Tree-ring reconstructed summer temperature anomalies for temperate East Asia since 800 C.E.

**Edward R. Cook · Paul J. Krusic · Kevin J. Anchukaitis ·
Brendan M. Buckley · Takeshi Nakatsuka ·
Masaki Sano · PAGES Asia2k Members**

Abstract We develop a summer temperature reconstruction for temperate East Asia based on a network of annual tree-ring chronologies covering the period 800–1989 C.E. The East Asia reconstruction is the regional average of 585 individual grid point summer temperature reconstructions produced using an ensemble version of point-by-point regression. Statistical calibration and validation tests indicate that the regional average possesses sufficient overall skill to allow it to be used to study the causes of temperature variability and change over the region. The reconstruction suggests a moderately warm early medieval epoch (ca. 850–1050 C.E.), followed by generally cooler 'Little Ice Age' conditions (ca. 1350–1880 C.E.) and 20th century warming up to the present time. Since 1990, average temperature has exceeded past warm epochs of comparable duration, but it is not statistically unprecedented. Superposed epoch analysis reveals a volcanic forcing signal in the East Asia summer temperature reconstruction, resulting in pulses of cooler summer conditions that may persist for several years. Substantial uncertainties remain, however, particularly at lower frequencies, thus requiring caution and scientific prudence in the interpretation of this record.

Keywords Paleoclimate · Temperature · Asia · Tree-rings · Volcanic

PAGES Asia 2k Members are Moinuddin Ahmed, Hemant Borgaonkar, Xu Chenxi, Guoqiang Chu, Zexin Fan, Narayan Gaire, Quansheng Ge, Katsuhiko Kimura, Jonathan Palmer, Tatyana Papina, Margit Schwikowski, Xuemei Shao, Feng Shi, Olga Solomina, Zafer Usama, Chenzi Xu, Koh Yasue, Masumi Zaiki, Kan Zhao, and Jingyun Zheng.

E. R. Cook (✉) · B. M. Buckley
Lamont-Doherty Earth Observatory of Columbia University,
Palisades, NY 10964, USA
e-mail: drdendro@ldeo.columbia.edu

P. J. Krusic
Department of Physical Geography and Quaternary Geology,
Stockholm University, Stockholm, Sweden

K. J. Anchukaitis
Department of Geology and Geophysics, Woods Hole
Oceanographic Institution, Woods Hole, MA 02543, USA

T. Nakatsuka · M. Sano
Graduate School of Environmental Studies,
Nagoya University, Nagoya 464-8601, Japan

1 Introduction

Instrumental weather data suggest that at hemisphere-to-global scales the temperatures of the late 20th and early 21st century are anomalously warm when compared to the last 150 years (Morice et al. 2012). Detection and attribution exercises using general circulation models (GCM) further support a hypothesis that recent temperature trends can only be explained if the radiative influence of increasing anthropogenic greenhouse gas emissions are considered (IPCC 2007). However, the relatively short time span of the observational record complicates both detection as well as attribution at regional scales, as natural inter-annual to multidecadal variability may obscure forced temperature trends (c.f. Stott et al. 2010). Longer estimates of regional temperature variability and change would provide a longer-term context on recent trends, improve characterization of the magnitude and spectra of modes of natural low frequency variability, and enhance efforts to estimate climate sensitivity to radiative forcing.

To this end, Past Global Changes (PAGES), a core project of the International Geosphere-Biosphere

Programme (IGBP), developed the 2k Network project, with the goal of generating a global array of regional temperature reconstructions covering the last 2000 years (http://www.pages-igbp.org/science/foci/focus-2/themes/2k-network). One of those designated PAGES 2k regions is Asia (under the auspices of the Asia2k project) and the reconstruction presented here is a contribution to that effort. Specifically, we have produced a regional boreal summer (June through August) temperature reconstruction for temperate East Asia from 23°N to 54°N based on long annual tree-ring chronologies. This reconstruction covers the period 800–1989 C.E. and has been appropriately scaled and extended up to 2009 with instrumental data to enable comparison of recent temperature change to the past. It is expressed as anomalies from the 1961–1990 period and is provided in both high-resolution inter-annual form and after smoothing using a 10 year low-pass smoothing spline filter (Cook and Peters 1981).

The method of reconstruction, an ensemble version of point-by-point regression, is described here in considerable detail, followed by the application of various tests of reconstruction quality. These tests include estimates of reconstruction uncertainty, a range of calibration and validation statistics commonly used in dendroclimatic reconstruction (e.g. Cook et al. 1999), and 'frozen grid' analyses designed to test for the stability of the low-frequency signal in the regional average reconstruction (e.g. Cook et al. 2004a). Together, they indicate a modest level of skill and a stable low-frequency signal in the reconstruction, which enables it to be used for comparison and interpretation of present and past summer temperature change over the last 1210 years in temperate East Asia.

2 Materials and methods

2.1 Instrumental temperature target field

The instrumental temperature field used for reconstruction is based on the global CRU TS $0.5 \times 0.5°$ gridded monthly temperature data (Mitchell and Jones 2005) version 3.1, covering the period 1901–2009. For purposes of this project, the resolution of the original $0.5 \times 0.5°$ field was reduced to $2 \times 2°$, resulting in a total of 773 grid points to reconstruct. Figure 1 shows a map with the overall monsoon Asia domain and the locations of the target $2 \times 2°$ temperature grid points.

The quality of the interpolated observational temperature field over the monsoon Asia grid declines significantly prior to 1951, particularly over the large central portion of the domain where the interpolated values relax towards climatology as available stations diminish back in time. Support for this assumption is evident in Fig. 2, which
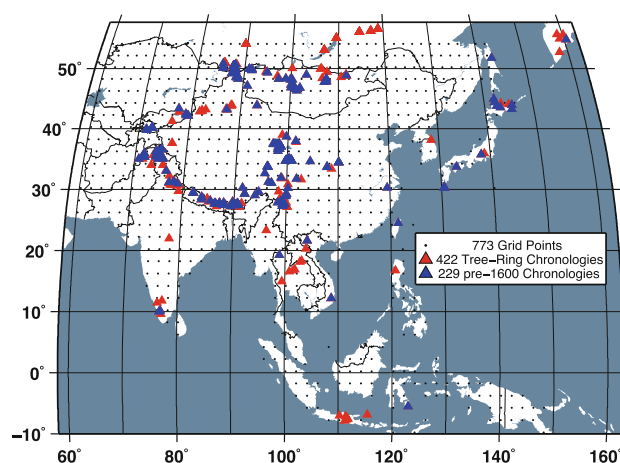

Fig. 1 Map of the Asia2k gridded temperature domain with the tree-ring network used for its reconstruction. There are 773 grid points of CRU TS 3.1 June-July-August average temperatures on a 2×2 grid (*black dots*) and a network of 422 annual tree-ring chronologies (*red filled triangles*) distributed irregularly across the temperature domain. Of the 422 chronologies, those that began on or before 1600 (*blue filled triangles*) were used for reconstruction. See the text for details

shows the changing number of single-station temperature records available for interpolation in the original CRU TS 3.1 $0.5 \times 0.5°$ grid by decade up to 1951. As the last map in the panel shows, the spatial coverage of temperature stations since 1951 is reasonably dense and well distributed spatially. However, prior to the 1940s the domain center (all of Mongolia, most of China, and all of Tibet) is devoid of local observations, which for the TS 3.1 gridded product must be estimated by interpolation from distant surrounding locations. While the method of interpolation used for producing TS 3.1 was state-of-the-art (Mitchell and Jones 2005), the estimated values in this large void must still exhibit some degradation in both accuracy and precision prior to 1951 due to the declining number of station records available and the increasing distances from the grid points being interpolated. In turn, this increase in gridded instrumental data uncertainty back in time must affect the interpretation of tests of reconstruction validity because it is impossible to know precisely how much the validation metrics are influenced by poor tree-ring estimates, poor instrumental data quality, or some unknown combination of both. It is important to note further that this paucity of instrumental data prior to the mid-20th century is not a problem unique to our domain. Indeed, a similar situation exists over vast areas of Africa, the Americas, Australia, and the oceans.

