

# 1 Ranking of tree-ring based temperature reconstructions of the past millennium

2  
3 Jan Esper <sup>a,\*</sup>, Paul J. Krusic <sup>b,c</sup>, Fredrik Ljungqvist <sup>d,e</sup>, Jürg Luterbacher <sup>f,g</sup>, Marco Carrer <sup>h</sup>, Ed Cook <sup>i</sup>,  
4 Nicole K. Davi <sup>h,j</sup>, Claudia Hartl-Meier <sup>a</sup>, Alexander Kirilyanov <sup>k,l</sup>, Oliver Konter <sup>a</sup>,  
5 Vladimir Myglan <sup>m</sup>, Mauri Timonen <sup>n</sup>, Kerstin Treydte <sup>o</sup>, Valerie Trouet <sup>p</sup>,  
6 Ricardo Villalba <sup>q</sup>, Bao Yang <sup>r</sup>, Ulf Büntgen <sup>o</sup>  
7

8 <sup>a</sup> Department of Geography, Johannes Gutenberg University, 55099 Mainz, Germany

9 <sup>b</sup> Department of Physical Geography and Quaternary Geology, Stockholm University, 10691  
10 Stockholm, Sweden

11 <sup>c</sup> Navarino Environmental Observatory, Messinia, Greece

12 <sup>d</sup> Department of History, Stockholm University, 10691 Stockholm, Sweden

13 <sup>e</sup> Bolin Centre for Climate Research, Stockholm University, 10691 Stockholm, Sweden

14 <sup>f</sup> Department of Geography, Justus-Liebig University, 35390 Giessen, Germany

15 <sup>g</sup> Centre for International Development and Environmental Research, Justus Liebig University  
16 Giessen, 35390 Giessen, Germany

17 <sup>h</sup> Università degli Studi di Padova, Dipartimento Territorio e Sistemi AgroForestali, 35020 Legnaro,  
18 Italia

19 <sup>i</sup> Tree Ring Laboratory, Lamont-Doherty Earth Observatory, Palisades, NY 10964, USA

20 <sup>j</sup> Department of Environmental Science, William Paterson University, Wayne, NJ 07470, USA

21 <sup>k</sup> Institute of Forest SB RAS, Akademgorodok, Krasnoyarsk, 660036, Russia

22 <sup>l</sup> Laboratory of Ecosystem Biogeochemistry, Siberian Federal University, Krasnoyarsk, 660041,  
23 Russia

24 <sup>m</sup> Institute for the Humanities, Siberian Federal University, Krasnoyarsk, 660041, Russia

25 <sup>n</sup> Natural Resources Institute Finland (Luke), Rovaniemi Unit, Rovaniemi, Finland

26 <sup>o</sup> Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland

27 <sup>p</sup> Laboratory of Tree-Ring Research, University of Arizona, Tucson, AZ 85721, USA

28 <sup>q</sup> Instituto Argentino de Nivología, Glaciología y Ciencias Ambientales, CCT-Mendoza, 5500  
29 Mendoza, Argentina

30 <sup>r</sup> Key Laboratory of Desert and Desertification, Cold and Arid Regions Environmental and  
31 Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China  
32  
33

34 Submitted as a **Review Paper** to *Quaternary Science Reviews*

35  
36  
37 \* Corresponding author: [esper@uni-mainz.de](mailto:esper@uni-mainz.de), Tel. +49 6131 3922296

1 **Abstract**

2 Tree-ring chronologies are widely used to reconstruct high- to low-frequency variations in growing  
3 season temperatures over centuries to millennia. The relevance of these timeseries in large-scale climate  
4 reconstructions is often determined by the strength of their correlation against instrumental temperature  
5 data. However, this single criterion ignores several important quantitative and qualitative characteristics  
6 of tree-ring chronologies. Those characteristics are (i) *data homogeneity*, (ii) *sample replication*, (iii)  
7 *growth coherence*, (iv) *chronology development*, and (v) *climate signal*. A new reconstruction-scoring  
8 scheme, based on these 5 characteristics, is designed and applied to 39 published, millennial-length  
9 temperature reconstructions from Asia, Europe, North America, and the Southern Hemisphere. By  
10 providing a more comprehensive set of criteria by which to evaluate tree-ring chronologies we hope to  
11 improve the development of large-scale temperature reconstructions spanning the past millennium. All  
12 the reconstructions and their corresponding scores are provided at [www.url-here.de](http://www.url-here.de).

13  
14

15 **Keywords**

16 Paleoclimate, Climate change, Proxy data, Dendrochronology, Dendroclimatology

17

## 1 **1. Introduction**

2 Tree-ring chronologies (TRCs) are an important source of information in large-scale temperature  
3 reconstructions (IPCC 2013, St. George 2014). The latter are used to estimate temperature variability at  
4 continental (Euro-Med 2k consortium 2015, Pages 2k consortium 2013, Trouet et al. 2013), hemispheric  
5 (Christiansen and Ljungqvist 2012, D'Arrigo et al. 2006, Esper et al. 2002a, Ljungqvist 2010, Ljungqvist  
6 et al. 2012, Mann et al. 2008, Schneider et al. 2015, Shi et al. 2013) and global scales (Mann and Jones  
7 2003) over the past 1000 years, enabling comparisons between climate variations during pre-industrial  
8 and industrial periods. The importance of TRCs in these reconstructions arises from the precise annual  
9 dating inherent to this proxy (Douglass 1941) and our mechanistic understanding of the influence of  
10 temperature on tree growth (Fritts 1976). The relative significance of tree-ring series increases back in  
11 time, as the overall number of annually resolved proxies rapidly declines towards the early centuries of  
12 the past millennium (Esper et al. 2004).

13  
14 TRCs are typically composed of tree-ring width (TRW) or maximum latewood density (MXD)  
15 measurement series from many trees (Fritts 1976). A TRC might extend back over the entire past  
16 millennium if one or more individual trees is 1000 years or more in age. Such longevity, however, is  
17 rare and restricted to only a few locations (OldList at: [www.rmtrr.org/oldlist.htm](http://www.rmtrr.org/oldlist.htm)). Most millennium-  
18 length TRCs are therefore produced by combining samples from living trees with older material from  
19 archeological and historical structures (hereafter: historical samples), dead wood on the ground (remnant  
20 samples), or wood buried under ground or preserved in lakes (sub-fossil samples). The successful  
21 combination of samples from living trees with historical/remnant/sub-fossil samples improves when the  
22 provenance of all samples is ecologically consistent. If not, sections of a millennium-length chronology  
23 can have different growth rates and climate signals than those sections dominated by samples from  
24 living trees (Tegel et al. 2010, Linderholm et al. 2014). For example, remnant samples from a sub-alpine  
25 site in the Alps are ideally combined with samples from living trees growing on the same slope, at the  
26 same elevation and aspect (Salzer et al. 2014b). Sub-fossil trees from a shallow lake are ideally  
27 combined with information from living trees surrounding the lake (Düthorn et al. 2013, 2015).

28  
29 Combining samples from living trees with historical/remnant/sub-fossil samples is not always  
30 straightforward. Habitat homogeneity in a TRC derived from living trees and in-situ remnant or sub-  
31 fossil wood from the same location may be high, but their combination with historical material can be  
32 more complicated. If, for example, the historical samples were obtained from an old building in a  
33 mountain valley, it often remains unclear from which position in the surrounding forests the samples  
34 were originally collected (Büntgen et al. 2006b). It is not uncommon that historical structures,  
35 particularly in alpine environments, contain recycled material of unknown origin, as a consequence of  
36 repairs and additions (Bellwald 2000, Kalbermatten and Kalbermatten 1997). Without detailed  
37 construction histories the researchers ability to trace the origin of samples is limited (Büntgen et al.  
38 2005, Wilson et al. 2004). The situation is further complicated if the samples combined in a TRC are  
39 from multiple locations spread over a large region, and if this region extends over several hundreds of  
40 kilometers. These problems, affecting the *Homogeneity* of a tree-ring dataset, are seemingly reduced in  
41 chronologies from only living trees, sampled at a single site.

42  
43 Another important characteristic of millennium-length TRCs includes the number and temporal  
44 distribution of TRW (or MXD) measurement series averaged in the mean chronology. Varying sample  
45 replication is often reported when describing a new TRC, but is usually disregarded in large-scale  
46 temperature reconstructions. Typically, the number of measurement series included in a TRC declines  
47 back in time, and might change from more than 100 living-tree samples in the 20th century to only a  
48 handful of samples (perhaps from a single historical structure) at the beginning of the last millennium.  
49 Acknowledging the effects of changing sample size by calculating temporally varying uncertainty  
50 estimates, is not usually considered outside the tree-ring community (IPCC 2013). However, this  
51 characteristic is important as the relevance an individual TRC in large-scale proxy networks is  
52 commonly based on the strength of instrumental calibration of only the well-replicated 20th century  
53 data, thereby overlooking any pre-instrumental replication changes.

54  
55 Similarly, the coherence among the TRW (MXD) series combined in a TRC, and temporal change  
56 thereof, is not considered in the non-dendrochronological literature (Frank et al 2007). The inter-series

1 correlation among TRW series is an important characteristic of a mean chronology, and is commonly  
2 computed to evaluate the temporal changes in the chronology's signal strength (Fritts 1976). It is rarely  
3 stable and can change, for example, at (i) the transition from living trees to series from  
4 historical/remnant/sub-fossil material, or (ii) from a cluster of measurement series of a certain building  
5 to another building, or (iii) by the proportion of juvenile, mature, and adult growth rings (Cook and  
6 Kairiukstis 1990). Gradual trends in inter-series correlation, as well as step changes, are common in  
7 long TRCs and bare important information on the reliability of dendroclimatic reconstructions during  
8 pre-instrumental periods. Measures that assess the affect of changing sample size and inter-series  
9 correlation include the *Expressed Population Signal* and *Subsample Signal Strength* (Wigley et al.  
10 1984). However, these metrics are not widely recognized beyond the tree-ring community, and their  
11 combination with other uncertainties, e.g. from the unexplained variance of the calibration model or the  
12 choice of the detrending model, remains challenging (Esper et al. 2007).

