GEOPHYSICAL RESEARCH LETTERS, VOL. 34, L16709, doi:10.1029/2007GL030571, 2007

Click Here for Full Article

Adjustment for proxy number and coherence in a large-scale temperature reconstruction

David Frank,[1] Jan Esper,[1] and Edward R. Cook[2]

Received 3 May 2007; revised 25 June 2007; accepted 23 July 2007; published 29 August 2007.

[1] Proxy records may display fluctuations in climate variability that are artifacts of changing replication and interseries correlation of constituent time-series and also from methodological considerations. These biases obscure the understanding of past climatic variability, including estimation of extremes, differentiation between natural and anthropogenic forcing, and climate model validation. Herein, we evaluate as a case-study, the Esper et al. (2002) extra-tropical millennial-length temperature reconstruction that shows increasing variability back in time. We provide adjustments considering biases at both the site and hemispheric scales. The variance adjusted record shows greatest differences before 1200 when sample replication is quite low. A reduced amplitude of peak warmth during Medieval Times by about 0.4°C (0.2°C) at annual (40-year) timescales slightly re-draws the longer-term evolution of past temperatures. Many other regional and large-scale reconstructions appear to contain variance-related biases. **Citation:** Frank, D., J. Esper, and E. R. Cook (2007), Adjustment for proxy number and coherence in a large-scale temperature reconstruction, *Geophys. Res. Lett.*, *34*, L16709, doi:10.1029/2007GL030571.

1. Introduction

[2] Characterizations of the past and present climate system, that are for example, necessary to differentiate between the effects of natural and anthropogenically induced forcing, require records that are as accurate as possible, and in particular, do not possess time dependent biases that alone could obscure the understanding of the spatial extent and magnitude of warmth during Medieval Times or the occurrences of recent extremes in comparison to pre-industrial conditions. However, the variance of proxy time-series often reveals significant temporal dependence with records showing monotonic increases, decreases, and long-term fluctuations (see Figure S1[1]). What causes these variance changes, be it methodological derivatives, changes in proxy qualities and quantities, or a true reflection of climate, needs to be explored before many attributes of the climate system can faithfully be addressed.

[3] Inherent to paleoclimatology is the fact that the number of relevant datasets does not remain stable through time. Although hundreds of proxy records are available to estimate regional to hemispheric-scale climate change in recent centuries, this number drops to a handful of annually resolved records by the beginning of the past millennium [*Esper et al.*, 2004; *Jones and Mann*, 2004; *Luterbacher et al.*, 2004]. These individual records are in turn of variable quality (typically decreasing) back in time due to changes in sample replication (e.g., in tree-ring records), decreasing resolutions (e.g., ice cores), and the inability to confirm the inferred climatic response prior to the instrumental period in all proxies. These factors result in changing uncertainties and robustness of paleoclimatic data, with the number of series averaged together directly impacting the local variance [*Wigley et al.*, 1984; *Osborn et al.*, 1997]. It is thus critical that proxy and instrumental data analyses – from individual records to large assemblages – account for and minimize potential artifacts that may result as replication and quality vary.

[4] Herein, we explore changes in the local variance – one of the many aspects that reflect and contribute to reconstruction uncertainty. We seek to i) further general awareness to the common attributes of temporal changes in the variance of proxy time-series, ii) explore reasons for variance changes, and iii) develop and advocate methods to help minimize variance artifacts. We focus our analysis on a tree-ring based reconstruction of Northern Hemisphere extra-tropical temperatures [*Esper et al.*, 2002] (hereinafter referred to as ECS). In this analysis, we attempt to identify and minimize the influence of changes in the proxy network that may bias its variance structure – thereby providing a "methodological update". We suggest this update improves ECS through a more detailed consideration of the changes within the dataset.

[5] We first provide background on basic theory and correction procedures to minimize variance biases due to changing sample replication and interseries correlation. The ECS dataset and reconstruction are introduced, followed by a description of and results from the adjustment methods employed, and we close with a discussion.

2. Variance Corrections: Theory and Practice

[6] The variance of the mean of a collection of time-series (S_n^2) is a function of the mean variance of the individual time-series ($\bar{S}_i^2$), their sample replication (n), and mean interseries correlation ($\bar{r}$) [*Wigley et al.*, 1984].

$$S_n^2 = \bar{S}_i^2\left[\frac{\bar{r}(n-1)+1}{n}\right]$$

If any terms do not remain time-stable, the resulting average will possess temporally dependent (possibly artificial) variance changes. Changes in n are, however, inherent in paleoclimatology. The reciprocal of the square-bracketed

[1]Swiss Federal Institute for Forest, Snow, and Landscape Research, Birmensdorf, Switzerland.
[2]Lamont-Doherty Earth Observatory, Palisades, New York, USA.

