Tree-Ring-Reconstructed Summer Temperatures from Northwestern North America during the Last Nine Centuries*

KEVIN J. ANCHUKAITIS,† ROSANNE D. D’ARRIGO, AND LAIA ANDREU-HAYLES‡

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

DAVID FRANK AND ANNE VERSTEGE

Swiss Federal Research Institute WSL, Birmensdorf, Switzerland, and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

ASHLEY CURTIS, BRENDAN M. BUCKLEY, GORDON C. JACOBY, AND EDWARD R. COOK

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

(Manuscript received 9 March 2011, in final form 31 May 2012)

ABSTRACT

Northwestern North America has one of the highest rates of recent temperature increase in the world, but the putative “divergence problem” in dendroclimatology potentially limits the ability of tree-ring proxy data at high latitudes to provide long-term context for current anthropogenic change. Here, summer temperatures are reconstructed from a Picea glauca maximum latewood density (MXD) chronology that shows a stable relationship to regional temperatures and spans most of the last millennium at the Firth River in northeastern Alaska. The warmest epoch in the last nine centuries is estimated to have occurred during the late twentieth century, with average temperatures over the last 30 yr of the reconstruction developed for this study [1973–2002 in the Common Era (CE)] approximately 1.3° ± 0.4°C warmer than the long-term preindustrial mean (1100–1850 CE), a change associated with rapid increases in greenhouse gases. Prior to the late twentieth century, multidecadal temperature fluctuations covary broadly with changes in natural radiative forcing. The findings presented here emphasize that tree-ring proxies can provide reliable indicators of temperature variability even in a rapidly warming climate.

1. Introduction

Instrumental climate observations indicate that surface air temperatures in northwestern North America are rising faster than nearly anywhere else in the world (ACIA 2004). Hartmann and Wendler (2005) found...
Developing accurate estimates of past temperature variability requires us to address four fundamental challenges to characterizing the full magnitude of past climate change from tree-ring proxies. First, we require a tree-ring proxy chronology with a robust, stable, and biophysically plausible relationship to our climate metric of interest. Those chronologies that demonstrate evidence of the divergence problem (Briffa et al. 1998; Pisaric et al. 2007; D’Arrigo et al. 2008) could potentially underestimate temperature change during putative prior warm periods. The divergence problem is not, however, ubiquitous, and tree-ring proxies at many sites do continue to track local temperatures (e.g., Wilmking et al. 2005; Wilson et al. 2007; Briffa et al. 2008; Büngeten et al. 2008; Esper and Frank 2009; Esper et al. 2010). Second, some latitudinal treeline species can be relatively short lived and may therefore potentially limit the spectral fidelity of any reconstruction due to the “segment length curse” (Cook et al. 1995), which when using data-adaptive growth curves fit to the individual series mathematically constrains the resolved frequency range of the resulting chronology to be no lower than that corresponding to the mean length of the individual series. Third, and related to both of these issues, is the need to avoid imposing time series end effects that can artificially modify trends over the last few decades of the tree-ring proxy chronology through the process of detrending and standardization, which itself can lead to an apparent but erroneous divergence in the proxy–climate relationship that is purely statistical in nature (Melvin and Briffa 2008; Esper and Frank 2009; Briffa and Melvin 2011). Finally, it is necessary for this particular application to avoid calibration methods that result in variance loss in the final reconstruction and could therefore underestimate the actual magnitude of past centennial-scale climate variability (e.g., see Ammann et al. 2010).

Here, we describe a divergence-free maximum late-wood density (MXD) Picea glauca chronology from the Firth River (Fig. 1) in northeastern Alaska that spans much of the past millennium. Our objective is to develop a proxy chronology and subsequent reconstruction that minimizes the statistical hazards described above. The resulting reconstruction is used to assess variability and long-term changes in regional temperatures, can be integrated in regional or hemisphere-scale multiproxy assessments, and may be applied to assess the local response to radiative forcing (e.g., Voulgarakis and Shindell 2010).