The poor spatial coverage of station-level temperature data over monsoon Asia prior to 1951 is similar to that encountered by Cook et al. (2010a) for the observational climate data used in producing the gridded 'Dai PDSI' data (Dai et al. 2004). For this reason, we will use the same
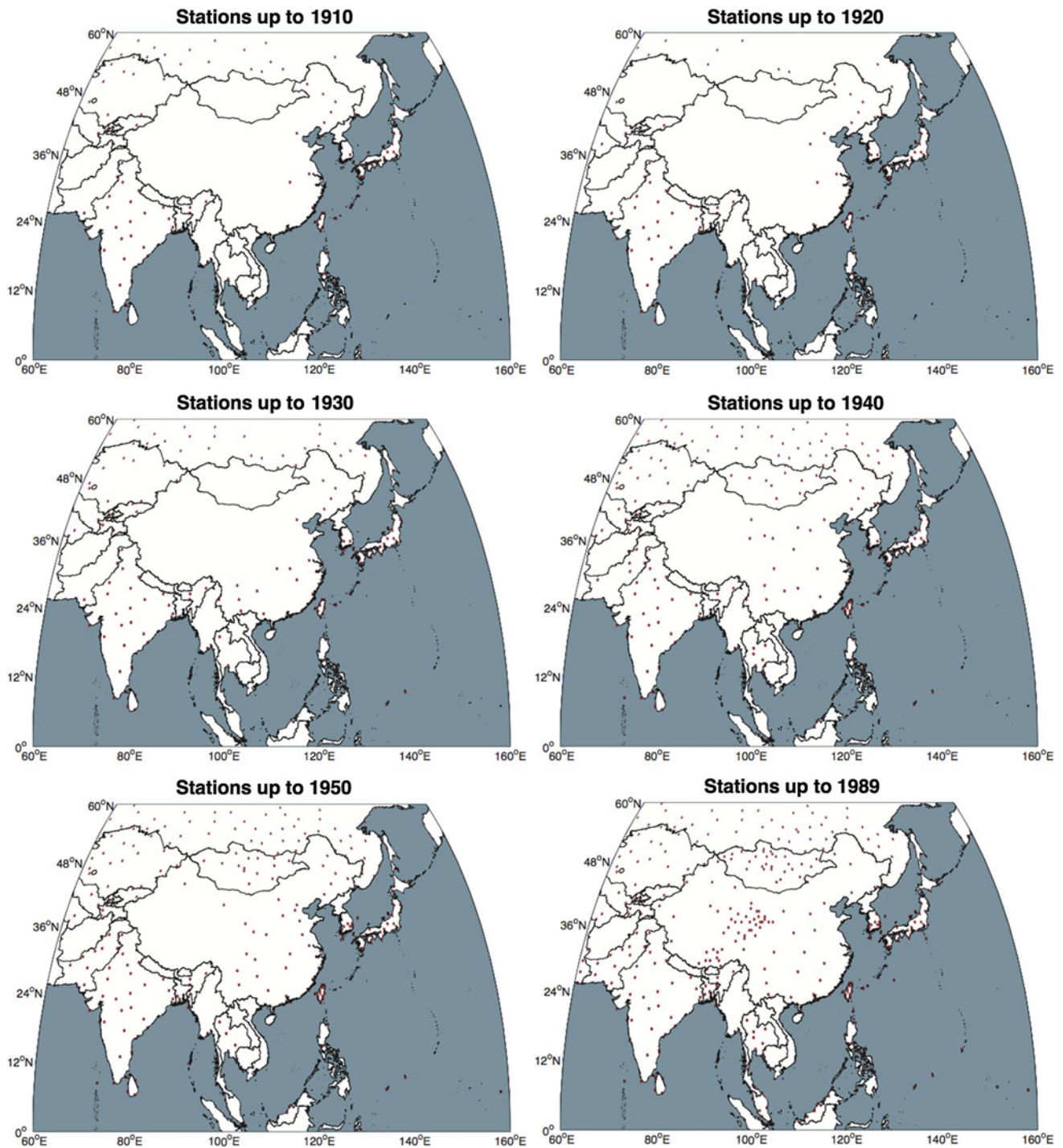
Fig. 2 Number of station temperature records available for the CRU TS 3.1 gridded temperature field by decade since 1901. The number of available stations drops precipitously prior to 1930

calibration and validation periods as in that study: 1951–1989 for calibration and 1920–1950 for validation, and the same arguments used for the choice of these periods applies. That is, the calibration period (1951–1989) was chosen to incorporate the best period of instrumental data and the largest number of tree-ring chronologies for reconstruction. The validation period was a compromise between using the best quality pre-1951 data (e.g. 1941–1950; see Fig. 3), while providing enough years of data for reasonably robust validation tests. Our selected 31-year validation period satisfies both requirements to an acceptable degree.
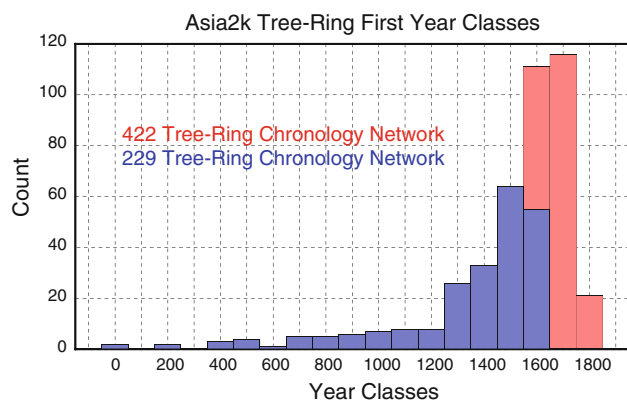
Fig. 3 Distribution of starting years of the tree-ring network shown in Fig. 1. The bins cover 100 year intervals centered on each century (e.g. 1250–1349, 1350–1449, etc). The *red columns* are for the full 422-chronology network and the *blue columns* are for the 229 chronologies that began on or before 1600. It is the latter that were used for reconstructing summer temperatures over the Asia2k domain. In either case, the large majority of the tree-ring chronologies begin after 1250, which restricts the ability to skillfully reconstruct temperatures back to early medieval times some 1,000 years ago

2.2 Tree-ring network

The tree-ring chronology network available for reconstruction of summer temperatures is shown in Fig. 1 as a series of filled red triangles. It is comprised of 323 annual tree-ring chronologies previously identified as potential predictors of the Palmer Drought Severity Index (PDSI) over monsoon Asia (Cook et al. 2010a) and 99 new tree-ring records contributed by members of the PAGES Asia2k project, for a total of 422 chronologies. Of these, 229 chronologies begin on or before 1600 and are indicated by the filled blue triangles in Fig. 1. These include 49 new chronologies contributed to this project by Asia2k members, an increase of 27 percent over the 180 in the original Monsoon Asia data set used to reconstruct drought before 1600 (Cook et al. 2010a). Thus, there is a substantial amount of new tree-ring data in the temperature reconstructions described here. Nonetheless, there remains a large degree of potential proxy predictor overlap between the Monsoon Asia Drought Atlas and the Asia2k temperature reconstructions described here. This overlap means that interpreting any joint features of these two reconstructed fields must be done with caution because they are not based on fully independent tree-ring data.

Considerable care was applied in processing the tree-ring chronologies for climate reconstruction to preserve centennial time scale variability consistent with attempting to avoid the 'segment length curse' (Cook et al. 1995), including the partial use of the 'signal free detrending' (Melvin and Briffa 2008) that optimizes the preservation of common low-to-medium frequency variability, particular at the end of the series. The tree-ring chronologies consist

of a diverse set of sites obtained over more than a decade of fieldwork in the region. Consequently, the predictor network consists of several tree species from a range of sites with known differences in climate response. Note also the very irregular distribution and relatively sparse number of tree-ring chronologies available for reconstruction, particularly south of the Himalaya in the lowland tropics. Although neither property is desirable for our specific application, and almost certainly limits the quality of the temperature reconstruction that is currently achievable, this new tree-ring network is actually denser than the 327-chronology network used to successfully reconstruct drought over monsoon Asia (Cook et al. 2010a).

Figure 3 shows the frequency distribution of starting years for the 422 tree-ring chronologies in the network, binned by 100 year periods centered on each century (e.g. 1250–1349, 1350–1449, etc.). The 229-chronology subset, those that begin on or before 1600, is also indicated. The large majority of the tree-ring chronologies begin after 1250, which restricts our ability to skillfully reconstruct temperatures back to early medieval times. However, a number of series do extend back through the 800–1250 C.E. interval and these will serve as the records to provide us with temperature estimates covering the past 1210 years.

As mentioned earlier, we target here a reconstruction of past summer temperatures. While it is arguable that annual temperatures would be better to reconstruct for the purpose of assessing temperature change over eastern and monsoon Asia, it is also the case that the direct effects of temperature on tree growth are largely confined to the photosynthetically active warm-season months, thus making summer temperature a logical and more causally defensible choice for reconstruction.

2.3 Reconstruction method

We used Point-by-Point Regression (PPR) (Cook et al. 1999) to reconstruct summer temperature from our tree-ring network. PPR has been used to successfully reconstruct drought over North America (Cook et al. 1999, 2004b, 2010b), monsoon Asia (Cook et al. 2010a), and northwestern Africa (Touchan et al. 2011), but this is the first time PPR is used to reconstruct a gridded temperature field. As first described by Cook et al. (1999), PPR sequentially reconstructs individual grid points of climate over a field using principal components regression. In our case, there are 773 $2 \times 2°$ grid points of CRU TS 3.1 summer temperatures over the monsoon Asia domain to reconstruct from the network of 422 annual tree-ring chronologies. Cook et al. (1999) showed that the underlying spatial patterns in the gridded data can be well preserved using this method.

As originally formulated for reconstructing drought over North America (Cook et al. 1999), PPR required two user-defined input parameters that were optimized for reconstructing the gridded climate field: a search radius and a screening probability. The 'optimal' search radius was designed to locate only those candidate tree-ring predictors that were likely to be reasonably physically related to climate at the grid point location being reconstructed. Once this 'level-1' pool of candidate predictors was determined, those tree-ring series were then tested (or screened) for statistical association with the grid point climate variable of interest. This screening probability was designed to retain for use in PPR only those candidate tree-ring predictors that were likely to contain a signal related to climate. After considerable experimentation, Cook et al. (1999) determined that there was a near-global joint optimum of both parameters that maximized the average calibration and validation statistics over the field: a 450 km search radius and a 90 % ($1 - \alpha$, 2-tailed) screening probability. The 450 km search radius is likely related to the manner in which the correlation between neighboring PDSI grid points decays as a function of distance, with 450 km encouragingly approximating the e-folding ($1/e$) distance for that precipitation dominated drought metric (see also Touchan et al. 2011).

The original PPR method was used in unaltered form by Cook et al. (1999, 2004b, 2010b) for reconstructing PDSI in North America. However, the basic PPR methodology had to be modified for reconstructing drought over monsoon Asia (Cook et al. 2010a) because the tree-ring chronology network and target PDSI grid were markedly inferior to the predictor/predictand fields used for reconstruction over North America. This made the previously determined 'optimal' search radius and screening probability impractical to use and also spurred the development of an ensemble-based version of PPR. In similar fashion, we make the following changes here to PPR for reconstructing the summer temperature field over eastern and monsoon Asia:

1. Instead of using multiple search radii, as done by Cook et al. (2010a) to bracket the likely search radius and allow it to vary over space, a single search radius was objectively determined by directly estimating the e-folding correlation decay distance over the target domain. This was accomplished by correlating all possible pairwise grid points of summer temperature and plotting those correlations against the distance between the grid point pairs being correlated (Fig. 4). In so doing, the e-folding ($1/e$) distance was determined to be approximately 1500 km, a result consistent with that determined by Jones et al. (1997) based on independent evaluations of hemispheric and global temperature data.