13  
14 Another important TRC characteristic is the degree to which a chronology retains the full spectrum of  
15 pre-instrumental temperature variance, which is affected by the method used in chronology  
16 development, and the age-structure of the underlying data (Cook et al. 1995). Recent assessments of  
17 large datasets showed that instrumental meteorological measurements and tree-ring timeseries contain  
18 different frequency spectra (Bunde et al. 2013, Büntgen et al. 2015, Franke et al. 2013, Zhang et al.  
19 2015), and that TRCs are limited in capturing millennial scale temperature trends (Esper et al. 2012b).  
20 To minimize the loss of long-term information, dendrochronologists apply detrending techniques that  
21 are specifically designed to preserve low frequency variance. The preferred approach is the Regional  
22 Curve Standardization (RCS) method, introduced to dendroclimatology by Briffa et al. (1992).  
23 However, RCS demands a large number of TRW (MXD) measurement series, and requires the  
24 underlying data to represent a combination of short segments (trees) distributed more or less evenly  
25 throughout the entire chronology (Esper et al. 2003a). For example, if a TRC is composed of only very  
26 old living trees, the chronology's biological age will steadily increases towards the present. This causes  
27 the biologically younger rings to be concentrated at the beginning of the past millennium, and the older  
28 rings in the modern period. This age structure limits the comparison of tree-rings of the same age over  
29 time, which is the backbone of RCS and related tree-ring detrending techniques (Melvin and Briffa  
30 2008).

31  
32 The basic characteristics of TRC *Homogeneity*, *Replication*, *Growth Coherence*, and *Chronology*  
33 *Development* are well known to dendroclimatologists. However, they are not usually recognized in the  
34 multi-proxy paleoclimate community and rarely, if ever, considered in large-scale temperature  
35 reconstructions derived from these data (IPCC 2013, Pages 2k consortium 2013). Promoting these  
36 characteristics, and stimulating their consideration in addition to the classical calibration against  
37 instrumental climate data, is the main objective of this article. We review the characteristics of a total  
38 of 39 published temperature reconstructions reaching back to AD 1000 and use the results to rank the  
39 timeseries.

## 40 41 42 **2. Data and Methods**

### 43 *2.1 Temperature reconstructions*

44 A survey of tree-ring based temperature reconstructions, reaching back with a minimum replication of  
45 three TRW (or MXD) measurement series to AD 1000, returned 39 records (Table 1). Fourteen records  
46 are from Asia, 13 from Europe, 8 from North America, and 4 from the Southern Hemisphere (SH). The  
47 reconstructions are not evenly distributed over the hemispheres, but are clustered in Fennoscandia, the  
48 European Alps, northern Siberia, high Asia, the Rocky Mountains, southwestern US, southern South  
49 America, and Tasmania/New Zealand. The TRCs are located in regions characterized by different  
50 summer warming trends over the past 100 years (see the colored areas in Fig. 1). Compared to the  
51 Northern Hemisphere (NH), the SH is clearly underrepresented with only four records.

52  
53 There are precedents of long TRCs with an inferred climate signal (e.g. LaMarche 1973, 1974), but the  
54 first instrumentally calibrated, millennium-length record is the summer (previous-year December to  
55 current-year February; pDec-Feb) temperature reconstruction from very old (living) *Fitzroya*  
56 *cupressoides* growing in the Rio Alerce valley in southern Argentina (Villalba 1990; Table 1). Other

1 records developed at that time were later updated by including new measurement series and/or  
2 reprocessed using new methods. A good example is the Torneträsk MXD chronology that was originally  
3 developed in the 1980s (Schweingruber et al. 1988), calibrated and reprocessed in the early 1990s (Briffa  
4 et al. 1990, 1992), updated in the early 2000s (Grudd 2008), and recently again updated and reprocessed  
5 (Melvin et al. 2013). In those instances where there are multiple versions of a reconstruction, we cite  
6 the most recently published account as it contains references to all previous work.

7  
8 The millennium-length temperature reconstructions are derived from various conifer species  
9 representing nine genera, with *Pinus* (n=14 records) and *Larix* (n=12) being most common. Seven  
10 reconstructions, including the early *Fitzroya cupressoides* record from Argentina (Villalba 1990), are  
11 produced from only living trees, whereas the majority of chronologies (n=32) are composed of tree-ring  
12 series from living trees combined with series from historical samples (e.g. Lötschental TRC from  
13 Switzerland; Büntgen et al. 2006a), remnant samples (e.g. Polar Ural TRC from Russia; Briffa et al.  
14 2013), and sub-fossil samples (e.g. Oroko Swamp TRC in New Zealand, Cook et al. 2002). Some of  
15 these chronologies are composed of samples collected in well-constrained, ecologically homogeneous  
16 (Schweingruber 1996) sites (e.g. Dzehlo in Russia, Myglan et al. 2012b), whereas others combine data  
17 from different sites (e.g. Yamal in Russia, Briffa et al. 2013), and even from several valleys within a  
18 larger region (e.g. Karakorum in Pakistan; Esper et al. 2002b).

19  
20 All TRCs included in this survey have either been calibrated against instrumental climate data and  
21 transferred into temperature units, or interpreted by the original authors as a temperature proxy.  
22 Interestingly, the different methods used to transfer TRW and MXD data into temperature units (Briffa  
23 et al. 1983, Cook et al. 1994, Esper et al. 2005) resulted in vastly different reconstructed temperature  
24 ranges, varying by only a few tenths to several degrees Celsius over the past millennium (see the varying  
25 ranges of the thin black curves in Figure 2; see Esper et al. 2012a for a regional example). Also, the  
26 season of maximum response to temperature (e.g. June-August, May-September, etc.) and the  
27 reconstructed climate target (e.g. mean, maximum, and minimum temperature) differ among the records  
28 (last column in Table 1).

29  
30 Surprisingly, despite these differences in (i) location and regional 20th century temperature trends (Fig.  
31 1), (ii) species composition and sample sources (historical/remnant/sub-fossil), (iii) seasonality of the  
32 temperature signal, and (iv) transfer technique and reconstructed variance; the simple arithmetic mean  
33 of each "continent" (acknowledging that the records do not spatially represent NH continents) coheres  
34 astonishingly well over the past 1000 years (Fig. 2e). Correlations range from  $r = 0.42$  between Asia  
35 and North America to  $r = 0.48$  between Europe and Asia, and increase at decadal resolution to 0.66  
36 (Asia/N-America) and 0.82 (Europe/Asia). This large-scale coherence indicates that some common  
37 external forcing affects this dendrochronological network (Fernández-Donado et al. 2013, Pages 2K  
38 PMIP3 group 2015) and confirms the paleoclimatic significance of tree-ring data over the past  
39 millennium.

## 40 41 2.2 TRC characteristics and metrics

42 In this section, we describe the five basic TRC characteristics *Data Homogeneity* (2.2.1), *Sample*  
43 *Replication* (2.2.2), *Growth Coherence* (2.2.3), *Chronology Development* (2.2.4), and *Climate Signal*  
44 (2.2.5), commonly used by dendrochronologists to evaluate a chronology for climate reconstruction,  
45 and explain how statistical metrics of these characteristics are used in an ordinal scoring scheme that is  
46 understandable to non-specialists. In those instances when raw TRW and MXD data were publically  
47 available or contributed by the authors (raw data at: [www.url-here.se](http://www.url-here.se)), we have re-calculated the metrics  
48 of interest. When the original cross-dated measurements were not available (see last column in Table  
49 7), we have estimated their scores based on information provided in the original articles. Such estimates  
50 are highlighted in red in the tables that follow. The calibration scores, resulting from the TRC's  
51 correlation against temperature data (2.2.5 *Climate signal*), were taken from the original articles. In the  
52 instances where no measure of calibration was detailed in the original article, we used nearby gridded  
53 data to provide an estimate of climate calibration.

54  
55 For each characteristic (2.2.1 to 2.2.5) we used the ordinal scoring scheme to rank the reconstructions.  
56 To aid reconstruction comparison, results of the TRC scores are stratified into four groups: class-A

1 (highlighted in green in Tables 2-7), class-B (light green), class-C (light blue), and class-D (blue).  
2 Except for the first characteristic (2.2.1 *Data Homogeneity*), we highlighted the ten top-ranked TRCs in  
3 green (ranks 1-10), the TRCs ranking 11-20 in light green, the TRCs ranking 21-30 in light blue, and  
4 the TRCs ranking 31-39 in blue. This systematic was changed with *Data Homogeneity* (5 green, 9 light  
5 green, 16 light blue, 9 blue) to account for the larger number intermediate TRCs. The individual ranks  
6 for each characteristic (2.2.1 to 2.2.5) were finally summed into an overall score.

### 7 8 2.2.1 *Data Homogeneity*

9 Of the five characteristics introduced here, *Data Homogeneity* is the most descriptive, i.e. it is based on  
10 a combination of qualitative traits rather than quantitative measures. *Homogeneity* integrates  
11 information on the (i) source of wood samples, (ii) type of chronology, (iii) number of species, (iv)  
12 temporal clustering, and (v) a remark (results shown in Table 2). "Source" includes information on the  
13 origin of wood samples and the number of sampling sites. We use "Sub-fossil" for samples from lakes,  
14 bogs, etc., "Remnant" for dead wood on the ground, and "Historic" for samples from old buildings  
15 archaeological structures. The *Homogeneity* score also considers whether the samples originate from  
16 one, several, or multiple sites, as far as this information could be obtained from the original publication  
17 or via personal communication with the authors. "Chronology type" differentiates between "C" for  
18 records composed of living plus relict (sub-fossil/remnant/historical) material, and "L" for records  
19 composed of samples from only living trees. The "Number of Species" in a TRC is typically one, but  
20 occasionally may be two. "Temporal clustering" refers to cases where the contribution of data from  
21 distinct homogeneous sites dominates specific periods of the past 1000 years (a condition that might  
22 require the application of multiple RCS runs, Melvin et al. 2013). Finally, we included a "Remark"  
23 section summarizing specific features that are relevant to the *Homogeneity* score in support of the  
24 reconstruction's ranking.