0094-8276/07/2007GL030571$05.00

[1]Auxiliary materials are available in the HTML. doi:10.1029/2007GL030571.

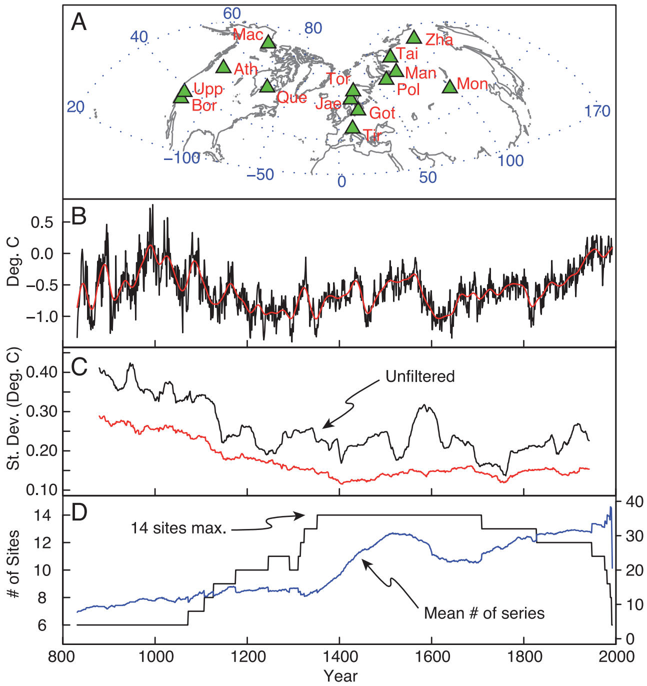

Figure 1. Site locations, reconstruction, and data characteristics for ECS. (a) Map with locations of the 14 sites. (b) ECS reconstruction scaled to annual land-only 20–90°N temperatures over the 1856–1979 period and 40-year splined smoothing. (c) Running STDEV of the ECS dataset for unfiltered and 40-year high-pass filtered data computed in a 100-year moving window. (d) Sample size information including the number of site chronologies and the average number of series per site.

term, referred to as the effective independent sample size, n_{eff}, represents the theoretical number of orthogonal time-series that would provide the same signal as the true (non-orthogonal) dataset. As $\bar{r}$ approaches zero or unity, n_{eff} approaches n or unity, respectively; as n approaches infinity (n_∞), n_{eff} approaches $1/\bar{r}$. *Osborn et al.* [1997] show how the variance can be adjusted for temporal changes in n and $\bar{r}$. This method requires that series are stationary and centered around a mean of zero. Multiplication of a mean time-series by $(n_{eff}/n_\infty)^{1/2} = (\bar{r}{*}n_{eff})^{1/2}$ should result in a series that does not contain variance artifacts related to n_{eff} fluctuations. This, however, depends upon assumptions made when calculating n_{eff}.

[7] Most simply, changes in n are quantified and a time independent estimate of $\bar{r}$ is made, as is, for example, in developing the variance adjusted gridded temperature datasets [*Brohan et al.*, 2006]. However, *Osborn et al.* [1997] additionally discuss how $\bar{r}$ may vary, both as a function of frequency and time. From analyses with tree-ring data, we often noticed significant temporal dependence of $\bar{r}$ at the site (roughly defined as a group of samples from about the same geographical, ecological, and climatic region, generally collected for the same purpose) level, that may result from differences in sample homogeneity in recent and relict wood, a higher percentage of correlations computed between different samples from the same trees during early portions of chronologies, and the influence of mean tree-age upon a chronology's signal.

[8] In producing site chronologies, and their subsequent large-scale mean, consideration of changes in n and $\bar{r}$ were not made in ECS. This has likely biased ECS at two levels – the mean site chronologies and their large-scale mean – thereby impacting the reconstructed course of extra-tropical temperatures.

3. ECS Data Set and Reconstruction

[9] *Esper et al.*'s [2002] reconstruction utilized a collection of 14 tree-ring sites (Figure 1a). We consider the same measurement series after the biological age-trend has been removed *via* Regional Curve Standardization (RCS) [*Briffa et al.*, 1992]. Features of ECS include a warming trend in the past century corresponding to that observed in instrumental data, extended periods of cooler conditions – tending to be more pronounced than those found in many other large-scale reconstructions – reflecting "Little Ice Age" conditions, and high values around 1000 associated with the "Medieval Warm Period" (MWP) (Figure 1b).