Fig. 4 Correlation decay with distance between 773 pairwise grid points of summer temperature used for reconstruction over monsoon Asia. A 40 % robust LOESS smooth has been used to indicate the overall shape of the decay in correlation. From this, the e-folding ($1/e$) decay in correlation equates to a distance of 1,500 km. This objectively defines the search radius used in PPR here for reconstructing summer temperatures from tree rings over monsoon Asia

Therefore, we used an initial search radius of 1,500 km at all grid points for locating the 'level-1' pool of candidate predictors for generating the summer temperature reconstructions. If at least five chronologies could not be found within 1,500 km of a given grid point, the search radius at that grid point was enlarged by 50 km increments until at least five chronologies were found. This is the 'dynamic search radius' feature first described in Cook et al. (1999). It enables climate grid points in tree-ring network voids to be reconstructed using PPR when the void is too large for at least five chronologies to be found within the initial search radius. In our case, the dynamic enlargement of the search radius was used for 229 out of 773 grid points of which 121 where located in the tropical regions south of 23°N where the tree-ring network is particularly sparse (e.g. in India) or the grid point locations were a long distance from the main Asian continent (e.g. in Indonesia). The rest came from far eastern Asia where the sparse tree-ring network or long distances from the main Asian continent again played a role.

2. As in Cook et al. (2010a), an ensemble-based approach was used to attempt to reduce the noise level in the reconstructions whereby the fixed screening probability was replaced by the variable weighting of the 'level-1' candidate tree-ring chronologies found with the 1,500 km search radius. The weighting (*wgt*), applied multiplicatively to the tree-ring chronologies,
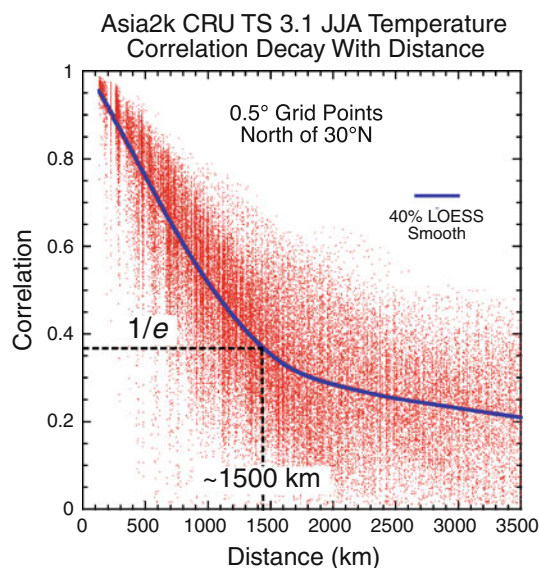
is related to some power p of the correlation r (sign ignored) of each tree-ring chronology with the grid point temperatures being reconstructed, i.e. $wgt = r^p$. This is expressed as $wTR = uTR \times r^p$, where wTR is the weighted tree-ring series and uTR is the original unweighted tree-ring series in standard normal deviate form. By varying p over a range from 0 to 2, a variable series of weightings is applied to the every suite of tree-ring chronologies found within a given search radius, thus perturbing the covariance matrix of predictors used in PPR. Figure 5 shows the functional forms of r^p used for reconstructing temperature from tree rings. A total of eight powers were applied (0, 0.10, 0.25, 0.50, 0.67, 1.0, 1.5, and 2.0), thus creating an ensemble of eight reconstructed temperature fields, which were then averaged to produce an ensemble mean field, and the regional summer temperature reconstruction presented here. The monotonicity of the r^p weighting preserves the rank order importance of the chronologies in the covariance matrix for $p > 0$, but varies their relative importance in the principal components analysis phase of PPR through scaling effects on their variances. For comparison, the fixed r cutoff based on the 90 % screening probability used by Cook et al. (1999) is shown in Fig. 5. The r^p weighting for $p > 0$ based on that fixed cutoff ranges from 0.07 to 0.88 for the same chronology correlated with temperature at that level. This illustrates how the r^p weighting affects the relative scaling of the same chronologies used in the covariance matrix in PPR.

Having made these determinations on how to reconstruct summer temperatures over monsoon Asia using PPR, three more data-based decisions were made that directly affected our results:

3. The chronologies found within each grid point's 1500 km search radius, and retained as predictors in weighted PPR, had to be positively correlated with summer temperature (i.e. $r > 0$) over the calibration period, with r^p naturally down-weighting the more weakly correlated predictors depending on the choice of p, and no down-weighting applied for the limiting case of $p = 0$. The reason for requiring $r > 0$ is our belief that the directly correlated effects of temperature on tree growth (e.g. direct temperature effects on physiological processes and photosynthetic rates) are likely to produce a more reliable expression of past temperature variability compared to inversely correlated temperatures effects on growth, which are more likely related to moisture stress (that is, an inverse temperature effect due to evapotranspiration demand). Those tree-ring chronologies that did not pass the $r > 0$ threshold showed no empirical evidence of
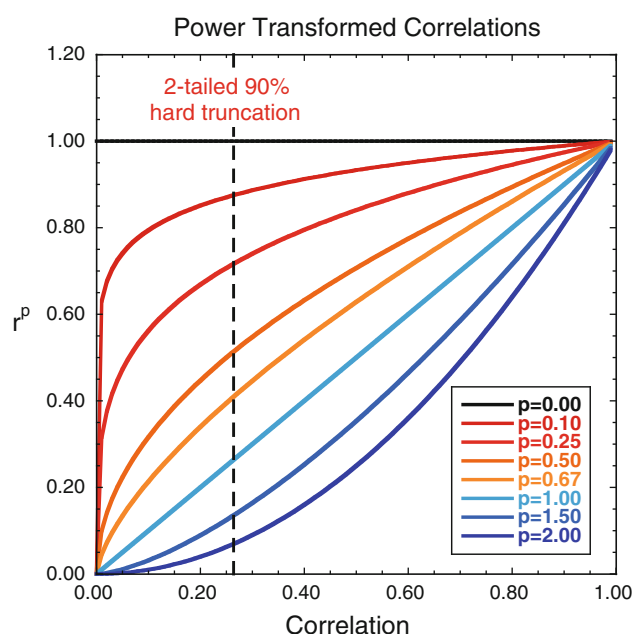

Fig. 5 Plot showing the functional forms of the rp weighting applied to the tree-ring chronologies found within a given search radius for a given grid point of temperature being reconstructed. For comparison, the *vertical dashed line* shows the fixed 90 % screen probability cutoff determined previously by Cook et al. (1999) to work well for reconstructing PDSI over North America

temperature sensitivity and were not used as predictors for the reconstruction. This was the only predictor variable screening applied to the tree-ring chronologies used in PPR, and it was entirely restricted to the 1951–1989 calibration period. The 1920–1950 validation period temperature data was not used in any data screening exercise. To do otherwise would have biased any tests of reconstruction validity and compromised their usefulness as an estimate of reconstruction skill.

4. To minimize the impact of autocorrelation or trend on the determination of which tree rings were correlated positively ($r > 0$) and therefore likely to be physiologically associated with temperature variability over the calibration period, a simple sign-based screening test was applied after first-order autocorrelation ('red noise') was removed from the calibration period tree-ring and temperature data by fitted first-order autoregressive (AR1) models. In addition, three correlation tests were used to determine if the tree rings were truly positively correlated with summer temperature: the parametric Pearson product-moment correlation, the non-parametric Spearman rank correlation, and a robust form of the Pearson correlation based on the bi-square weight function (Gnanadesikan 1997). If all three correlations were positive a tree-ring variable retained as a predictor in PPR. The r^p weighting applied to each chronology was based on the average

r of those three estimates. However, once the tree-ring predictors were selected in this way for $r > 0$ based on the AR1 pre-whitened data, the actual reconstructions were produced using the original tree-ring chronologies and temperature data with all of their original persistence or trend intact. This procedure was applied on both t_0 and $t + 1$ tree-ring predictors following the use of lagged tree-ring variables as candidate predictors in Cook et al. (2010a). While statistical screening is less desirable than pure biophysically-informed selection, it is the only practical solution for a large predictor network developed by multiple investigators over more than a decade for multiple purposes with limited metadata.

5. Only those tree ring series that began on or before 1600 C.E. were used as candidates for reconstruction based on the 1,500 km search radius and the $r > 0$ screening. The locations of those sites are shown in Fig. 1 as filled blue triangles. Doing so eliminated the shortest tree-ring chronologies that tend to be more deficient in low-frequency variance due to the 'segment length curse' (Cook et al. 1995). This reduced the total number of chronologies available for use in the network from 422 to 229, a reduction of ~46 %. While this sizable reduction in the number of chronologies used might appear counter to development of the best reconstruction possible over the domain, it was done to insure that the final PAGES Asia2k reconstruction would provide a reasonably homogeneous estimate of the centennial-scale temperature variability over the last millennium that was not weighted excessively toward the shorter, more spectrally limited, chronologies. The domain map of tree-ring chronologies beginning on or before 1600 also indicates that this subset of longer chronologies is spatially unbiased relative to the distribution of all chronologies from the Himalaya north to Mongolia. Thus, using only those chronologies that begin on or before 1600 for reconstruction should not bias the large-scale results. Finally, the minimum acceptable sample size for each annual mean tree-ring index value in each chronology was set to four. This shortened the lengths of the chronologies used for reconstruction, but in so doing eliminated the least reliable inner parts of those chronologies based on fewer than 4 samples. We did not apply the $EPS < 0.85$ cutoff (Wigley et al. 1984) for eliminating the weaker inner portions of the chronologies because (1) it is a 'rule of thumb' with no firm statistical basis for any specific cutoff and (2) each grid point reconstruction was based on multiple chronologies, thus making the declining sample sizes in individual chronologies much less of an issue.