### 25 26 2.2.2 *Sample Replication*

27 The temporal distribution of TRW (or MXD) measurement series in the reconstructions differs  
28 dramatically over the past millennium (Fig. 3). These changes are considered in the second metric by  
29 combining information on (i) mean replication, (ii) maximum replication, (iii) minimum replication, and  
30 (iv) the 11th/20th century ratio of measurement numbers. "Mean replication" is the average number of  
31 measurement series (core samples or radii from disks) over the last millennium, considering all years  
32 from AD 1000 to the most recent year of a reconstruction. "Maximum replication" and "Minimum  
33 replication" refer to the maximum and minimum numbers of measurement series, which are typically  
34 reached in the modern and the early periods of a reconstruction, respectively (see the black curves in  
35 Fig. 3). The "11th/20th century ratio" acknowledges this exemplary replication curve shape, as well as  
36 its significance in the reconstruction: all TRCs are calibrated over the well-replicated 20th century, but  
37 the reconstruction period extends back to the, often weakly replicated, 11th century. The metric equals  
38 the mean 11th century replication, divided by the mean 20th century replication, multiplied by 100. To  
39 produce the final *Replication* score, the first three values are summed ( $i+ii+iii$ ), and the resulting sum  
40 multiplied by (iv). If the reconstruction is produced using MXD data the *Replication* score ( $(i+ii+iii)*iv$ )  
41 is multiplied by 2 to account for MXD's increased signal strength and higher production costs. Note that  
42 these choices, as well as those described below for the other TRC characteristics, are not statistically  
43 validated, but made with the intention of combining descriptive measures commonly used in  
44 dendrochronology into an ordinary scoring system that can be used to compare and rank reconstructions.

### 45 46 2.2.3 *Growth Coherence*

47 Another important characteristic influencing the temporally changing skill of tree-ring based climate  
48 reconstructions, is the correlation between the TRW (MXD) measurement series (Frank et al. 2007,  
49 Osborn et al. 1997, Wigley et al. 1984). For those reconstructions where the raw data are available, we  
50 calculated the inter-series correlation (abbreviated "Rbar" in the dendrochronological literature; Cook  
51 and Kairiukstis 1990) for 100-year segments, sliding in 10-year steps along the chronology (Fig. 4). The  
52 resulting timeseries reveal substantial differences among the reconstructions (the black curves in Fig.  
53 4), as well as a minor tendency towards reduced values back in time, particularly in some of the records  
54 from Europe and Asia. These characteristics are considered here in the *Growth Coherence* score by  
55 summing the (i) average inter-series correlation over the past millennium (mean Rbar), (ii) maximum  
56 inter-series correlation in a single 100-year period (max. Rbar), and (iii) minimum inter-series

1 correlation in a single 100-year segment (min. Rbar). The sum (*i+ii+iii*) is multiplied by (*iv*) the  
2 11th/20th century Rbar ratio (in %).

#### 3 4 2.2.4 Chronology development

5 A key component in the process of building a TRC is the detrending method used to remove tree-age  
6 related growth trends from the raw measurement series (Bräker 1981, Cook and Kairiukstis 1990, Cook  
7 et al. 1995). As detailed above, application of RCS (Esper et al. 2003a) is currently accepted as the  
8 preferred method to preserve low frequency variance in TRCs. We acknowledge this view by (*i*)  
9 assigning TRCs produced using RCS a “1”, and TRCs produced using individual-series detrending  
10 methods (e.g. ratios from negative exponential curves or smoothing splines) a “2” (Cook and Peters  
11 1997). However, RCS only works well if the underlying measurement series are derived from a  
12 composite of living and relict trees, ideally including young and old tree-rings evenly distributed  
13 throughout the past millennium (Esper et al. 2014). TRCs that are composed this way are characterized  
14 by age curves that are nearly horizontal over the past 1000 years (Fig. 5). In practice this is rarely the  
15 case. The age curves of some TRCs composed of very old living trees in, for example, North America  
16 and Asia are particularly steep (increasing-age towards present). In contrast, in Europe, where the  
17 majority of reconstructions are derived from composite chronologies of historical and living-tree  
18 samples, the mean age curves are relatively flat (the blue curve in Fig. 5e).

19  
20 We score these attributes by considering (*ii*) the maximum difference between the highest and lowest  
21 value in the age curve over the past millennium, and (*iii*) the slope of a linear regression fit to the age  
22 curve. We further consider (*iv*) the maximum retained low frequency information, ranging from multi-  
23 centennial = 1, to centennial = 2, to decadal = 3. For the final *Chronology Development* score we  
24 multiply (*i*) the method score (1 for RCS, 2 for individual-series detrending), with (*ii*) the (square root  
25 of the) max.-min. age difference, (*iii*) the (absolute) slope of the linear regression (times 100), and (*iv*)  
26 the maximum retained low frequency score (1 to 3, for multi-centennial, centennial, and decadal).

#### 27 28 2.2.5 Climate signal

29 This final score considers some of the classic metrics used in paleoclimatic research, such as the  
30 correlation against monthly instrumental temperature data, averaged over the season of maximum  
31 response (see the last column in Table 1). However, as the period of overlap between instrumental and  
32 proxy data varies considerably among the reconstructions – largely due to the lengths of observational  
33 data available to researchers – we score *Climate Signal* by (*i*) the square root of the number of years of  
34 overlap between the TRC and instrumental record, multiplied by the residual between, (*ii*) the  
35 correlation against climate data and (*iii*) a split calibration/verification difference. The latter metric is a  
36 standard criterion in dendroclimatology used to benchmark the temporal robustness of the relationship  
37 between proxy and instrumental data (Cook and Kairiukstis 1990). However, the split  
38 calibration/verification differences are not always reported. In those instances, we estimated the split  
39 calibration/verification difference based on our calculations using gridded temperature data. Finally, we  
40 include an additional adjustment (*iv*) to account for a calibration period that was intentionally shortened  
41 to avoid potential divergence issues (for details see Büntgen et al. 2008, D’Arrigo et al. 2008, Esper and  
42 Frank 2009, Esper et al. 2010, Wilson et al. 2007). If such problems are reported in the original article,  
43 and the calibration period was truncated, we used 0.5 as a multiplier (1 if no such problem was detected).  
44 The final *Climate Signal* score was derived by: square root *i* \* (*ii* - *iii*) \* *iv*.

### 45 46 47 3. Results and Discussion

#### 48 3.1 Overall TRC ranking

49 Our assessment of 39 millennial-length TRCs’ *Homogeneity*, *Replication*, *Growth Coherence*,  
50 *Chronology Development*, and *Climate Signal* is presented in Tables 2-6. The final ranking (Table 7),  
51 derived from the sum of these metrics, reveals that no reconstruction consistently dominates in the top  
52 group (class-A, dark green dots in the tables) in all five categories. Four records (N-Scan, E-Canada,  
53 Finland, Dzhelo) score high (class-A or class-B) in four out of the five categories, and one record  
54 (Yamal) scores high in three. However, each of these, overall best-ranked reconstructions, scores less  
55 well (class-C: light blue dot) on at least one criterion, mostly *Homogeneity* (four records).

1 There are ten records (W-Himalaya, Tatra, Karakorum, Great Basin, S-Finland, Tien Shan, Jämtland,  
2 Wulan, Gulf of Alaska, French Alps) with weak scores (class-C or class-D) in four metrics. W-Himalaya  
3 is the only reconstruction scoring in class-D in four: *Homogeneity*, *Replication*, *Growth Coherence*, and  
4 *Chronology Development*. This low scoring, combined with the class-B rank in *Climate Signal*, places  
5 the W-Himalaya reconstruction at the bottom of table 7. However, the raw data are not available for this  
6 reconstruction (see the last column) and several scores had to be estimated (highlighted in red in tables  
7 3-6). The same is true for several other reconstructions, and it seems advisable to emphasize the coarse  
8 categorization into four classes (A to D) rather than the precise ordering in our tables.

### 9 10 3.2 Detailed TRC rankings

11 Four reconstructions (Dzhelo, Tasmania, Rio Alerce, Qamdo, Mongolia) rank at the top in *Homogeneity*  
12 (Table 2). The data used in these TRCs include samples from living trees, as well as remnant and sub-  
13 fossil material from a single site or valley (with one exception; Rio Alerce includes only living trees).  
14 These top-ranked records are followed by a group of nine reconstructions that were sampled from  
15 slightly less homogeneous conditions, including data at moist and dry micro-sites, from different  
16 elevations, and measured using different techniques, for example. Despite their less than ideal  
17 *Homogeneity* score, these reconstructions are still more homogenous compared to a number of TRCs  
18 (n=16) that integrate data from multiple sites in different valleys, regions, and/or elevations (light blue  
19 dots in Table 2). Such large differences in habitat can introduce substantial growth rate variations that  
20 are difficult to differentiate from long-term temperature variations. In addition, the climate signal might  
21 change between samples from different elevations and micro-sites. These potential biases are likely  
22 most severe in the nine TRCs ranking last (dark blue dots in Table 2). The two end members (Karakorum  
23 and W-Himalaya) are produced using living trees only, sampled from multiple sites, located in different  
24 valleys and at different elevations, with distances up to 100 km between sites, and including two tree  
25 species (Karakorum). Clearly these TRCs contain a less homogeneous sample composition compared  
26 to the top-ranked records that include samples from only one, well-constrained site.