[10] Superimposed upon, or embedded within, the reconstructed temperatures is a tendency for increased variability back in time, particularly prior to about 1400 (Figure 1c). The maximum number of site chronologies (14) spans 1352–1708 with sample replication decreasing to ten site chronologies prior to 1246 and six chronologies prior to 1072 (Figure 1d). In addition, there is a general decrease back in time in the number of tree-ring measurement series available at each site. Between 1500 and 1992 the average chronology contains just over 28 series, decreasing to about 14 series over 831–1499.

4. Methods

[11] Following the above descriptions, we applied variance adjustment corrections to all 14 ECS site chronologies (Figure 1a), and their subsequent large-scale average in this two-stage averaging process. We utilize both time dependent estimates of $\bar{r}$, referred to hereafter as "RUNNINGr" adjustments, and a more conventional time independent single estimate of $\bar{r}$ calculated over the full dataset, referred to as "MEANr" adjustments. $\bar{r}$ was calculated in 100-year moving windows for the RUNNINGr adjustment. This approach appears to yield reasonable estimates of temporal variations in $\bar{r}$ found in site data, with tests showing only slight sensitivity to the window size (Figure S3b). In averaging the site chronologies together to form the reconstruction, we only applied the MEANr adjustment due to the relatively low between site $\bar{r}$, for which we assume that most fluctuations are unrelated to changes in underlying data properties. We follow the procedures as in ECS, however, we do not explicitly consider the latitude of sites in the calculation of n_{eff}.

5. Variance Adjusted ECS Reconstruction

[12] Site level variance corrections, as shown in figure S2 for the Tirol dataset, were performed for all 14 of the site chronologies. Results summarizing the variance inflation present in unadjusted site records are shown in Figure 2a. Variance corrections are generally less than 20%, but may rise dramatically during a chronology's earliest portions where sample size decreases to a minima (Figure 1d); a

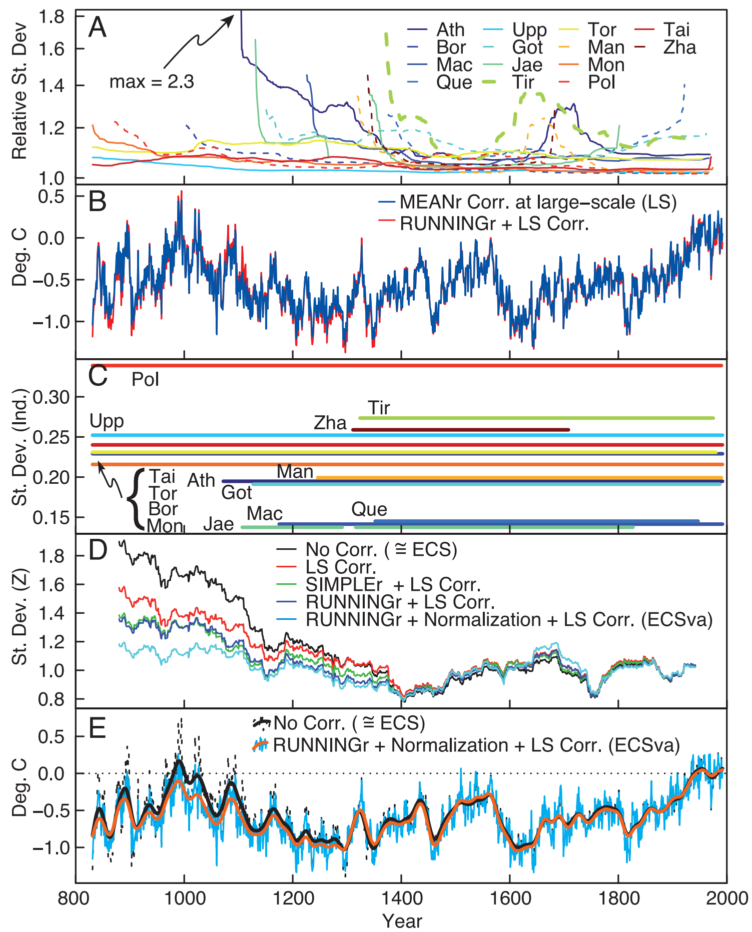

Figure 2. Variance adjustments for ECS dataset. (a) Corrections as in Figure S2e (note log scale). (b) Variance corrected series using MEANr for the large-scale average only and also in addition RUNNINGr at the site level. (c) STDEV of chronologies with horizontal bars showing chronology time-spans. (d) Running 100-year STDEV for various correction stages calculated with 40-year high-pass fractions after 1856–1979 normalization. (e) Variance adjusted ECS record (ECSva) along with the unadjusted mean record. Dashed line shows 1961–1990 anomaly reference period mean. Calibration as in Figure 1b.

two-fold increase is noted for the Athabasca site chronology – the most extreme case. A few records show notable changes between 1600–1700 where replication minima during the transition from living to relict material exist.