3 Results

3.1 Field calibration and validation results

PPR was applied as described above to the reconstruction of gridded summer temperatures over our target domain. Figure 6 shows maps of the calibration period coefficient of determination or R^2 (CRSQ) and cross-validation reduction of error (CVRE). The latter is a 'leave-one-out' procedure analogous to R^2 based on Allen's PRESS statistic (Allen 1971) and is a more conservative measure of explained variance than CRSQ. In extremely weak calibration cases CVRE can actually go negative, a clear indication of no calibration skill. The validation period statistics are the square of the Pearson correlation (VRSQ), the reduction of error statistic (VRE), and the coefficient of efficiency (VCE), all of which are measures of explained variance between actual and estimated values when positive. Note that all three (especially VRE and VCE) can go negative as well, which again indicates a lack of skill in the reconstructed values. As a rule of thumb (because no theoretical significance tests are available for VRE or VCE), if VRE > 0 or VCE > 0 then the reconstruction has some skill in excess of the calibration or verification period climatologies of the instrumental data (Cook et al. 1999). In the case of VRSQ, values significant at the 95 % level based on positive correlations only ($p < 0.05$, 1-tailed) were used as indications of significant reconstruction skill. See Cook et al. (1999, 2010a) for discussions of these statistics.

The calibration/validation maps show clear areas of white, which identify those grid point locations that did not calibrate or validate based on the statistical tests described above. It is clear that the area exhibiting the least amount of model skill is in the lowland tropics south of the Himalaya, where very little reconstruction skill is evident. This might be a consequence of removing from the predictor pool those chronologies that began after 1600, as most of those from the lowland tropics were eliminated. However, even if they had been left in, they would not have contributed to any understanding of past temperatures before 1601. In contrast, the region north of roughly 23°N shows evidence of some reconstruction skill over a reasonably coherent portion of the domain. We view this result as sufficient to justify producing an estimate of past boreal summer temperatures for this region back to 800 C.E.

Figure 7 provides more details about the relative contributions of the 229 tree-ring chronologies used as predictors in our reconstruction. The circle diameters are scaled to reflect in relative terms the number of times the chronologies were used for grid point reconstruction. This shows how the tree-ring chronologies were used multiple times in reconstructing individual grid points due to the
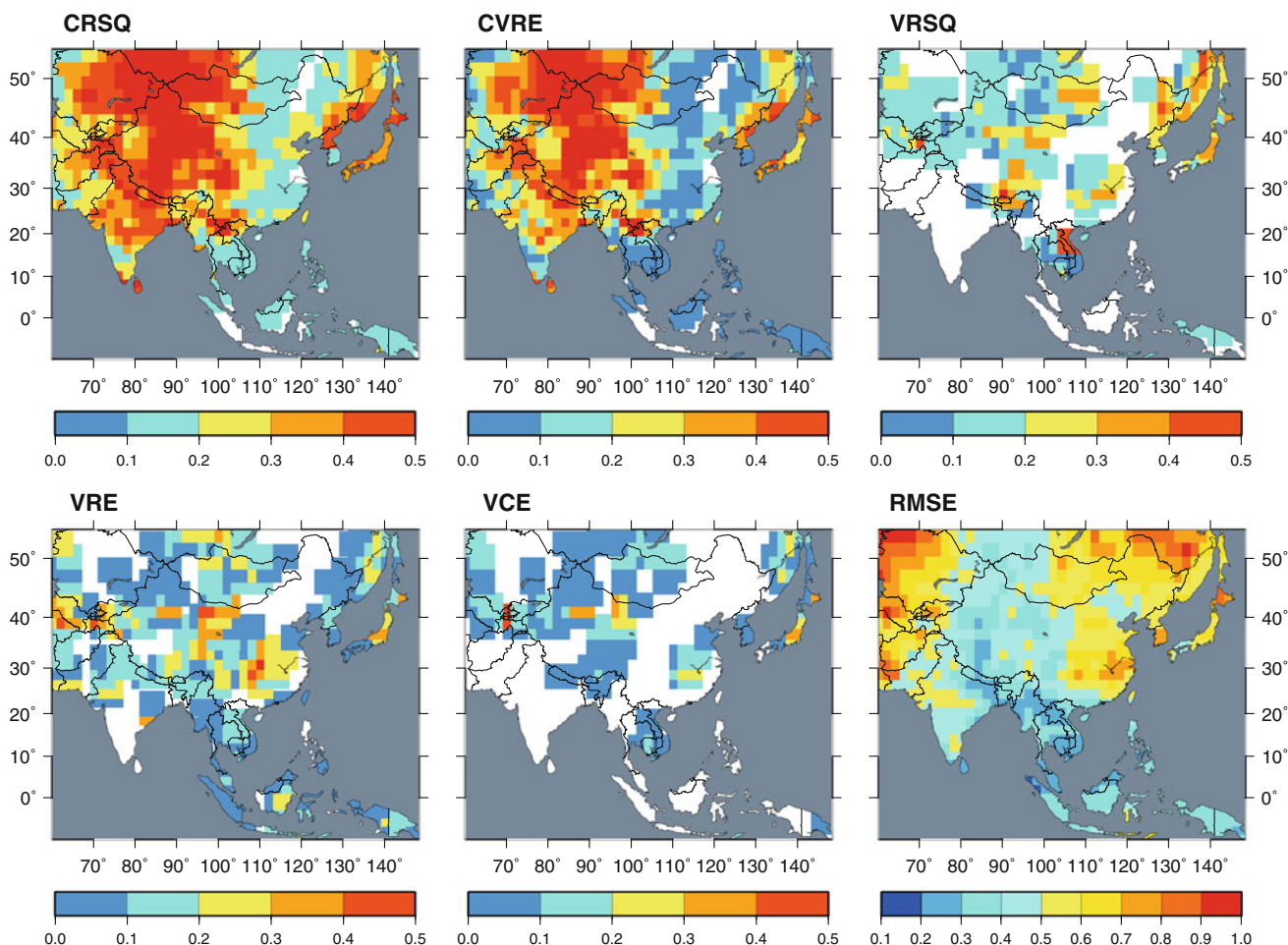
Fig. 6 Calibration and validation maps for the ensemble mean of the gridded summer temperature reconstructions from tree rings. All except the RMSE statistics are on the same scale for easy comparison. *White spaces* in the fields indicate grid points where either the calibration and/or validation statistics did not pass the tests described

in the text. It is clear that the lowland tropics south of the Himalayas and Tibetan Plateau have relatively little reconstruction skill. In contrast, the region north of roughly 23°N shows some evidence of reconstruction validity

way they fell within the 1,500 km search radius of several adjacent grid points. The variable circle sizes also reflect the fact that some tree-ring chronologies were correlated with gridded temperatures more often than others. In addition, each circle has been split and colored to show the average correlation with summer temperature for the t_0 and $t + 1$ tree-ring variables used.

This information is expressed in a non-spatial way in Fig. 8, which shows boxplots of the number of times the 229 chronologies were used for reconstruction and the correlations with temperature by those series for t_0 and $t + 1$. The median number of times used is 75 for t_0 and 47 for $t + 1$. The smaller number for $t + 1$ simply reflects the fact that this lag variable was less often related to summer temperature ($r > 0$) compared to the unlagged t_0 variable. The boxplots of correlations are perhaps the more interesting result. The median correlation is only 0.16 for t_0 and 0.12 for $t + 1$. Both are very weak by most statistical

standards suggesting a useless predictor for reconstruction if pertaining to a single chronology. Yet, the joint use of multiple chronologies in PPR has produced a credible level of both calibration and validation skill (Fig. 6). This shows that, even when weakly correlated with summer temperature, each retained chronology is contributing enough independent information to the reconstruction to be useful. It also shows the value of using large multi-species tree-ring networks for climate reconstruction.

3.2 A boreal summer temperature anomaly reconstruction since 800 C.E.

Figure 9a shows our boreal summer temperature reconstruction for temperate eastern Asia north of 23°N, the best validated portion of the overall temperature grid. It is the average of all grid point reconstructions from that region (585 in total) and is expressed as anomalies from the