27  
28 The reconstructions scoring well in *Homogeneity* are not necessarily top-ranked in *Replication* (Table  
29 3). To appear in the top group in *Replication*, it is necessary not only to include a large number of TRW  
30 or MXD measurement series, but have these samples evenly distributed throughout the past millennium.  
31 Bumps from very high to very low replications in certain periods, as well as large differences between  
32 20th and 11th century replications, result in a lower score. Among the records doing well in *Replication*  
33 are two TRCs from Central Asia (Mongun and Dzhelo) and one from New Zealand (Oroko Swamp).  
34 These records score particularly well in the 11th/20th century ratio, reaching values >100%. Other  
35 reconstructions, such as the Alps (larch) and Swiss/Austrian Alps TRCs include many samples (530 and  
36 253 respectively over the past millennium), but contain a dramatic replication decline from the 20th to  
37 the 11th centuries (down to 2% and 25%, respectively), limiting the skill of these timeseries in the early  
38 period of the past millennium. The TRCs scoring weakest in *Replication* (Tatra, Boreal Plateau, Rio  
39 Alerce, Tien Shan) are characterized by low minimum replications ( $n \leq 5$  series) and small 11th/20th  
40 century ratios ( $\leq 15\%$ ). These records might perform well when calibrated against 20th century  
41 instrumental temperature data, but there is considerable risk that this 20th century skill does not persist  
42 over the past millennium, simply because the number of samples changes dramatically back in time.

43  
44 Since more than the sheer number of measurement series is important, we also considered the  
45 reconstructions' inter-series correlations (Table 4). The three TRCs scoring best in this category  
46 (Indigirka, Yamal, Taimyr) are all located in northern Siberia, where growth variations among trees are  
47 synchronized by harsh climatic conditions during a rather short growing season. These top-ranked  
48 records are characterized by inter-series correlations that do not fall below  $R_{bar}=0.20$  at any time over  
49 the past millennium (minimum correlation in Table 4) and reach values >100% in their 11th/20th century  
50 ratio. Other mid-ranked TRCs, such as Polar Ural (class-B) and Jämtland (class-C), display either a very  
51 low minimum  $R_{bar}$  values (-0.20 in Polar Ural) or substantially decreasing  $R_{bar}$  values from the 20th  
52 century back to the 11th century (42% in Jämtland). Another interesting example of a class-C TRC is  
53 Oroko Swamp, which is characterized by only minor  $R_{bar}$  changes back in time (92%), but an overall  
54 low mean inter-series correlation ( $R_{bar}=0.18$ ). Finally, the TRCs scoring weakest (Tatra, S-Finland,  
55 Central Alps) are characterized by severe correlation declines, down to  $\leq 20\%$  back in the 11th century,  
56 and either a low mean  $R_{bar}$  values (0.20 estimated for S-Finland and Central Alps) or negative minimum



1 Rbar values (-0.07 in Tatra). In these cases it seems advisable to anticipate substantial changes in the  
2 chronologies' signal strength over the past millennium, as the coherence among their constituent  
3 measurement series is extremely variable. If the inter-series correlation drops significantly, reductions  
4 in TRC variance, and a tendency towards the long-term mean are to be expected.

5  
6 The three top-ranked reconstructions in the *Chronology Development* category are all from Northern  
7 Europe (N-Scan, Finland, Torneträsk (TRW)), followed by records from the Alps (Lötschental) and  
8 Canada (E-Canada) (Table 5). These reconstructions, as well as the other class-A and class-B TRCs  
9 (green and light green in Table 5, total n=20), are all composed of a mixture of living trees and  
10 historical/remnant/sub-fossil samples, facilitating the application of RCS for optimal conservation of  
11 low frequency variance (Autin et al. 2015, Briffa et al. 1992, Esper et al. 2003a). The top-scoring  
12 Northern European records are, however, additionally characterized by small age ranges (<110 years)  
13 and only minor (positive and negative) linear trends in the mean age curves. The top-ranked N-Scan  
14 record is reported to contain millennial scale temperature variance (Esper et al. 2012b), a feature also  
15 seen in the Taimyr reconstruction from Northern Siberia. The subsequent mid-ranked TRCs are  
16 characterized by age ranges from ~150-300 years, as well as linear trend angles ranging from ~5-30  
17 degrees. Some class-C records were standardized using individual detrending methods, including the  
18 Swiss/Austrian Alps, Lauenen, and Mongolia reconstructions, an approach more commonly found in  
19 the TRCs towards the bottom of table 5. The application of individual detrending methods has been  
20 shown to systematically limit the low frequency variance retained in TRCs (Cook et al. 1995). This  
21 limitation is reflected in the maximum frequency metric included here, indicating that six  
22 reconstructions (Rio Alerce, Wulan, Gulf of Alaska, Mongun, S-Chile, Lauenen) maximally retain  
23 decadal scale temperature variance. These records, as well as some of the individually detrended TRCs,  
24 should not be used with the objective of reconstructing the full spectrum of temperature variance over  
25 the past millennium (e.g. Mann et al. 2008).

26  
27 By comparison to *Homogeneity*, *Replication*, *Growth Coherence*, and *Chronology Development*,  
28 measures of climate signal strength are widely recognized in the paleoclimatic community. However, a  
29 good correlation between tree-ring proxy and instrumental temperature data alone is a fairly incomplete  
30 description of reconstruction skill. For example, if a TRC includes many more samples during the 20th  
31 century (*Replication* metric), or the samples originate from different valleys (*Homogeneity*), or the mean  
32 age curve declines severely back in time (*Chronology Development*), the 20th century calibration  
33 statistics provide little information about the signal strength over past centuries. That being said, we  
34 here assess climate signal strength based on the length of the calibration period, the correlation strength  
35 with instrumental data, the calibration/verification difference and any, seemingly arbitrary, truncation  
36 of the calibration period.

37  
38 The reconstructions scoring best for *Climate Signal* are all from regions where sufficient instrumental  
39 data are available to calibrate over periods of 100 years and longer (Table 6). The three top-ranked  
40 records (Torneträsk (MXD), N-Scan, Alps (larch)) all correlate at  $\geq 0.70$  against instrumental  
41 temperature data, with only minor differences (<0.10) between calibration and verification periods.  
42 Other reconstructions, with calibration period correlations  $\geq 0.70$ , albeit over shorter periods (53 years  
43 in Qamdo, 57 years in Taimyr), contain larger calibration/verification differences (0.18 in Taimyr) and  
44 appear in class-B. These reconstructions certainly meet the criteria for a successful TRC calibration, but  
45 they may contain a marginally verifiable climate signal. This is either because the  
46 calibration/verification differences are large (e.g. 0.63 in Qilian), the calibration period was truncated  
47 due to some inconsistency (e.g. Tatra, see the fourth column in Table 6), or the overall correlation is  
48 low (e.g. 0.17 in Upper Wright Lakes). However, a weak calibration result does not necessarily mean  
49 that a TRC contains no climate signal, but might indicate that the instrumental station record is too short  
50 (Esper et al. 2010), of poor quality (Böhm et al. 2001, 2010, Parker et al. 1994), or too remote (Cook et  
51 al. 2013).

52  
53 Perhaps a good example, highlighting the importance of using several categories to evaluate a TRC, is  
54 the case of the Alps (larch) record. The Alps TRC correlates well ( $r=0.70$ ) over 140 years of regional  
55 instrumental temperatures, and thus ranks #3 in the *Climate Signal* metric (calibration/verification  
56 difference is 0.07, calibration period not truncated). However, these calibration statistics were obtained

1 over the period 1864-2003 during which the TRC's mean replication is 1379 series. Concurrently, the  
2 average number of TRW series in the 11th century reaches only 22, which produces an 11th/20th century  
3 ratio of 2% (see Table 3). Though certainly an extreme example, it nicely demonstrates how a large-  
4 scale reconstruction produced focusing on robust 20th century climate signals, can result in an  
5 overestimation of statistical skill over the past millennium.  
6

### 7 *3.3 Ranking implications*

8 Over the past decades a number of statistically valid methods have been developed to describe TRC's  
9 signal strength. Examples include the *Expressed Population Signal* (Wigley et al. 1984), bootstrap  
10 confidence intervals (Briffa et al. 1992), ensemble calibration technique (Frank et al. 2010), and reduced  
11 sample calibration trials (Esper et al. 2012b). All of these dendro-specific statistics help estimate the  
12 temporally varying skill of tree-ring based climate reconstructions, but the methods are largely  
13 inapplicable to other proxy archives, and are not used in large-scale, multi-proxy reconstructions (Pages  
14 2k consortium 2013).  
15

16 By providing an assessment and ranking of the TRCs, we attempt to bridge the gap between the tree-  
17 ring, modeling, and multi-proxy communities. While some of the scores and metrics used here have not  
18 been rigorously validated, we believe that the development of an intuitive ranking system that can be  
19 universally applied to all TRCs will foster the judicious use of tree-ring data in large-scale  
20 reconstructions. For example, if NH temperature variability during medieval times is of interest, it is not  
21 meaningful to include TRCs with only a few samples during the 11th century, i.e. researchers might  
22 want to avoid reconstructions with low *Replication* scores (Table 3). Similarly, if the full spectrum of  
23 past temperature variability is of interest, one might want to include only those TRCs retaining  
24 centennial to millennial scale variance, i.e. exclude records with low *Chronology Development* scores  
25 (Table 5).  
26

27 These arguments lead to a list of recommendations:

- 28 R1 Avoid integrating TRCs that emphasize decadal scale variance (scoring low on "Maximum  
29 frequency") when intending to reconstruct low frequency (centennial to millennial scale)  
30 temperature variance.
- 31 R2 Avoid overrating TRCs that average many TRW (MXD) measurement series in the 20th century,  
32 but only few series at the start.
- 33 R3 Pay attention to the sample composition of millennium-length TRCs. Do data sources change  
34 through time (different sites, buildings, valleys, etc.)?
- 35 R4 Consider TRC replication and Rbar changes when interpreting tree-ring based climate  
36 reconstructions.
- 37 R5 Differentiate between composite TRCs that integrate data from varying sources  
38 (living/remnant/historical/sub-fossil) and TRCs that integrate data from only old living trees, and  
39 acknowledge potential biases due to changing tree ages over the past millennium.
- 40 R6 Do not only focus on the calibration statistics from comparisons with instrumental climate data, as  
41 this perspective can give the false impression that reconstruction skill persists throughout the past  
42 millennium.  
43

44 We acknowledge that some of the metrics presented here contain partly redundant information, e.g.  
45 lower replication or reduced Rbar values typically result in weaker correlations with instrumental  
46 climate. There are also other TRC characteristics that could be used to assess tree-ring based temperature  
47 reconstructions, though these appeared difficult to quantify with simple metrics. Examples include the  
48 TRC serial correlation (Meko 1981) and climate signal after trend removal (von Storch et al. 2004). For  
49 instance, an assessment of serial correlation in both tree-ring and instrumental temperature data might  
50 reveal a larger lag-1 autocorrelation in a TRC (likely due to biological memory effects; Esper et al.  
51 2015), potentially indicating a coherence deficiency and reduced skill of a long-term climate  
52 reconstruction. Similarly, an assessment of the climate signal after removing low frequency variance  
53 (e.g. increasing 20th century temperature trend), from the instrumental and proxy data, increases the  
54 degrees of freedom of the calibration statistics and supports the estimation of signal strength in the high  
55 frequency domain. However, correctly evaluating these properties in a large network of millennium-

1 length TRCs, including several records for which the underlying measurement data are not available, is  
2 not feasible.