[13] We performed two variance adjustments to account for the changes in the ECS dataset. The first only employed the MEANr correction for large-scale averaging; the second employed both the RUNNINGr correction at the site level and the MEANr correction for the large-scale average (Figure 2b). These corrections have greatly reduced the trend towards increased variability in the early portion of ECS (Figures 1b and 1c), and together demonstrate the influence of the two averaging steps: the combination of detrended measurement series to form site chronologies and the combination of site chronologies to form a large-scale composite. Correction for the large-scale averaging was most important and reflects the low $\bar{r}$ values between site chronologies that make their variance highly sensitive to replication changes. However, despite correction at both the site and large-scale levels, an increase in variance back in time is still observed (Figure 2d).

[14] Inspection of the mean variance of the site chronologies as a function of their length showed that the sites that extend towards the beginning of the ECS record tend to have higher variances, so that as the shorter chronologies drop-out beginning around 1300, the remaining chronologies tend to inflate the variance of the mean (Figure 2c). This perhaps includes a species specific component, with the longest chronologies composed of pine and larch and the shorter chronologies, spruce – reported to be more "complacent" [*Schweingruber*, 1996]. Correction for the different mean chronologies levels of variability was achieved by giving the records uniform standard deviations (STDEV) over their common period (1352–1708). This variance normalization (along with the adjustments for changing n and $\bar{r}$) yields a record largely void of significant variance trends present in the ECS record (Figure 2d).

[15] We suggest that the corrections based on mathematical expectation of averaging result in a reconstruction more closely representing climatic variability and its temporal structure than ECS. The variance adjusted record (hereafter, ECSva) that has been corrected at both the site and large-scale levels, with additional adjustment for differences in

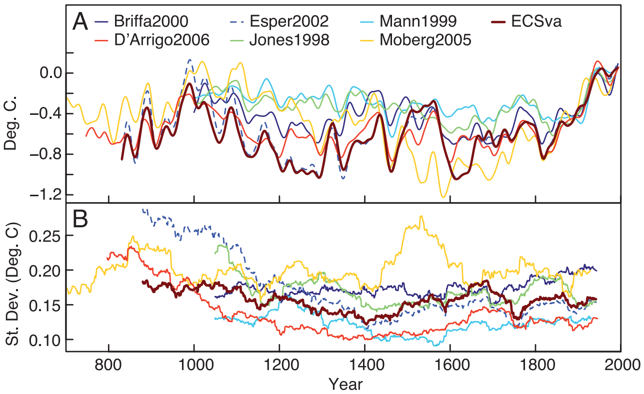

Figure 3. Large-scale temperature reconstructions shown for (a) the 40-year low pass records, and (b) the STDEV of the 40-year high-pass reconstructions in a 100-year moving window. Calibration as in Figure 1b.

the mean chronologies variances, is shown in figure 2e (see also Figure S5). Most changes occur prior to about 1200, when it was already cautioned [*Esper et al.*, 2002; *Cook et al.*, 2004] that the reconstruction's quality is reduced. Adjustment impacts are time dependent with increasing importance back in time. At 1000, the STDEV of ECS is inflated by approximately 40% in comparison to the adjusted ECSva record. About 50% of this variance increase was eliminated by adjusting the large-scale averaging for n variations, 35% for the mean variance level correction, and the remainder from site level changes.

[16] The variance adjustments applied result in subtle yet notable changes. Peak reconstructed temperatures during the MWP are reduced by about 0.4°C at annual and 0.2°C at 40-year timescales, suggesting the 0.3°C estimate of recent warmth exceeding those of the MWP by *Cook et al.* [2004] was conservative. However, due to calibration uncertainties related to the appropriate target season, region and methodology [*Esper et al.*, 2005] absolute temperature estimates should be regarded only as estimates relative to the specific calibration approach.