Fig. 7 The relative contributions of the 229 tree-ring chronologies used for reconstruction. The *circle* diameters are scaled to reflect in relative terms the number of times the chronologies were used for grid point reconstruction. In addition, each *circle* has been split and colored to show the average correlation with summer temperature for the t_0 and $t + 1$ tree-ring variables used. This figure shows how the tree-ring chronologies were used multiple times in reconstructing individual grid points due to the way that they fell within the 1,500 km search radius of several adjacent grid points. The variable *circle* sizes also reflect the fact that some tree-ring chronologies were correlated with temperature more often than others
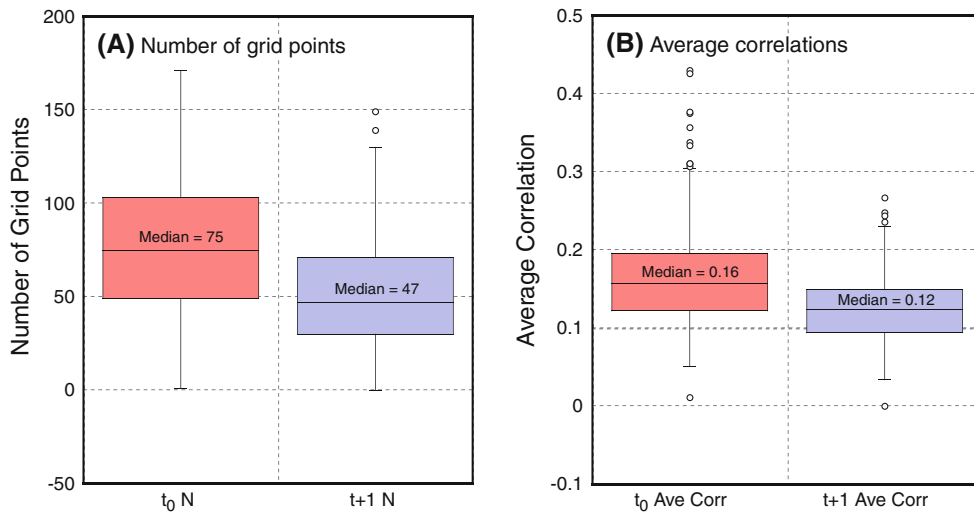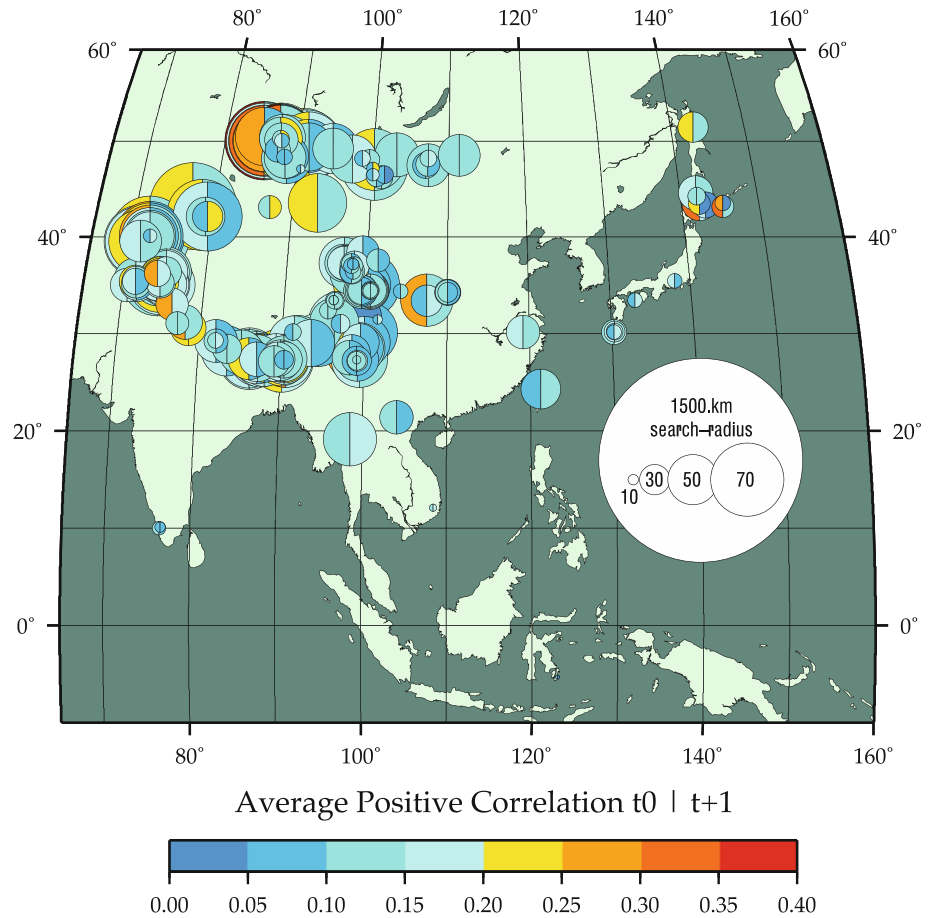

Fig. 8 *Boxplots* of the number of times each of the 229 tree-ring chronologies were used for reconstruction at different grid points (**a**) and their correlations with summer temperature (**b**). The median number of times each chronology was used is 75 for t_0 and 47 for $t + 1$. The smaller number for $t + 1$ simply reflects the way in which that lag variable was less often related to summer temperature ($r > 0$) compared to the unlagged t_0 variable. The median correlation of those tree-ring variables used is only 0.16 for t_0 and 0.12 for $t + 1$. Both are very weak by most statistical standards, yet the joint use of multiple chronologies in PPR has produced a credible level of both calibration and validation as indicated in Fig. 6
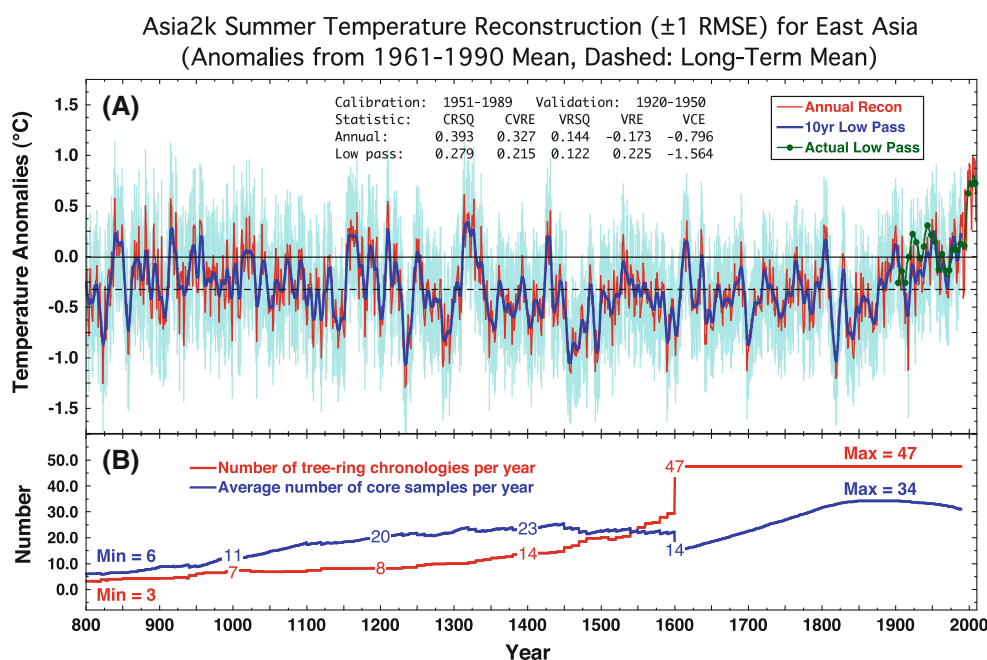
Fig. 9 Asia2k summer temperature reconstruction from tree rings for temperate East Asia with 1 RMSE uncertainties (**a**) and the changing number and average replication of the chronologies used (**b**). The annually resolved values are in *red*, the 10 year low-pass spline filtered values are in *dark blue*, and the uncertainties are in *pale blue*. The time period covered is 800–2009 C.E., but the 1990–2009 data are instrumental data for comparison to the past estimated temperatures. The *solid horizontal line* is the zero-anomaly value based on the 1961–1990 period. The *dashed horizontal line* is the long-term mean, which is −0.32°C below the zero-anomaly mean. The calibration and validation statistics indicate modest skill (1920–1950: VRSQ = 0.14 for inter-annual and 0.12 for 10 year low-pass values; VRE = 0.225 for 10 year low-pass values), but VCE is uniformly negative due to the progressive departure of the two curves prior to 1951 indicated by the *blue* and *green* 10 year low-pass filtered curves. The number of chronologies declines back in time as expected, but the average replication per chronology remains reasonably high back to 1000 C.E

1961–1990 mean. The full period covers 800–2009 C.E., with the data since 1990 being instrumental for easy comparison of recent warming to earlier reconstructed temperatures. The pre-1990 reconstruction has also been scaled to match the amplitude of the instrumental data over the calibration period to make comparisons with the extension to 2009 compatible.

To further facilitate meaningful comparisons, uncertainties in the form of ± 1 RMSE (root mean square error) has been added to the rescaled, regression-based, reconstructed values. These estimates of RMSE are simply averages of the individual ensemble member RMSEs averaged yearly over all grid points having reconstructed data back to 800 C.E. Estimated this way, average RMSE ranges from 0.47°C to 0.56°C, which is greater than the standard deviation of the reconstruction itself (0.33°C). This reflects the weak-to-moderate calibration skill of the individual grid point reconstructions indicated in Fig. 6. It also makes it difficult to compare changes in summer temperature from current conditions back to 800 C.E. because the reconstruction uncertainties are large. The RSMEs used here are also not exact because the lost variance due to regression has been artificially added back to the reconstruction to produce less biased past estimates for

comparison to current summer temperature change. Doing so underestimates the uncertainties in the rescaled reconstruction because the RMSEs do not reflect the added variance associated with rescaling. However, we are more interested in comparing changes in mean temperature over time here, with the RMSEs providing simple lower-limit estimates of true uncertainty. See Ammann et al. (2010) for a discussion related to correcting for signal attenuation in paleoclimate reconstructions due to noise in the climate proxies used.

Calibration and validation period statistics for this regional average are inset in the Fig. 9 plot. For the annually resolved data reconstructed directly from PPR, they indicate a modest level of calibration skill (CRSQ = 0.39, CVRE = 0.33), with some validation skill also indicated (VRSQ = 0.14, $p < 0.05$, 1-tailed), but VRE and VCE are both negative. The cause of this latter problem appears to be due in part to the way in which the tree-ring estimates fall progressively below (and are therefore are colder than) the actual data from 1950 back to 1920 (and beyond to 1901). This difference is most easily seen between the blue and green 10yr low-pass filtered curves plotted in Fig. 9. The calibration and validation statistics based on the 10 year low-pass filtered values are

also provided for comparison to those based on the annual values to show how well the reconstruction might be interpreted as an expression of temperature variability on decadal and longer time scales. The results are weaker overall, although not inconsistent with those based on the annual values.

The biggest difference in the validation statistics after smoothing is the positive VRE (from -0.17 to $+0.23$), which implies some skill in the reconstruction at lower frequencies. This is due in part to a tendency for positive bias in VRE when changes in the mean level of the instrumental data from the calibration to the validation period are likewise expressed at some level in the reconstruction (Cook et al. 1999). This particular source of bias is missing from the VCE statistic, which is why it remains negative after smoothing. This does not mean that the positive VRE after smoothing is completely misleading, however. For the evaluation of forecast (or hindcast in our case) skill, as the reduction of error statistic was originally designed to evaluate (Lorenz 1956), the positive VRE after smoothing suggests some level of lower-frequency hindcast skill in the reconstruction in excess of climatology as defined by the calibration period mean. Being based on smoothed data, this estimate of skill must be viewed with substantial caution, however.

Figure 9b shows how the number of chronologies (including both t_0 and $t + 1$) used for reconstruction has changed back in time. It is by construction constant at 47 back to 1,600 after which it systematically declines to 14 by 1,400, 8 by 1,200, 7 by 1,000, and 3 by 800. The average replication (number of core samples per year) in the chronologies used is reasonably high back to 1,200 (an average of 20 samples per chronology) after which it declines to 11 by 1,000 and 6 by 800. Without question, these declines in the number of chronologies and average replication per chronology used for reconstruction will affect reconstruction fidelity, but it is difficult to say by exactly how much and in what ways. The estimated reconstruction uncertainties ($\pm 0.50°C$ on average) in Fig. 9a tell us about the precision of the estimates, but they provide little insight into how reconstruction accuracy might have been affected.