3  
4 Our review clearly indicates that solely focusing on the calibration statistics overlooks a number of  
5 additional, important characteristics inherent to tree-ring based climate reconstructions. When  
6 evaluating large TRC networks it is important to keep in mind that the 20th century instrumental data  
7 (i) contain gaps, breakpoints, and biases (Hinkel et al. 2003, Landsberg 1981, Oke 2007), (ii) are of  
8 substantially varying length depending on the study region (e.g. in Europe versus central Asia; Cook et  
9 al. 2013), and (iii) are recorded at greatly differing distances from the tree-ring sampling sites. The  
10 suitability of a station record is additionally influenced by the topography (flat or mountainous), the  
11 elevational difference between tree and station sites, and regional synoptic weather patterns. The use of  
12 gridded climate data does not necessarily overcome these shortcomings as they rely on the same (Jones  
13 et al. 1999) or even fewer (Krusic et al. 2015) station data.

14  
15 Finally, it is important to acknowledge the seasonality of the climate signal (typically related to the  
16 regional growing season, Table 1) and question the use of a seasonally restricted reconstruction, as  
17 reported in the original literature, in attempting to reconstruct mean annual temperatures (Mann et al.  
18 1999, 2008). While we acknowledge that the calibration/verification statistics (Table 6), as well as  
19 means by which the chronology is transformed to measures of climate are also important (e.g. Bürger  
20 et al. 2006, Christiansen 2011, Christiansen et al. 2009, Esper et al. 2005, Juckes et al. 2007, Lee et al.  
21 2008, Smerdon et al. 2011, 2015, Rutherford et al. 2005, von Storch et al. 2004, Zorita et al. 2003), we  
22 believe that a revised consideration of basic TRC characteristics as well as the development of new  
23 TRCs is key to improving the extensive network of large-scale reconstructions and our understanding  
24 of past climate variability.

#### 25 26 27 **4. Conclusions**

28 Thirty-nine millennium-length temperature reconstructions are ranked based on a rating scheme that  
29 considers basic TRC characteristics commonly considered by dendrochronologists. The TRC  
30 characteristics were grouped into five composite scores: *Homogeneity*, *Replication*, *Growth Coherence*,  
31 *Chronology Development*, and *Climate Signal*. It is argued that consideration of these characteristics,  
32 beyond the tree-ring community, will improve the development of large-scale temperature  
33 reconstructions that utilize tree-ring data from different regions and continents. Similarly, the rankings  
34 produced for each score will support this objective, as they facilitate the selection process of TRCs when  
35 addressing paleoclimatic objectives. For example, researchers might not want to include TRCs resting  
36 on only a few trees during the 11th century, in a study addressing the magnitude and spatial extent of  
37 warmth during medieval times. This, and other recommendations are expressed towards the end of this  
38 review paper.

39  
40 A systematic comparison of TRC characteristics permitted ranking the 39 millennium-length  
41 temperature reconstructions into four groups (class-A to class-D) for each of the five metrics. No  
42 reconstruction scores top in all five metrics, but each record has its particular strengths and weaknesses.  
43 Nevertheless, there are some reconstructions that overall perform better than others. These include N-  
44 Scan and Finland from Europe; E-Canada from North America; Yamal and Dzehlo from Asia.  
45 Reconstructions performing less well include W-Himalaya and Karakorum from Asia; Tatra and S-  
46 Finland from Europe; and Great Basin from North America. The rankings presented here can be used  
47 to select and exclude particular records for producing hemispheric scale reconstructions. The fact that  
48 some of the records appear more often towards the end of a ranking table does not mean they cannot be  
49 used for climate reconstruction purposes, but indicates that the users of these data need to be aware of  
50 potential weaknesses that may inadvertently affect their experiment. This review of millennial-length  
51 TRCs will be updated as new reconstructions are produced. Updates will be published online at:  
52 [www.url-here.se](http://www.url-here.se)

#### 53 54 **Acknowledgements**

1 We thank all the tree-ring data producers for sharing their chronologies and measurement series. JL  
2 acknowledges the German Science Foundation project “Attribution of forced and internal Chinese climate  
3 variability in the Common Era”.

4  
5

## 6 **References**

- 7 Autin, J., Gennaretti, F., Arseneault, D., Bégin, Y., 2015. Biases in RCS tree ring chronologies due to  
8 sampling heights of trees. *Dendrochronologia* 36, 13–22.
- 9 Bellwald, I., 2000. Der Rote Segensonntag 1900. Der Dorfbrand von Wiler. Ein Rückblick aus dem  
10 Jahre 2000. *Gem. Wiler, Kippel*.
- 11 Böhm, R., Auer, I., Brunetti, M., Maugeri, M., Nanni, T., Schöner, W., 2001. Regional temperature  
12 variability in the European Alps: 1760-1998 from homogenized instrumental time series. *Int. J.*  
13 *Climatol.* 21, 1779–1801.
- 14 Böhm, R., Jones, P.D., Hiebl, J., Frank, D., Brunetti, M., Maugeri, M., 2010. The early instrumental  
15 warm-bias: a solution for long central European temperature series 1760–2007. *Clim. Change* 101, 41–  
16 67.
- 17 Bräker, O.U., 1981. Der Alterstrend bei Jahrringdichten und Jahrringbreiten von Nadelhölzern und sein  
18 Ausgleich. *Mitteil. Forstl. Bundesversuchsanst. Wien* 142, 75–102.
- 19 Briffa, K.R., Jones, P.D., Wigley, T.M.L., Pilcher, J.R., Baillie, M.G.L., 1983. Climate reconstruction  
20 from tree rings: Part 1, basic methodology and preliminary results for England. *J. Climatol.* 3, 233–242.
- 21 Briffa, K.R., Bartholin, T.S., Eckstein, D., Jones, P.D., Karlén, W., Schweingruber F.H., Zetterberg, P.,  
22 1990. A 1,400-year tree-ring record of summer temperatures in Fennoscandia. *Nature* 346, 434–439.
- 23 Briffa, K.R., Jones, P.D., Bartholin, T.S., Eckstein, D., Schweingruber, F.H., Karlen, W., Zetterberg, P.,  
24 Eronen, M., 1992. Fennoscandian summers from AD 500: temperature changes on short and long  
25 timescales. *Clim. Dyn.* 7, 111–119.
- 26 Briffa, K.R., Shishov, V.V., Melvin, T.M., Vaganov, E.A., Grudd, H., Hantemirov, R.M., Eronen, M.,  
27 Naurzbaev, M.M., 2008. Trends in recent temperature and radial tree growth spanning 2000 years across  
28 northwest Eurasia. *Philosoph. Trans. Royal Soc. B* 363, 2269–2282.
- 29 Briffa, K.R., Melvin, T.M., Osborn, T.J., Hantemirov, R.M., Kirilyanov, A.V., Mazepa, V.S., Shiyatov,  
30 S.G., Esper, J., 2013. Reassessing the evidence for tree-growth and inferred temperature change during  
31 the Common Era in Yamalia, Northwest Siberia. *Quat. Sci. Rev.* 72, 83–107.
- 32 Büntgen, U., Esper, J., Frank, D.C., Nicolussi, K., Schmidhalter, M., 2005. A 1052-year tree-ring proxy  
33 for Alpine summer temperatures. *Clim. Dyn.* 25, 141–153.
- 34 Büntgen, U., Frank, D.C., Nievergelt, D., Esper, J., 2006a. Summer temperature variations in the  
35 European Alps, A.D. 755-2004. *J. Clim.* 19, 5606–5623.
- 36 Büntgen, U., Bellwald, I., Kalbermatten, H., Schmidhalter, M., Frank, D.C., Freund, H., Bellwald, W.,  
37 Neuwirth, B., Nüsser, M., Esper, J., 2006b. 700 years of settlement and building history in the  
38 Lötschental/Switzerland. *Erdkunde* 60, 96–112.
- 39 Büntgen, U., Frank, D.C., Wilson, R.J.S., Carrer, M., Urbinati, C., Esper, J., 2008. Testing for tree-ring  
40 divergence in the European Alps. *Glob. Change Biol.* 14, 2243–2453.
- 41 Büntgen, U., Frank, D., Carrer, M., Urbinati, C., Esper, J., 2009. Improving Alpine summer temperature  
42 reconstructions by increasing sample size. *Trace* 7, 36–43.
- 43 Büntgen, U., Tegel, W., Nicolussi, K., McCormick, M., Frank, D., Trouet, V., Kaplan, J.O., Herzig, F.,  
44 Heussner, K.U., Wanner, H., Luterbacher, J., Esper, J., 2011. 2500 years Of European climate variability  
45 and human susceptibility. *Science* 331, 578–582.
- 46 Büntgen, U., Neuschwander, T., Frank, D., Esper, J., 2012. Fading temperature sensitivity of Alpine  
47 tree growth at its Mediterranean margin and associated effects on large-scale climate reconstructions.  
48 *Clim. Change* 114, 651–666.
- 49 Büntgen, U., Kyncl, T., Ginzler, C., Jaks, D.S., Esper, J., Tegel, W., Heussner, K.U., Kyncl, J., 2013.  
50 Filling the Eastern European gap in millennium-length temperature reconstructions. *Proc. Nat. Acad.*  
51 *Sci.* 5, 1773–1778.

