, solar and CO_2 forcing on the temperatures in the Dalton Minimum (1790–1830): A model study, *Clim. Dyn.*, *25*(2-3), 205–218.

Wahl, E. R., and C. M. 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, *Clim. Change*, in press.

Wahl, E. R., D. M. Ritson, and C. M. Ammann (2006), Comment on "Reconstructing past climate from noisy data", *Science*, *312*(5773), 529.

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

Zorita, E., F. González-Rouco, and S. Legutke (2003), Testing the Mann et al. (1998) approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G coupled climate model, *J. Clim.*, *16*, 1378–1390.

M. Küttel, J. Luterbacher, N. Riedwyl, H. Wanner, and E. Xoplaki, NCCR Climate and Institute of Geography, University of Bern, Hallerstrasse 12, CH-3012 Bern, Switzerland. (kuettel@giub.unibe.ch)

E. Zorita, Institute for Coastal Research, GKSS-Research Centre, Max Planck Strasse 1, D-21502 Geesthacht, Germany.