One way to investigate how reconstruction accuracy might change over time is through the creation of frozen grid average reconstructions based on only those grid point series that begin on or before some starting year prior to 1600. Doing so will reflect the sensitivity of the reconstruction based on the full grid to that based on a reduced number of grid points related to the declining number of tree-ring chronologies available for reconstruction (see Figs. 3 and 9b). It will also reveal the potential for geographic bias in the reconstruction prior to 1600 because the spatial distribution of the longest chronologies used is not random with respect to the predictor network or predictand field.

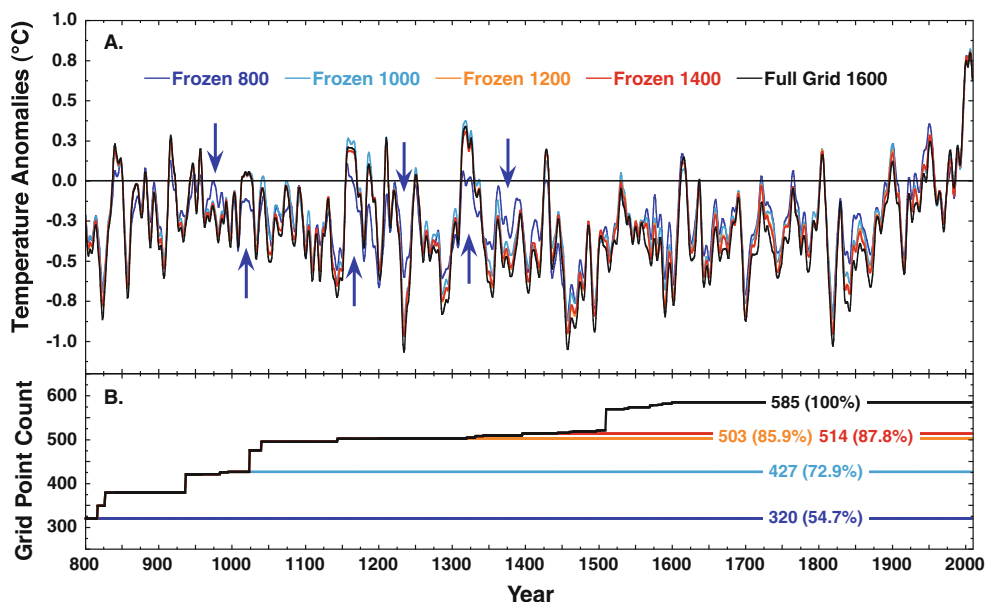To this end, four subsets of reconstructions with frozen starting years of 1400, 1200, 1000, and 800 have been

Fig. 10 Frozen grid experiments (**a**) in which four subsets of reconstructions with frozen starting years of 1400, 1200, 1000, and 800 are compared to the full grid reconstruction, all after 10 year low-pass filtering to facilitate comparison. The change in the number and percentage of grid points used in each frozen grid reconstruction (**b**) are indicated below. Over all, there is an excellent degree of similarity between the low-frequency signals of the reconstructions back to a frozen grid starting year of 1000. In contrast, the frozen grid reconstruction based on a starting year of 800 shows occasional large differences from the others as indicated by the *blue arrows*

compared to the full domain reconstruction (Fig. 10a), all after 10 year low-pass filtering to facilitate visual comparison. The change in the number and percentage of grid points used in each frozen grid reconstruction are indicated in Fig. 10b. This comparison indicates a high degree of low-frequency coherence in the reconstruction back to 1000 because it is still based on 427 out of 585 grid points (72.9 % of the total). Even back to 937 C.E. there are still 421 grid points. In contrast, the reconstruction based on only 320 grid points, all beginning at or before 800 C.E. (54.7 %) of the total), show some clear differences from the others, especially where indicated by the blue arrows. This intermittent bias is most likely related to the decreased number of grid points and their more restricted geographic locations. So overall, the low-frequency variability in the full-grid reconstruction is reasonably robust to changing numbers of grid points and their geographic contributions back to 1000 (or perhaps even 937) C.E. This result does not tell us if the low-frequency signal in the full-grid reconstruction is a truly accurate expression of past summer temperatures over temperate East Asia. Rather, whatever level of accuracy the reconstruction does possess, it appears to be reasonably robust to the data and method used for reconstruction.

Having documented the relatively consistent low-frequency signal in the reconstruction back to 1000 C.E., it is useful now to determine where the signal in the unfiltered series is most strongly expressed across the spatial domain. Figure 11 shows the pattern of significant correlation ($p < 0.10$, 1-tailed) between the regional reconstruction and the original $0.5 \times 0.5°$ gridded CRU TS 3.1 temperature data. These maps are produced for both the calibration and validation periods and indicate that the region north of 23°N is most strongly reflected by the full regional average. These maps are also provided in three ways based on the original, detrended, and 1st-differenced data to test the degree to which the patterns of significant correlation are sensitive to trend or autocorrelation in the data.

These maps provide two insights:

1. The portion of the region that appears to be most reliably expressed in the reconstruction is north of 36°N. This result was somewhat unexpected given the significant calibration and validation statistics shown in Fig. 6 between 23°N and 36°N. However, the total number of grid points in the region north of 36°N is 394 compared to 191 southward to 23°N. Thus, the area that calibrates and validates best in terms of positive correlation in Fig. 11 reflects in part the areal weighting of the average reconstruction itself. The small region of negative correlation south of 30°N might reflect the influence of the Asian monsoon on

Correlations between Asia2k Summer Temperature Reconstruction and CRU TS 3.1 Gridded Temperatures (p<0.10, 1-tailed)

(A) Calibration period: 1951–1989
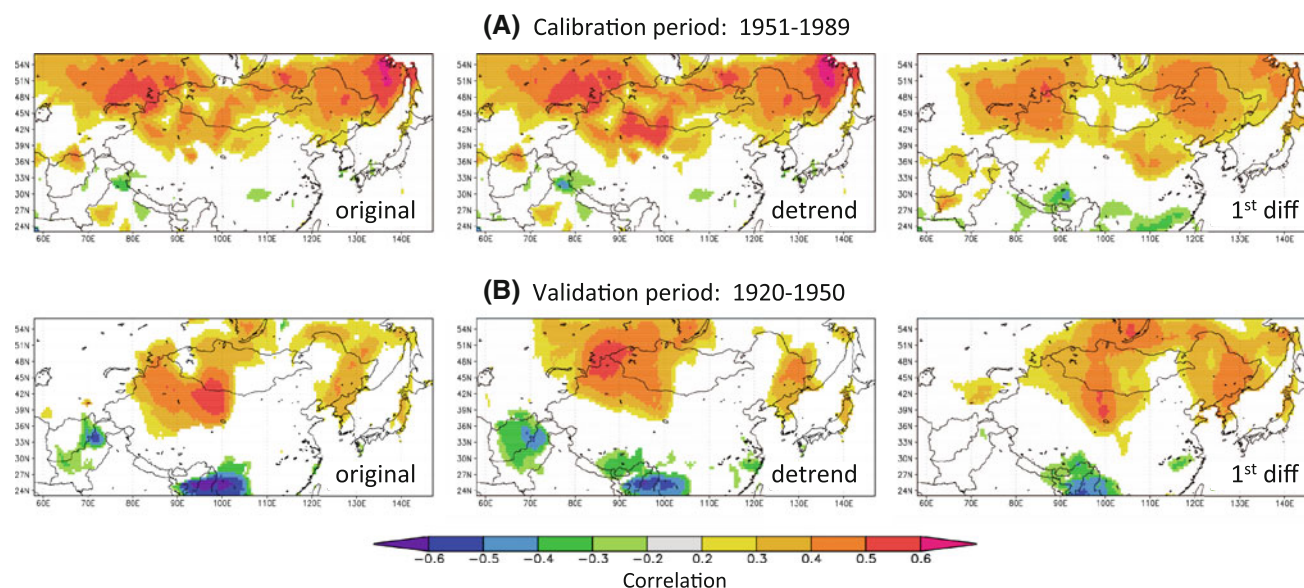
(B) Validation period: 1920–1950

Fig. 11 Maps of correlation between the regional Asia2k summer temperature reconstruction (Fig. 9) and CRU TS 3.1 temperature data. The correlations are for the calibration (a) and validation (b) periods used for reconstruction. For each period, the correlations were calculated for the original data, the data after linear detrending, and the data after they have been 1st-differenced. The latter two data transformations test for sensitivity of trend or positive autocorrelation on the pattern of significant ($p < 0.10$, 1-tailed) correlations. The 1-tailed test is based on the expectation that only positive correlations over the region north of 23°N are relevant. The correlation maps were produced using KNMI's Climate Explorer (http://climexp.knmi.nl)

the CRU TS 3.1 temperature field, but its very weak presence in the calibration period correlation field calls into question the importance of this interpretation.

2. The patterns of significant correlation are relatively insensitive to the presence of trend or autocorrelation. That is, the geographic patterns of significant correlation remain more or less intact even after detrending and 1st-differencing were applied to the data. This finding applies to both the calibration and validation period data and is further evidence of meaningful skill in the reconstruction presented here.

4 Discussion

The summer temperature reconstruction presented here is a credible, albeit noisy, estimate of past summer temperatures over temperate East Asia since 800 C.E. The calibration and validation statistics are statistically significant over much of the grid north of 23°N, but significant uncertainties of various kinds remain. For example, it is difficult to say if the lack of low frequency matching seen between the actual and reconstructed 10yr low-pass filtered values prior to 1951 (Fig. 9A) is due solely to the tree-ring estimates or related at some level to insufficient local instrumental data in the pre-1951 period, as discussed earlier. The latter cannot be readily dismissed. For this reason, we believe that this reconstruction may reflect the true temperature field better than the validation statistics indicate and is therefore worth further evaluation as our current best estimate of how summer temperatures may have varied over temperate East Asia since 800 C.E.