- 1 Büntgen, U., Trnka, M., Krusic, P.J., Kyncl, T., Kyncl, J., Luterbacher, J., Zorita, E., Ljungqvist, F.C.,  
2 Auer, I., Konter, O., Schneider, L., Tegel, W., Štěpánek, P., Brönnimann, S., Hellmann, L., Nievergelt,  
3 D., Esper, J., 2015. Tree-ring amplification of the early nineteenth-century summer cooling in central  
4 Europe. *J. Clim.* 28, 5272–5288.
- 5 Bürger, G., Fast, I., Cubasch, U., 2006. Climate reconstruction by regression—32 variations on a theme.  
6 *Tellus* 58, 227–235.
- 7 Bunde, A., Büntgen, U., Ludescher, J., Luterbacher, J., von Storch, H., 2013. Is there memory in  
8 precipitation? *Nat. Clim. Change* 3, 174–175.
- 9 Christiansen, B., 2011. Reconstructing the NH mean temperature: Can underestimation of trends and  
10 variability be avoided? *J. Clim.* 24, 674–692.
- 11 Christiansen, B., Schmith, T., Thejll P., 2009. A surrogate ensemble study of climate reconstruction  
12 methods: Stochasticity and robustness. *J. Clim.* 22, 951–976.
- 13 Christiansen, B., Ljungqvist, F.C., 2012. The extra-tropical Northern Hemisphere temperature in the last  
14 two millennia: reconstructions of low-frequency variability. *Clim. Past* 8, 765–786.
- 15 Cook, E.R., Kairiukstis, L.A., 1990. *Methods of Dendrochronology – Applications in the Environmental*  
16 *Science*. Kluwer, Dordrecht.
- 17 Cook, E.R., Briffa, K.R., Jones, P.D., 1994. Spatial regression methods in dendroclimatology: a review  
18 and comparison of two techniques. *Int. J. Climatol* 14, 379–402.
- 19 Cook, E.R., Briffa, K.R., Meko, D.M., Graybill, D.A., Funkhouser, G., 1995. The ‘segment-length  
20 curse’ in long tree-ring chronology development for palaeoclimatic studies. *Holocene* 5, 229–237.
- 21 Cook, E.R., Peters, K., 1997. Calculating unbiased tree-ring indices for the study of climatic and  
22 environmental change. *Holocene* 7, 361–370.
- 23 Cook, E.R., Buckley, B.M., D’Arrigo, R.D., Peterson, M.J., 2000. Warm-season temperatures since  
24 1600BC reconstructed from Tasmanian tree rings and their relationship to large-scale sea surface  
25 temperature anomalies. *Clim. Dyn.* 16, 79–91.
- 26 Cook E.R., Palmer, J.G., Cook, B.I., Hogg, A., D’Arrigo, R.D., 2002. A multi-millennial palaeoclimatic  
27 resource from *Lagarostrobos colensoi* tree-rings at Oroko Swamp, New Zealand. *Glob. Plan. Change*  
28 33, 209–220.
- 29 Cook, E.R., Krusic, P.J., Anchukaitis, K.J., Buckley, B.M., Nakatsuka, T., Sano, M., Pages Asia2k  
30 Members, 2013. Tree-ring reconstructed summer temperature anomalies for temperate East Asia since  
31 800 CE. *Clim. Dyn.* 41, 2957–2972.
- 32 D’Arrigo, R.D., Jacoby, G., Frank, D., Pederson, N., Cook, E.R., Buckley, B.M., Nachin, B., Mijiddorj,  
33 R., Dugarjav, C., 2001. 1738 years of Mongolian temperature variability inferred from a tree-ring width  
34 chronology of Siberian pine. *Geophys. Res. Lett.* 28, 543–546.
- 35 D’Arrigo, R., Wilson, R., Jacoby, G. 2006. On the long-term context for late 20th century warming. *J.*  
36 *Geophys. Res.* 111, D03103, doi: 10.1029/2005JD006352.
- 37 D’Arrigo, R.D., Wilson, R., Liepert, B., Cherubini, P., 2008. On the ‘divergence problem’ in northern  
38 forests: a review of the tree-ring evidence and possible causes. *Global Planet. Change* 60, 289–305.
- 39 Douglass, A.E., 1941. Crossdating in dendrochronology. *J. Forestry* 39, 825–832.
- 40 DÜthorn, E., Holzkämper, S., Timonen, M., Esper, J., 2013. Influence of micro-site conditions on tree-  
41 ring climate signals and trends in Central and Northern Sweden. *Trees* 27, 1395–1404.
- 42 DÜthorn, E., Schneider, L., Konter, O., Schön, P., Timonen, M., Esper, J., 2015. On the hidden  
43 significance of differing micro-sites in dendroclimatology. *Silva Fennica* 49, doi: org/10.14214/sf.1220.
- 44 Esper, J., Cook, E.R., Schweingruber, F.H., 2002a. Low-frequency signals in long tree-ring  
45 chronologies for reconstructing of past temperature variability. *Science* 295, 2250–2253.
- 46 Esper, J., Schweingruber, F.H., Winiger, M. 2002b. 1,300 years of climate history for Western Central  
47 Asia inferred from tree-rings. *Holocene* 12, 267–277.
- 48 Esper, J., Cook, E.R., Krusic, P.J., Peters, K., Schweingruber, F.H., 2003a. Tests of the RCS method for  
49 preserving low-frequency variability in long tree-ring chronologies. *Tree-Ring Res.* 59, 81–98.

- 1 Esper, J., Shiyatov, S.G., Mazepa, V.S., Wilson, R.J.S., Graybill, D.A., Funkhouser, G., 2003b.  
2 Temperature-sensitive Tien Shan tree-ring chronologies show multi-centennial growth trends. *Clim.*  
3 *Dyn.* 8, 699–706.
- 4 Esper, J., Frank, D.C., Wilson, R.J.S., 2004. Climate reconstructions: low frequency ambition and high  
5 frequency ratification. *EOS* 85, 113–130.
- 6 Esper, J., Frank, D.C., Wilson, R.J.S., Briffa, K.R., 2005. Effect of scaling and regression on  
7 reconstructed temperature amplitude for the past millennium. *Geophys. Res. Lett.* 32, doi:  
8 10.1029/2004GL021236.
- 9 Esper, J., Frank, D.C., Büntgen, U., Verstege, A., Luterbacher, J., Xoplaki, E., 2007. Long-term drought  
10 severity variations in Morocco. *Geophys. Res. Lett.* 34, doi: 10.1029/2007GL030844.
- 11 Esper, J., Frank, D., 2009. Divergence pitfalls in tree-ring research. *Clim. Change* 94, 261–266.
- 12 Esper, J., Frank, D., Büntgen, U., Verstege, A., Hantemirov, R.M., Kirilyanov, A.V., 2010. Trends and  
13 uncertainties in Siberian indicators of 20th century warming. *Glob. Change Biol.* 16, 386–398.
- 14 Esper, J., Büntgen, U., Timonen, M., Frank, D.C., 2012a. Variability and extremes of Northern  
15 Scandinavian summer temperatures over the past two millennia. *Glob. Plan. Change* 88-89, 1-9.
- 16 Esper, J., Frank, D.C., Timonen, M., Zorita, E., Wilson, R.J.S., Luterbacher, J., Holzkämper, S., Fischer,  
17 N., Wagner, S., Nievergelt, D., Verstege, A., Büntgen U., 2012b. Orbital forcing of tree-ring data. *Nat.*  
18 *Clim. Change* 2, 862–866.
- 19 Esper, J., Dũthorn, E., Krusic, P., Timonen, M., Büntgen, U., 2014. Northern European summer  
20 temperature variations over the Common Era from integrated tree-ring density records. *J. Quat. Sci.* 29,  
21 487–494.
- 22 Esper, J., Schneider, L., Smerdon, J., Schõne, B., Büntgen, U., 2015. Signals and memory in tree-ring  
23 width and density data. *Dendrochronologia* 35, 62–70.
- 24 Euro-Med 2k consortium, 2015. European summer temperatures since Roman times. *Environ. Res.*  
25 *Lett.*, in press.
- 26 Fernández-Donado, L., González-Rouco, J.F., Raible, C.C., Ammann, C.M., Barriopedro, D., García-  
27 Bustamante, E., Jungclauss, J.H., Lorenz, S.J., Luterbacher, J., Phipps, S.J., Servonnat, J.,  
28 Swingedouw, D., Tett, S.F.B., Wagner, S., Yiou, P., Zorita, E., 2013. Large-scale temperature  
29 response to external forcing in simulations and reconstructions of the last millennium. *Clim. Past* 9,  
30 393–421.
- 31 Franke, J., Frank, D., Raible, C., Esper, J., Brõnnimann, S., 2013. Spectral biases in tree-ring climate  
32 proxies. *Nat. Clim. Change* 3, 1–5.
- 33 Frank, D., Esper, J., Cook E.R., 2007. Adjustment for proxy number and coherence in a large-scale  
34 temperature reconstruction. *Geophys. Res. Lett.* 34, doi: 10.1029/2007GL030571.
- 35 Frank, D.C., Esper, J., Raible, C.C., Büntgen, U., Trouet, V., Joos, F., 2010. Ensemble reconstruction  
36 constraints of the global carbon cycle sensitivity to climate. *Nature* 463, 527–530.
- 37 Fritts, H.C., 1976. *Tree Rings and Climate*. Academic press, London.
- 38 Gennaretti, F., Arseneault, D., Nicault, A., Perreault, L., Bégin, Y., 2014. Volcano-induced regime  
39 shifts in millennial tree-ring chronologies from northeastern North America. *Proceed. Nat. Acad. Sci.*  
40 111, 10077–10082.
- 41 Graumlich, L.J., 1993. A 1000-year record of temperature and precipitation in the Sierra Nevada.  
42 *Quat. Res.*, 39, 249–255.
- 43 Grudd, H., 2008. Tornetråsk tree-ring width and density AD 500–2005: a test of climatic sensitivity  
44 and a new 1500-year reconstruction of north Fennoscandian summers. *Clim. Dyn.* 31, 843–857.
- 45 Helama, S., Fauria, M.M., Mielikäinen, K., Timonen, M., Eronen, M., 2010. Sub-Milankovitch solar  
46 forcing of past climates: mid and late Holocene perspectives. *GSA Bulletin*; 122, 1981–1988.
- 47 Helama, S., Vartiainen, M., Holopainen, J., Mäkelä, H.M., Kolström, T., Meriläinen, J. 2014. A  
48 palaeotemperature record for the Finnish Lakeland based on microdensitometric variations in tree  
49 rings. *Geochronometria* 41, 265–277.