6. Large-Scale Records

[17] A comparison with other large-scale temperature reconstructions (Figure 3a) indicates an increasingly familiar picture of reconstruction coherence and divergence. However, the general consensus from these records reflects the greatest warmth during the recent century and around 1000. The methodology used in developing these reconstructions varies considerably, but some have specifically included corrections for changes in n at the site and/or the subsequent large-scale levels. For example, *Jones et al.* [1998] used one of the methods suggested by *Osborn et al.* [1997], but do not provide more details. In addition, they normalized the series prior to averaging; this turned out to be an important factor to the variance increase in ECS. To the best of our knowledge, none of these approaches allowed for changes in $\bar{r}$ in their adjustments.

[18] Many of these hemispheric reconstructions display systematic trends in their long-term variance behavior, as shown by the running STDEVs for the high-passed large-scale reconstructions (Figure 3b). The long-term increase is notable in the ECS reconstruction, although perhaps surprisingly, considering their nested approach and a MEANr type adjustment at the site levels, is also observed in *D'Arrigo et al.*'s [2006] reconstruction. Similarly, *Jones et al.* [1998] shows an increase at the beginning of the record. *Mann et al.* [1999] and particularly *Briffa* [2000] tend to show the most stable variability, whereas *Moberg et al.*'s [2005] record is characterized by a notable peak in the 16th century, which occurs when this reconstruction diverges most substantially from the others. While the highest frequency fraction of most reconstructions is not highly correlated with inter-annual temperatures likely due to the poor spatial representativity [*Esper et al.*, 2005; *Cook et al.*, 2004], this overview suggests that many other reconstructions may be subject to similar biases as the original ECS record.

7. Discussion

[19] Changes in variability are common features found in both regional (Figures S1 and 2) and large-scale temperature reconstructions (Figure 3) that affect the estimates of seasonal to annual climatic extremes, impact long-term trends, and affect assessments of natural vs. anthropogenic climatic forcing. The variance adjustment procedures we have applied appear successful at reducing biases in the ECS reconstruction. Following *Osborn et al.* [1997], the methods outlined eliminate variance biases based on mathematical expectation, rather than empirical approaches more prone to eliminate true climatically related changes in variance. Such variance adjustments should be applicable to all proxy archives, including corals, ice cores, and documentary evidence.

[20] The variance adjustment requires the time-series to have a stationary mean centered around zero. *Jones et al.* [2001] in applying these methods to instrumental measurements first 30-year high-pass filtered the data with correlations computed and corrections applied only on this fraction. In contrast to the instrumental series at grid box scales, tree-ring data generally have substantially lower $\bar{r}$ making them more sensitive to changes in n. The ECS data also possess noise at both short and long time scales - with dependence on the spatial region of interest; at large-scales the ECS dataset exclusively possesses signal at wavelengths greater than 20-years [*Cook et al.*, 2004]. To consider the relationships at all of these wavelengths and also average regionally specific variations, we simply set the long term mean of each series to zero prior to corrections. We have assumed that the series are stationary, although if this is not correct, the adjustments performed might underestimate the magnitude of climatic extremes during periods of lowered sample depth, including the MWP.

[21] The RUNNINGr procedure utilized has the advantage of allowing a temporally dependant estimate of $\bar{r}$ that enables adjustment for variance changes related to, for example, age dependence on trees' interseries correlation, or shifts in data characteristics as may occur in records composed of a living tree site with good site control and an older portion composed of material where the site locations are generally less known and more widely distributed. Importantly, in contrast to the more standard MEANr-type adjustments, the RUNNINGr approach is less sensitive to stationarity assumptions, as the individual series are not required to have the same relationship over their full time

period but rather only over the window length. Examples shown with individual tree-ring chronologies, indicate the utility of the RUNNINGr adjustments in this regard. Potential RUNNINGr weaknesses include sensitivity in estimating $\bar{r}$ to window length, end effects in its computation, and the possibility of random statistical or climatically induced fluctuations.

[22] These results and additional examples (Figure S1) suggest that changes in the local variance in climatic time-series have a variety of sources – unfortunately implying that there is no panacea to produce records free of variance biases, and leading to additional challenges in the study of extreme event probabilities or anthropogenic signatures of climatic forcing. Since the most suitable solution seems to be careful inspection of records to find potential underlying causes for (artificial) changes in variance, we do not advocate a priori employment of empirical approaches to equalize the variability. We also suggest that truncation of chronologies below a minimum replication threshold, of say 5 series, would minimize many biases related to changes in n (but not $\bar{r}$).