As a working hypothesis, the reconstruction indicates that instrumental summer temperatures since 1990 (0.53° C ± 0.09) (and especially since 1998: 0.74° C ± 0.07) have been on average warmer than any period of comparable duration over the previous 1190 years. This interpretation is considerably tempered, however, by the realization that during certain reconstructed past warm epochs (e.g. 1154–1169: 0.19°C ± 0.52 and 1312–1328: 0.28°C ± 0.52) their RMSE uncertainties overlap with the average instrumental temperatures (and their one standard error uncertainties) since 1990. This means that we cannot say with any statistical certainty that current summer warmth over temperate East Asia is unprecedented based on this reconstruction. Although it appears to be unusual, a comparable period of warmth may have happened before. This statement cannot be interpreted to reflect any doubt about the existence of modern greenhouse warming, as the forcing and boundary conditions of today's climate system associated with elevated greenhouse gas concentrations in the atmosphere are radically different from the past.

Rather, it simply suggests that current climate warming over temperate East Asia may not yet have emerged in a statistically significant sense from the long-term natural background variability of the climate system. In turn, the fact that past natural climate variability over East Asia could have resulted in average summer temperatures comparable to todays warmth means that the climate system has the capacity for unusually warm temperatures even without the additional forcing from anthropogenic greenhouse gases.

More generally, the East Asia temperature reconstruction indicates that average summer conditions during early medieval times (ca. 850–1050 C.E.: −0.18 ± 0.56) were comparable to the 20th century up to 1989 (1900–1989: −0.08 ± 0.47). Between those two warmer periods is a prolonged period of generally cooler conditions lasting roughly from 1350 to 1880 C.E. (−0.44 ± 0.49), which is punctuated by briefer periods of much colder conditions (e.g. 1453–1497 C.E.: −0.76 ± 0.50). Overall, these long-term patterns of warmer and cooler conditions are consistent with the spatial signatures of the 'Medieval Warm Period' *sensu lato* and the 'Little Ice Age' as reconstructed by Mann et al. (2009) and Ljungqvist et al. (2012) over East Asia. However, the magnitude of temperature change between these two epochs is highly uncertain in our reconstruction as indicated by large RMSE in Fig. 9.

The reliance on the somewhat artificial 1961–1990 period for calculating temperature anomalies can result in a distorted appearance of how temperatures have changed over the past 1210 years because most of the reconstructed values fall below the 1961–1990 anomaly mean. This can be readily appreciated by calculating and plotting the long-term reconstruction mean (800–1989 C.E.) for comparison, as shown on Fig. 9. The long-term mean (dashed line) is −0.32°C cooler than the 1961–1990 period used for anomaly estimation and is almost one standard deviation (0.33°C) below the anomaly period mean. This rather large difference emphasizes the degree to which the 20th century has warmed since 1961. Viewed from the perspective of this long-term mean, both the early medieval (ca. 850–1050 C.E.: 0.13 ± 0.56) and 20th century (1900-1989: 0.24 ± 0.47) periods stand out more strongly as epochs of generally elevated warmth, with the latter more persistently so in the 20th century as temperatures emerged from the Little Ice Age. In either case, post-1990 summer temperatures have been unusually warm relative to the past, but again this change cannot be viewed as statistically unprecedented because of the high uncertainties described herein.

It is not the purpose of this paper to conduct formal attribution tests of the causes of the estimated changes in East Asian temperatures since 800 C.E. and doing so would be fraught with considerable uncertainties in any case due

to the limitations of this reconstruction. Even so, there is evidence that large tropical volcanic eruptions have had an influence on summer temperatures in our reconstruction. To determine the presence or absence of a volcanic forcing signal, we conducted a superposed epoch analysis (SEA, Haurwitz and Brier 1981) using modified block resampling to assess statistical significance in the presence of auto-correlation (Adams et al. 2003). Because of ambiguity in the chronology and strength of past explosive volcanism (c.f. Gao et al. 2006; Schmidt et al. 2012), we consider two lists of volcanic eruptions to use as key years for conducting the SEA. From the forcing series used by Ammann et al. (2007), we identify events equal to or larger than the radiative forcing due to the eruption of Mount Pinatubo in 1991. Fischer et al. (2007) used historical and geological information to identify event years from well-characterized eruptions since the late 16th century. Figure 12 shows the SEA results based on the list of 10 volcano years from Ammann et al. (2007) and 14 volcano years from Fischer et al. (2007). In both cases, a statistically significant ($p < 0.01$) reduction in summer temperature is indicated beginning in the same year as the eruption and persisting for 2–4 more years. We estimate mean peak summer cooling due to these tropical eruptions of 0.3 to 0.4 °C (event year standard deviation is approximately 0.15 °C). Some of the apparent persistence in cooling following

eruption years may also be due to the closely spaced eruptions in the 13th and 18th centuries. A slight although statistically non-significant tendency for the appearance of mild cooling in the year prior to the designated event year may represent initial temperature anomalies before the onset of peak cold anomalies—for instance, Tambora erupted in boreal spring of 1815 C.E., but peak cooling and 'Year 0' (as defined by Fischer et al. (2007)) is 1816 C.E. Alternatively, the considerable uncertainty in the timing of eruptions and their radiative forcing may also contribute (Schmidt et al. 2012). It is also important to recognize that volcanic eruptions do not result in globally or even regionally uniform summer cooling (e.g. Briffa et al. 1998; Jones et al. 2003; Schneider et al. 2009; Anchukaitis et al. 2012) and that the state of the ENSO system in particular may in part determine the sign and magnitude of the resulting temperature anomalies in Asia (Zhang et al. 2012). These results are generally consistent with previous studies (c.f. Briffa et al. 1998; Breitenmoser et al. 2012), thus reinforcing our understanding of explosive volcanism as an important radiative forcing factor of climate over Asia (c.f. Anchukaitis et al. 2010).

Our reconstruction is comprised of only tree-ring data, as this proxy provides the time-certainty and precise dating necessary to perform split-sample calibration and valida-tion against the instrumental record. These data are also
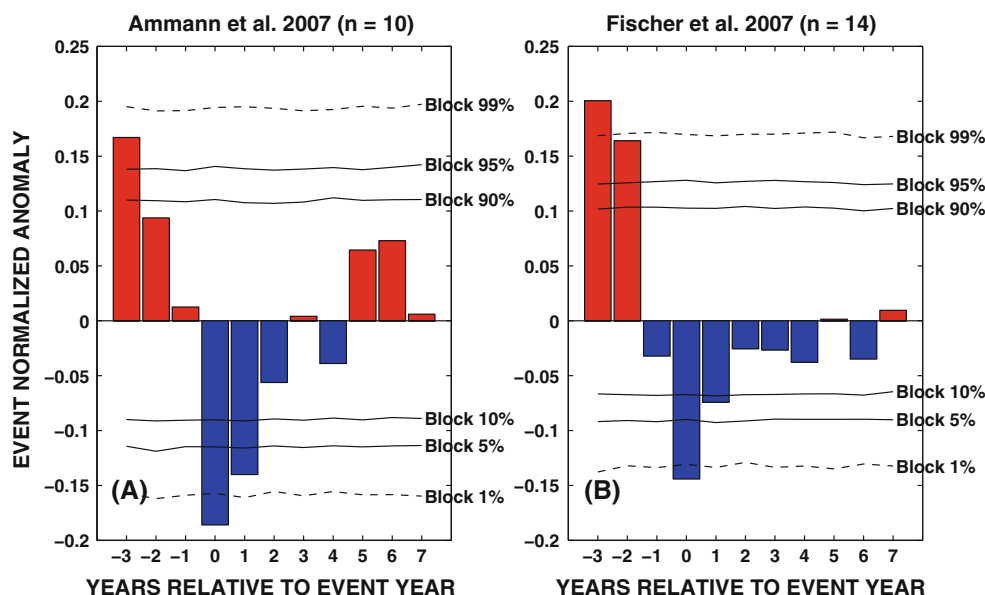

Fig. 12 Results of superposed epoch analysis (SEA) testing the impact of explosive volcanism on summer temperatures over temperate East Asia. Two key lists of volcanic eruption years, based on Ammann et al. (2007) [1258,1269,1278,1453,1600,1641,1809, 1816,1884, and 1903 C.E.] and Fischer et al. (2007) [1587, 1596,1600,1641,1674,1809,1816,1823,1832,1835,1884,1903,1964,1-983] were used as tests of robustness and a modified block resampling method (Adams et al. 2003) was used for testing statistical signifi-cance. For the SEA, reconstructed temperature values in each event

'window' were centered and divided by the maximum anomaly in that window in order to avoid giving any single eruption undue weight in the mean calculation (Adams et al. 2003). The results indicate a statistically significant cooling effect in the same year as the eruption, followed by 3–4 years of additional below average temperatures. Closely spaced eruptions in the 13th and 18th century may also be responsible for some of the apparent persistence in surface temper-ature declines