- 1 Hinkel, K.M., Nelson, F.E., Klene, A.E., Bell, J.H., 2003. The urban heat island in winter at Barrow,  
2 Alaska. *Int. J. Climatol.* 23, 1889–1905.
- 3 IPCC, 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to*  
4 *the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge  
5 University Press, Cambridge.
- 6 Jones, P.D., New, M., Parker, D.E., Martin, S., Rigor, I.G., 1999. Surface air temperature and its  
7 variations over the last 150 years. *Rev. Geophys.* 37, 173–199.
- 8 Juckes, M.N., Allen, M.R., Briffa, K.R., Esper, J., Hegerl, G.C., Moberg, A., Osborn, T.J., Weber,  
9 S.L., 2007. Millennial temperature reconstruction intercomparison and evaluation. *Clim. Past* 3, 591–  
10 609.
- 11 Kalbermatten, H., Kalbermatten L., 1997. *Blatten. Was alte Menschen, alte Häuser und Schriften*  
12 *erzählen.* Druckerei Bloch, Arlesheim.
- 13 Krusic, P.J., Cook, E.R., Dukpa, D., Putnam, A.E., Rupper, S., Schaefer, J., 2015. 638 years of summer  
14 temperature variability over the Bhutanese Himalaya. *Geophys. Res. Lett.* 42, doi:  
15 10.1002/2015GL063566.
- 16 LaMarche, V.C. Jr., 1973. Holocene climatic variations inferred from treeline fluctuations in the  
17 White Mountains, California. *Quat Res* 3, 632–660.
- 18 LaMarche, V.C. Jr., 1974. Paleoclimatic inferences from long tree-ring records. *Science* 183, 1043–  
19 1048.
- 20 Landsberg, H.E., 1981. *The urban climate.* Academic press, London.
- 21 Lara, A., Villalba, R., 1993. A 3620-year temperature record from *Fitzroya cupressoides* tree rings in  
22 southern South America. *Science* 260, 1104–1106.
- 23 Lee, T.C.K., Zwiers, F.W., Tsao, M., 2008. Evaluation of proxy-based millennial reconstruction  
24 methods. *Clim. Dyn.* 31, 263–281.
- 25 Linderholm, H.W., Gunnarson, B.E., 2005. Summer temperature variability in central Scandinavia in  
26 the last 3600 years. *Geogr. Ann.* 87, 231–241.
- 27 Linderholm, H.W., Zhang, P., Gunnarson, B.E., Björklund, J., Farahat, E., Fuentes, M., Rocha, E.,  
28 Salo, R., Seftigen, K., Stridbeck, P. and Liu, Y., 2014. Growth dynamics of tree-line and lake-shore  
29 Scots pine (*Pinus sylvestris* L.) in the central Scandinavian Mountains during the Medieval Climate  
30 Anomaly and the early Little Ice Age. *Fron. Ecol. Evol.* 2, doi: 10.3389/fevo.2014.00020.
- 31 Liu, Y., An, Z., Linderholm, H.W., Chen, D., Song, H., Cai, Q., Sun, J., Tian, H., 2009. Annual  
32 temperatures during the last 2485 years in the mid-eastern Tibetan Plateau inferred from tree rings.  
33 *Sci. China Series D: Earth Sci.* 52, 348–359.
- 34 Ljungqvist, F.C., 2010. A new reconstruction of temperature variability in the extra-tropical Northern  
35 Hemisphere during the last two millennia. *Geogr. Ann.* 92, 339–351.
- 36 Ljungqvist, F.C., Krusic, P.J., Brattström, G., Sundqvist, H.S., 2012. Northern Hemisphere temperature  
37 patterns in the last 12 centuries. *Clim. Past* 8, 227–249.
- 38 Lloyd, A.H., Graumlich, L.J., 1997. Holocene dynamics of treeline forests in the Sierra Nevada.  
39 *Ecology* 78, 1199–1210.
- 40 Luckman, B.H., Wilson, R.J.S., 2005. Summer temperatures in the Canadian Rockies during the last  
41 millennium: a revised record. *Clim. Dyn.* 24, 131–144.
- 42 Mann, M.E., Jones, P.D., 2003. Global surface temperatures over the past two millennia. *Geophys.*  
43 *Res. Lett.* 30, doi: 10.1029/2003GL017814.
- 44 Mann, M.E., Zhang, Z., Hughes, M.K., Bradley, R.S., Miller, S.K., Rutherford, S., Ni, F., 2008.  
45 Proxy-based reconstructions of hemispheric and global surface temperature variations over the past  
46 two millennia. *Proc. Nat. Acad. Sci.* 105, 13252–13257.
- 47 Meko, D.M., 1981. Applications of Box-Jenkins methods of time series analysis to the reconstruction  
48 of drought from tree rings. Ph.D. Dissertation. University of Arizona, Tucson.

- 1 Melvin, T.M., Briffa, K.R., 2008. A “signal-free” approach to dendroclimatic standardisation.  
2 *Dendrochronologia* 26, 71–86.
- 3 Melvin, T.M., Grudd, H., Briffa, K.R., 2013. Potential bias in ‘updating’ tree-ring chronologies using  
4 Regional Curve Standardization: re-processing the Torneträsk maximum-latewood-density data.  
5 *Holocene* 23, 364–373.
- 6 Myglan, V.S., Oidupaa, O.C., Vaganov, E.A., 2012a. A 2367-year tree-ring chronology for the Altai-  
7 Sayan region (Mongun-Taiga Mountain Massif). *Archaeol. Ethnol. Anthropol. Eurasia* 40, 76–83.
- 8 Myglan, V.S., Zharnikova, O.A., Malysheva, N.V., Gerasimova, O.V., Vaganov, E.A., Sidorov, O.V.,  
9 2012b. Constructing the tree-ring chronology and reconstructing summertime air temperatures in  
10 southern Altai for the last 1500 years. *Geogr. Nat. Resour.* 33, 200–207.
- 11 Oke, T.R., 2007. Siting and exposure of meteorological instruments at urban sites. In: *Air Pollution*  
12 *Modeling and Its Application XVII*. Springer US, 615–631.
- 13 Osborn, T.J., Briffa, K.R., Jones, P.D., 1997. Adjusting variance for sample-size in tree-ring  
14 chronologies and other regional-mean time-series. *Dendrochronologia* 15, 89–99.
- 15 Pages 2k Consortium, 2013. Continental-scale temperature variability over the Common Era. *Nature*  
16 *Geosc.* 6, 339–346.
- 17 Pages 2k PMIP3 group, 2015. Continental-scale temperature variability in PMIP3 simulations and Pages  
18 2k regional temperature reconstructions over the past millennium. *Clim. Past Discuss.* 11, 2483–2555.
- 19 Parker, D.E., 1994. Effects of changing exposure of thermometers at land stations." *Int. J. Climatol.* 14,  
20 1–31.
- 21 Rutherford, S., Mann, M.E., Osborn, T.J., Bradley, R.S., Briffa, K.R., Hughes, M.K., Jones, P.D., 2005.  
22 Proxy-based Northern Hemisphere surface temperature reconstructions: sensitivity to methodology,  
23 predictor network, target season, and target domain. *J. Clim.* 18, 2308–2329.
- 24 Salzer, M.W., Kipfmüller, K.F., 2005. Reconstructed temperature and precipitation on a millennial  
25 timescale from tree-rings in the southern Colorado Plateau, USA. *Clim. Change* 70, 465–487.
- 26 Salzer, M.W., Bunn, A.G., Graham, N.E., Hughes, M.K., 2014a. Five millennia of paleotemperature  
27 from tree-rings in the Great Basin, USA. *Clim. Dyn.* 42, 1517–1526.
- 28 Salzer, M.W., Larson, E.R., Bunn, A.G., Hughes, M.K., 2014b. Climate response in near-treeline  
29 bristlecone pine. *Environ. Res. Lett.* 9, doi: 10.1088/1748-9326/9/11/114007.
- 30 Schneider, L., Smerdon, J.E., Büntgen, U., Wilson, R.J.S., Myglan, V.S., Kirilyanov, A.V., Esper, J.,  
31 2015. Revising midlatitude summer temperatures back to A.D. 600 based on a wood density network.  
32 *Geophys. Res. Lett.* 42, doi: 10.1002/2015GL063956.
- 33 Schweingruber, F.H., 1996. *Tree Rings and Environment: Dendroecology*. Haupt Verlag, Bern.
- 34 Schweingruber, F.H., Bartholin, T., Schär, E., Briffa, K.R., 1988. Radiodensitometric-  
35 dendroclimatological conifer chronologies from Lapland (Scandinavia) and the Alps (Switzerland).  
36 *Boreas* 17, 559–566.
- 37 Shi, F., Yang, B., Mairesse, A., von Gunten, L., Li, J., Bräuning, A., Yang, F., Xiao, X., 2013. Northern  
38 Hemisphere temperature reconstruction during the last millennium using multiple annual proxies. *Clim.*  
39 *Res.* 56, 231–244.
- 40 Sidorova, O.V., Naurzbaev, M.M., Vaganov, E.A., 2006. An integral estimation of tree-ring  
41 chronologies from subarctic regions of Eurasia. *Trace* 4, 84–91.
- 42 Smerdon, J.E., Kaplan, A., Zorita, E., González-Rouco, J.F., Evans, M.N., 2011. Spatial performance  
43 of four climate field reconstruction methods targeting the Common Era. *Geophys. Res. Lett.* 38, doi:  
44 10.1029/2011GL047372.
- 45 Smerdon, J.E., Coats, S., Ault, T.R., 2015. Model-dependent spatial skill in pseudoproxy experiments  
46 testing climate field reconstruction methods for the Common Era. *Clim. Dyn.*, doi: 10.1007/s00382-  
47 015-2684-0.
- 48 St. George, S., 2014. An overview of tree-ring width records across the Northern Hemisphere. *Quat.*  
49 *Sci. Rev.* 95, 132–150.