[23] *Osborn et al.* [1997] discussed that for large-spatial scales $\bar{r}$ may be different for low and high frequency components. Based on pseudo-proxy experiments, *Brohan et al.* [2006] determined their variance adjustments to be appropriate at the grid box scale, but slightly biased when grid boxes are averaged over hemispheric scales. Both of these studies suggest uncertainties in how variance should optimally be adjusted at different spatial scales, for different data types, and at lower frequencies.

[24] Even though the ECSva reconstruction falls within the ECS error bands, improvement of the central tendency and the shape of the reconstruction has likely been achieved by utilizing methods herein that reduce artifacts from changes in n and $\bar{r}$. However, despite any subtle improvements of the ECSva record over ECS, we emphasize that new and updated regional chronologies are key in producing more skillful large-scale reconstructions. Further methodological understanding and refinements will also contribute to these goals.

[25] Comparison of the ECSva record with other hemispheric-scale temperature reconstructions shows that the more extreme nature of the ECS series in the peak values around 1000 are reduced, whereas during the cool 17th centuries the ECS dataset remains largely unaltered and approaching the upper limit - along with borehole records [e.g., *Huang et al.*, 2000] - in estimating the temperature amplitude over the past 500 years [*Esper et al.*, 2005]. Future proxy reconstructions at the local to hemispheric scales should seek to identify and eliminate variance biases, while applying methods to retain potential changes in the actual variability [*Frank et al.*, 2005] so that robust assessments and attributions of natural and anthropogenic influences on the climate system can be derived.

[26] **Acknowledgments.** We thank Olivier Bouriaud, Ulf Büntgen, Jürg Luterbacher, Rob Wilson, and Elena Xoplaki and two anonymous reviewers for helpful suggestions and comments. Work was funded by the Swiss National Science Foundation (grant 2100-066628, NCCR-Climate) and the EU-project Millennium. Lamont-Doherty Earth Observatory Contribution 7051.

References

Briffa, K. R. (2000), Annual climate variability in the Holocene: Interpreting the message of ancient trees, *Quat. Sci. Rev.*, *19*, 87–105.

Briffa, K. R., P. D. Jones, T. S. Bartholin, D. Eckstein, F. H. Schweingruber, W. Karlen, P. Zetterberg, and M. Eronen (1992), Fennoscandian summers from AD 500: Temperature changes on short and long timescales, *Clim. Dyn.*, *7*, 111–119.

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

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

D'Arrigo, R., R. Wilson, and G. Jacoby (2006), On the long-term context for late twentieth century warming, *J. Geophys. Res.*, *111*, D03103, doi:10.1029/2005JD006352.

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

Esper, J., D. C. Frank, and R. J. S. Wilson (2004), Climate Reconstructions: Low-Frequency Ambition and High-Frequency Ratification, *Eos Trans. AGU*, *85*(12), 113.

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

Frank, D., R. Wilson, and J. Esper (2005), Synchronous variability changes in alpine temperature and tree-ring data over the last two centuries, *Boreas*, *34*, 498–505.

Huang, S. P., H. N. Pollack, and P. Y. Shen (2000), Temperature trends over the past five centuries reconstructed from borehole temperatures, *Nature*, *403*, 756–758.

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

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

Jones, P. D., T. J. Osborn, K. R. Briffa, C. K. Folland, E. B. Horton, L. V. Alexander, D. E. Parker, and N. A. Rayner (2001), Adjusting for sampling density in grid box land and ocean surface temperature time series, *J. Geophys. Res.*, *106*, 3371–3380.

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

Mann, M. E., R. S. Bradley, and M. K. Hughes (1999), Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophys. Res. Lett.*, *26*, 759–762.

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

Osborn, T. J., K. R. Briffa, and P. D. Jones (1997), Adjusting variance for sample-size in tree-ring chronologies and other regional-mean time-series, *Dendrochronologia*, *15*, 89–99.

Schweingruber, F. H. (1996), *Tree Rings and Environment: Dendroecology*, Haupt: Bern.

Wigley, T. M. L., K. R. Briffa, and P. D. Jones (1984), On the average of correlated time series, with applications in dendroclimatology and hydrometeorology, *J. Clim. Appl. Meteorol.*, *23*, 201–213.

E. R. Cook, Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades NY, 10964, USA.

J. Esper and D. Frank, Swiss Federal Research Institute WSL, Zuercherstrasse 111, CH-8903 Birmensdorf, Switzerland. (frank@wsl.ch)