Fig. 13 Comparison with a multiproxy China temperature reconstruction. Our tree-ring reconstruction (10 years low pass filter applied to the Frozen 800 CE grid, see Fig. 10) compared to the composite decadal reconstructions for China using ice core, lake sediment, peat bog, tree-ring, and historical records by Yang et al. (2002). We plot both the 'high resolution' and 'weighted' reconstructions
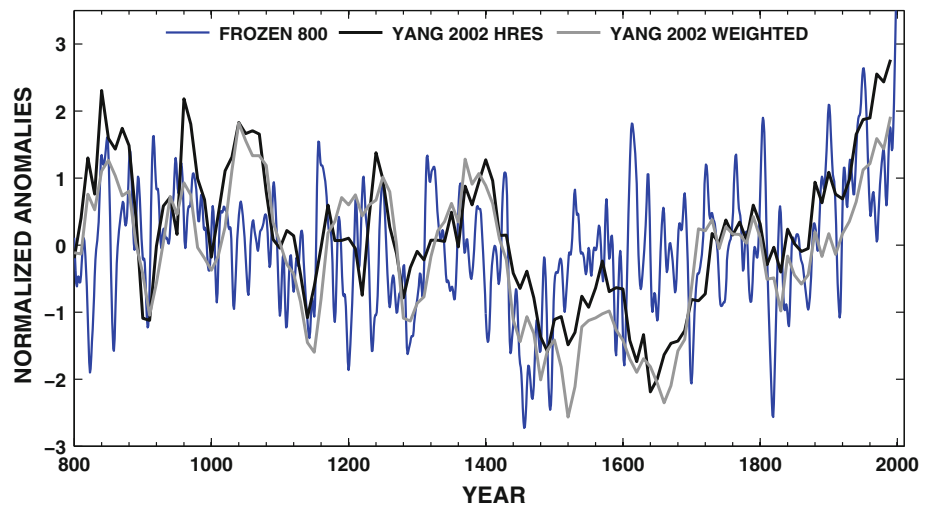
widely, albeit heterogeneously, distributed across our target domain. We can compare our reconstruction to the multiproxy estimates of past temperature in China by Yang et al. (2002). See Fig. 13. The two reconstructions show similar multicentury-scale behavior, with generally warmer conditions during the Medieval epoch, cooling into the Little Ice Age, and increase in temperature since the 19th century. There are some notable differences, however. Our tree-ring based reconstruction shows cooling concurrent with the mid 15th century eruption of Kuwae, while the multiproxy reconstruction suggests a delayed colder period in the late 15th or early 16th century. The reconstruction by Yang et al. (2002) also shows several decades of colder conditions in the 17th century that are not observed in our reconstruction. This difference might arise from the spatial distribution of the proxies, as this feature appears prominently in some proxies from eastern China and Japan, but not obviously in Himalayan ice cores. The multiproxy reconstruction appears to show more multidecadal to centennial variability; however, this may be a consequences of using fewer, lower resolution, and time-uncertain proxies.

5 Summary and conclusions

The boreal summer temperature reconstruction presented here for temperate East Asia is the primary contribution of the Asia2k project to the greater PAGES 2k Network project, whose goal is to generate a global array of regional temperature reconstructions covering the last 2000 years for climate change detection and attribution studies. The reconstruction covers the past 1210 years and is based solely on long annual tree-ring series from East Asia. Calibration and validation tests indicate that the reconstruction has a credible (if modest) level of skill, but substantial uncertainties of various kinds remain, particularly

at lower frequencies. This limits the value of the reconstruction for determining the degree to which rapid observed warming over East Asia since 1990 is statistically unprecedented because the estimated uncertainties are too large. Even so, the reconstruction has centennial time scale variability that is consistent with the occurrence of the Medieval Warm Period and Little Ice Age, as documented both in East Asia and elsewhere around the globe by others. In addition, our reconstruction further supports explosive volcanism as a significant forcing agent of regional summer temperature variability.

References

Adams JB, Mann ME, Ammann CM (2003) Proxy evidence for an El Nino-like response to volcanic forcing. Nature 426:274–278

Allen DM (1971) Mean square error of prediction as a criterion for selecting variables. Technometrics 13:469–475

Ammann CM, Joos F, Schimel DS, Otto-Bliesner BL, Tomas RA (2007) Solar influence on climate during the past millennium: results from transient simulations with the NCAR climate system model. Proc US Natl Acad Sci 104(10):3713–3718

Ammann CM, Genton MG, Li B (2010) Technical note: correcting for signal attenuation from noisy proxy data in climate reconstructions. Clim Past 6(2):273–279

Anchukaitis KJ, Buckley BM, Cook ER, Cook BI, D'Arrigo R, Ammann CM (2010) The influence of volcanic eruptions on the climate of the Asian monsoon region. Geophys Res Lett 37:L22703. doi:10.1029/2010GL044843

Anchukaitis KJ, Breitenmoser P, Briffa KR, Buchwal A, Büntgen U, Cook ER, D'Arrigo RD, Esper J, Evans MN, Frank D, Grudd H,

Gunnarson B, Hughes MK, Kirdyanov AV, Körner C, Krusic PJ, Luckman B, Melvin TM, Salzer MW, Shashkin AV, Timmreck C, Vaganov EA, Wilson RJ (2012) Tree rings and volcanic cooling. Nat Geosci 5:836–837. doi:10.1038/ngeo1645

Breitenmoser P, Beer J, Brönnimann S, Frank D, Steinhilber F, Wanner H (2012) Solar and volcanic fingerprints in tree-ring chronologies over the past 2000 years. Palaeogeogr Palaeoclimatol Palaeoecol 313–314:127–139

Briffa KR, Jones PD, Schweingruber FH, Osborn TJ (1998) Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. Nature 393:450–455

Cook ER, Peters K (1981) The smoothing spline: a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies. Tree-Ring Bull 41:45–53

Cook ER, Briffa KR, Meko DM, Graybill DA, Funkhouser G (1995) The segment length curse in long tree-ring chronology development for paleoclimatic studies. Holocene 5:229–237

Cook ER, Meko DM, Stahle DW, Cleaveland MK (1999) Drought reconstructions for the continental United States. J Clim 12(4):1145–1162

Cook ER, Esper J, D'Arrigo RD (2004) Extra-tropical northern hemisphere land temperature variability over the past 1000 years. Quat Sci Rev 23:2063–2074

Cook ER, Woodhouse CA, Eakin CM, Meko DM, Stahle D (2004) Long-term aridity changes in the western United States. Science 306(5698):1015–1018

Cook ER, Anchukaitis KJ, Buckley BM, D'Arrigo RD, Jacoby GC, Wright WE (2010) Asian monsoon failure and megadrought during the last millennium. Science 328(5977):486–489. doi:10.1126/science.1185188

Cook ER, Seager R, Heim RR, Vose RS, Herweijer C, Woodhouse CA (2010) Megadroughts in North America: placing IPCC projections of hydroclimatic change in a long-term paleoclimate context. J Quat Sci 25(1):48–61

Dai A, Trenberth KE, Qian T (2004) A global dataset of palmer drought severity index for 1870–2002: relationship with soil moisture and effects of surface warming. J Hydrometeorol 5(6):1117–1130

Fischer EM, Luterbacher J, Zorita E, Tett SFB, Casty C, Wanner H (2007) European climate response to tropical volcanic eruptions over the last half millennium. Geophys Res Lett 34(5):L05707. doi:10.1029/2006GL027992

Gao CC, Robock A, Self S, Witter JB, Steffenson JP, Clausen HB, Siggaard-Andersen ML, Johnsen S, Mayewski PA, Ammann C (2006) The 1452 or 1453 AD Kuwae eruption signal derived from multiple ice core records: greatest volcanic sulfate event of the past 700 years. J Geophys Res Atmos 111:D12107

Gnanadesikan R (1997) Methods for statistical data analysis of multivariate observations. Wiley, New York

Haurwitz MW, Brier GW (1981) A critique of the superposed epoch analysis method—its application to solar-weather relations. Mon Weather Rev 109(10):2074–2079

IPCC (2007) Summary for Policymakers. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge

University Press, Cambridge, United Kingdom and New York, NY, USA

Jones P, Moberg A, Osborn T, Briffa K (2003) Surface climate responses to explosive volcanic eruptions seen in long European temperature records and mid-to-high latitude tree-ring density around the Northern Hemisphere. In: Robock A, Oppenheimer C (eds) Volcanism and the Earth's Atmosphere. American Geophysical Union, Washington, pp 239–254

Jones PD, Osborn TJ, Briffa KR (1997) Estimating sampling errors in large-scale temperature averages. J Clim 10:2548–2568

Ljungqvist FC, Krusic PJ, Brattstrom G, Sundqvist HS (2012) Northern Hemisphere temperature patterns in the last 12 centuries. Clim Past 8:227–249. doi:10.5194/cp-8-227-2012

Lorenz EN (1956) Empirical orthogonal functions and statistical weather prediction. Statistical Forecasting Scientific Report 1, Massachusetts Institute of Technology, Department of Meteorology

Mann ME, Zhang Z, Rutherford S, Bradley RS, Hughes MK, Shindell D, Ammann C, Faluvegi G, Ni F (2009) Global signatures and dynamical origins of the little ice age and medieval climate anomaly. Science 326(5957):1256–1260

Melvin T, Briffa K (2008) A 'signal-free' approach to dendroclimatic standardisation. Dendrochronologia 26(2):71–86

Mitchell T, Jones P (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int J Climatol 25(6):693–712

Morice CP, Kennedy JJ, Rayner NA, Jones PD (2012) Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set. J Geophys Res 117:D08101

Schmidt GA, Jungclaus JH, Ammann CM, Bard E, Braconnot P, Crowley TJ, Delaygue G, Joos F, Krivova NA, Muscheler R, Otto-Bliesner BL, Pongratz J, Shindell DT, Solanki SK, Steinhilber F, Vieira LEA (2012) Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.1). Geosci Model Dev 5(1):185–191

Schneider DP, Ammann CM, Otto-Bliesner BL, Kaufman DS (2009) Climate response to large, high-latitude and low-latitude volcanic eruptions in the community climate system model. J Geophys Res 114(D15101):D15101. doi:10.1029/2008JD011222

Stott PA, Gillett NP, Hegerl GC, Karoly DJ, Stone DA, Zhang X, Zwiers F (2010) Detection and attribution of climate change: a regional perspective. Wiley Interdiscip Rev Clim Change 1(2):192–211

Touchan R, Anchukaitis KJ, Meko DM, Sabir M, Attalah S, Aloui A (2011) Spatiotemporal drought variability in northwestern Africa over the last nine centuries. Clim Dyn 37(1–2):237–252. doi:10.1007/s00382-010-0804-4

Wigley TML, Briffa KR, Jones PD (1984) On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. J Clim Appl Meteorol 23:201–213

Yang B, Braeuning A, Johnson KR, Yafeng S (2002) General characteristics of temperature variation in china during the last two millennia. Geophys Res Lett 29(9):1324. doi:10.1029/2001GL014485

Zhang D, Blender R, Fraedrich K (2012) Volcanoes and ENSO in millennium simulations: global impacts and regional reconstructions in East Asia. Theor Appl Climatol, 1–18