- 1 Støve, B., Ljungqvist, F.C., Thejll, P., 2012. A test for non-linearity in temperature proxy records. *J.*  
2 *Clim.* 25, 7173–7186.
- 3 Tegel, W., Vanmoerkerke, J., Büntgen, U., 2010. Updating historical tree-ring records for climate  
4 reconstruction. *Quat. Sci. Rev.* 29, 1957–1959.
- 5 Trouet, V., Diaz, H.F., Wahl, E.R., Viau, A.E., Cook, E.R., 2013. A 1500-year reconstruction of annual  
6 mean temperature for temperate North America on decadal-to-multidecadal time-scales. *Environ. Res.*  
7 *Lett.* 8, doi: 10.1088/1748-9326/8/2/024008.
- 8 Villalba, R., 1990. Climatic fluctuations in northern Patagonia during the last 1000 years as inferred  
9 from tree-ring records. *Quat. Res.* 34, 346–360.
- 10 von Storch, H., Zorita, E., Jones, J., Dimitriev, Y., González-Rouco, J.F, Tett, S., 2004. Reconstructing  
11 past climate from noisy data. *Science* 306, 679–682.
- 12 Wang, J., Yang, B., Qin, C., Kang, S., He, M., Wang, Z., 2014. Tree-ring inferred annual mean  
13 temperature variations on the southeastern Tibetan Plateau during the last millennium and their  
14 relationships with the Atlantic Multidecadal Oscillation. *Clim. Dyn.* 43, 627–640.
- 15 Wigley, T.M.L., Briffa K.R., Jones, P.D., 1984. On the average of correlated time series, with  
16 applications in dendroclimatology and hydrometeorology. *J. Clim. Appl. Meteorol.* 23, 201–213.
- 17 Wiles, G.C., D'Arrigo, R.D., Barclay, D., Wilson, R.S., Jarvis, S.K., Vargo, L., Frank, D., 2014. Surface  
18 air temperature variability reconstructed with tree rings for the Gulf of Alaska over the past 1200 years.  
19 *Holocene* 24, 198–208 .
- 20 Wilson, R.J.S., Esper, J., Luckman, B.H., 2004. Utilizing historical tree-ring data for  
21 dendroclimatology: a case study from the Bavarian Forest, Germany. *Dendrochronologia* 21, 53–68.
- 22 Wilson, R., D'Arrigo, R.D., Buckley, B., Büntgen, U., Esper, J., Frank, D., Luckman, B., Payette, S.,  
23 Vose, R., Youngblut, D., 2007. A matter of divergence: tracking recent warming at hemispheric scale  
24 using tree ring data. *J. Geophys. Res.* 112, D17103.
- 25 Yadav, R.R., Braeuning, A., Singh, J., 2011. Tree ring inferred summer temperature variations over the  
26 last millennium in western Himalaya, India. *Clim. Dyn.* 36, 1545–1554.
- 27 Yang, B., Qin, C., Wang, J.L., He, M.H., Melvin, T.M., Osborn, T.J., Briffa, K.R., 2014. A 3,500-year  
28 tree-ring record of annual precipitation on the northeastern Tibetan Plateau. *Proc. Nat. Acad. Sci.* 111,  
29 2903–2908.
- 30 Zhang, Y., Shao, X.M., Yin, Z.Y., Wang, Y., 2014. Millennial minimum temperature variations in the  
31 Qilian Mountains, China: evidence from tree rings. *Clim. Past* 10, 1763–1778.
- 32 Zhang, H., Yuan, N., Esper, J., Werner, J.P., Xoplaki, E., Büntgen, U., Treydte, K., Luterbacher, J.,  
33 2015. Modified climate with long term memory in tree ring proxies. *Environ. Res. Lett.* 10, doi:  
34 10.1088/1748-9326/10/8/084020.
- 35 Zhu, H., Zheng, Y., Shao, X., Liu, X., Xu, Y., Liang, E., 2008. Millennial temperature reconstruction  
36 based on tree-ring widths of Qilian juniper from Wulan, Qinghai Province, China. *Chin. Sci. Bull.* 53,  
37 3914–3920.
- 38 Zorita, E., González-Rouco, J.F., Legutke, S., 2003. Testing the Mann et al. (1998) approach to  
39 paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G coupled  
40 climate model. *J. Clim.* 16, 1378–1390.
- 41

1 **Table and Figure Captions**

2 **Table 1** Millennium-length tree-ring based temperature reconstructions. Superscript \* indicates  
3 reconstructions developed using MXD (instead of TRW). The Icefield reconstruction contains both  
4 MXD and TRW data. *Signal* specifies the seasonality of reconstructed temperatures, with *p* indicating  
5 previous-year months. *T* is temperature, *T<sub>max</sub>* is maximum temperature, *T<sub>min</sub>* is minimum temperature.

6 **Table 2** *Homogeneity* scores. Chronology type *C* refers to reconstructions derived from a composite of  
7 material from living trees, remnant, historical and/or sub-fossil wood. Type *L* refers to reconstructions  
8 derived from only living trees. Temporal clustering (*Yes*) indicates records composed of data from  
9 distinct sites or species concentrated in discrete periods over the past 1000 years.

10 **Table 3** *Replication* scores. The number of TRW (or MXD) measurement series included in the  
11 reconstructions. *11th/20th* is the ratio of the mean replication during the 11th century relative to the  
12 mean replication during the 20th century. Values in red are estimates.

13 **Table 4** *Growth Coherence* scores. Mean, maximum, and minimum correlations among the TRW (or  
14 MXD) series included in the reconstructions. *11th/20th* is the ratio of the correlation during the 11th  
15 century relative to the 20th century correlation. Values in red are estimates.

16 **Table 5** *Chronology Development* scores. *Detrending method* 1 = RCS (and Signal Free), and 2 =  
17 individual detrending. *Age range* is the difference between highest and lowest mean age curve point  
18 over the past millennium. *Age trend* is the slope of a linear regression fit to the mean age curve over the  
19 past millennium (times 100). *Maximum frequency* indicates the wavelength of lowest frequency  
20 information retained in a reconstruction including 1 = multi-centennial, 2 = centennial, and 3 = decadal.  
21 Values in red are estimates.

22 **Table 6** *Climate Signal* scores. *Length* is the period of overlap with instrumental temperature data in  
23 years. *Correlation* is the Pearson correlation coefficient between the tree-ring chronology and the  
24 instrumental data over the calibration period. *Calibration/verification difference* indicates the  
25 correlation range between different periods of overlap with instrumental data. *Truncation* = 0.5 if the  
26 calibration period was shortened (e.g. due to divergence), *truncation* = 1 if this is not the case. Values  
27 in red are estimates.

28 **Table 7** Ranking of 39 tree-ring based temperature reconstructions based on the *Homogeneity*,  
29 *Replication*, *Growth Coherence*, *Chronology Development*, and *Climate Signal* scores. Last column  
30 indicates which datasets are publicly available.

31  
32 **Fig. 1** Location of millennium-length tree-ring based temperature reconstructions (circles). Colors  
33 indicate the June-August temperature change between the mean of the period 1964-2013 minus the mean  
34 of the period 1914-1963 using GISS 1200 km gridded data.

35 **Fig. 2** Tree-ring based temperature reconstructions. Black curves are the 13 reconstructions from Europe  
36 (a), 14 from Asia (b), 8 from North America (c), and 4 from the Southern Hemisphere (d) shown as  
37 anomalies from their 20th century means. Note that the reconstructed temperature variance differs  
38 substantially among records, largely as a result of the differing calibration schemes used in the original  
39 publications. Colored curves are the arithmetic means calculated over the common period of all  
40 reconstructions in each region. e, Comparison of the mean timeseries from Europe, Asia, and North  
41 America.

42 **Fig. 3** TRC replication curves. Black curves show the changing numbers of TRW (or MXD)  
43 measurement series integrated in the temperature reconstructions from Europe (a), Asia (b), North  
44 America (c), and the Southern Hemisphere (d). The replication curve of the Alps (larch) reconstruction  
45 in (a) refers to the right axis. Colored curves are the arithmetic means calculated over the common  
46 period covered by all reconstructions in each region. e, Comparison of the mean curves.

47 **Fig. 4** TRC inter-series correlations. Black curves show the correlation coefficients among the TRW (or  
48 MXD) measurement series integrated in the local temperature reconstructions from Europe (a), Asia  
49 (b), North America (c), and the Southern Hemisphere (d). Correlations are calculated over 100-year  
50 periods shifted in 10-year steps throughout the past millennium. The earliest value is centered on 1050,  
51 the most recent value on 1950. Colored curves are the arithmetic means calculated for each region, and  
52 dashed lines indicate the mean values over the millennium. e, Comparison of the mean inter-series  
53 correlation curves.

54 **Fig. 5** TRC age curves. Black curves show the mean tree age of the TRW (or MXD) data integrated in  
55 the temperature reconstructions from Europe (a), Asia (b), North America (c), and the Southern

- 1 Hemisphere (**d**). Colored curves are the arithmetic means calculated over the common period covered
- 2 by all reconstructions in each region. **e**, Comparison of mean replication curves.