2. Materials and methods

The Firth River Picea glauca (white spruce) MXD chronology is composed of 246 radii from 103 living and subfossil trees growing on hill slopes in sparsely populated groves on either side of the open parklike valley and spanning 1073–2002 CE (930 yr) with a mean segment length of 191 yr (range 54–457 yr). These are some of the northernmost trees in North America (Fig. 1). The ring-width chronology was originally developed at the Lamont-Doherty Earth Observatory Tree Ring Laboratory following standard procedures (Stokes and Smiley 1968; Fritts 1976). Wood density was measured at the Swiss Federal Research Institute WSL in Birmensdorf, Switzerland, using a Dendro 2003 (Walesch) X-ray densitometer system (Eschbach et al. 1995) and following the procedures originally described by Schweingruber et al. (1978) and Schweingruber (1988).

The MXD data are significantly correlated ($p < 0.01$) with summer temperatures from weather stations and gridded observational records during the last century (Fig. 2). This finding is consistent with other studies of MXD (e.g., Schweingruber 1988; D’Arrigo et al. 1992; Davi et al. 2003; Frank and Esper 2005; D’Arrigo et al. 2009), and is due to the influence of summer climate on the thickness of the cell wall in the latewood of the annual ring (Yasue et al. 2000). We select for our reconstruction target the mean monthly July and August temperatures from the Goddard Institute for Space Studies (GISS) Surface Temperature Analysis (GISTEMP; Hansen et al. 2010) 1200-km grid cell associated with the Firth River site, since it is continuous, long, represents a large-scale regional mean, and can be readily compared to climate
model output. The GISTEMP values are temperature anomalies from the base period 1951–80. There is no evidence of a decay in the relationship between the proxy and seasonal climate [Figs. 3a,b; see also Andreu-Hayles et al. (2011)]; indeed, the relationship is somewhat stronger over the most recent decades as more proximal station data become available. The Firth River chronology correlates significantly with temperatures over a substantial portion of northwestern North America, with the strongest correlations observed close to the site, decaying to nonsignificant values in the central Canadian Arctic (Fig. 4a). Similar large-scale spatial correlation is observed for the Coppermine and Thelon MXD chronologies from the Canadian treeline (Figs. 4b,c; D’Arrigo et al. 2009). In practice, the selection of the reconstruction target has minimal influence on the interpretation of the relative transient climate history of the region.

To preserve lower-frequency variability in tree-ring proxy time series, dendroclimatologists have increasingly utilized regional curve standardization (RCS), which removes the biological age trend from each series using a common mean growth curve calculated from all individual trees (Mitchell 1967; Briffa et al. 1992) and that can therefore in theory preserve variance at periods longer than the mean length of these series. However, RCS can also potentially result in spurious low-frequency variability as well as modern trend distortion (Briffa and Melvin 2011). A “signal free” (SF) detrending procedure (Melvin 2004; Melvin and Briffa 2008) has recently been developed that seeks to prevent trend distortion by iteratively removing an estimated common temporal growth signal from the individual measurement series. With the common variance removed, growth curves can be fit to the now signal-free chronologies without also removing likely common climatic information. We find here that if the means of the individual series are not reset to unity at every iteration of the algorithm, then SF detrending can also potentially preserve variations at wavelengths longer than the mean segment length (see below), although caution is warranted because it is still possible to impart spurious low-frequency variability simply because the means and lengths of the individual tree-ring series are different (Melvin and Briffa 2008).

Although recognized in theory, biases induced by detrending methods cannot be known with certainty a priori or from their application to the actual tree-ring data alone. To evaluate whether traditional, RCS or the signal-free detrending of the Firth River chronology are able to preserve low-frequency variability without also imparting artifacts, we conducted a set of pseudoproxy experiments (cf. Evans et al. 1998; Mann and Rutherford 2002; Bunn et al. 2004; Smerdon et al. 2010; Smerdon 2011). As Sherwood et al. (2009) suggest, “methods for reconstructing climate should be evaluated by testing them on simulated datasets constructed as realistically as possible, including suspected proxy/instrumental artifacts or biases and sampling patterns.” Here, we generate “psuedoclimate” time series (C) meant to mimic a possible, realistic transient climate signal. We then fit a straight line (A) to each individual MXD series from the Firth River site. We use both of these to then create simulated density series (R) where the values in year t can be expressed as the linear combination of the (known) pseudoclimate signal, the (known) growth curve, and random noise (ε):

$$R_t = A_t + C_t + \epsilon_t.$$  

We then applied the same RCS, signal-free, and individual data adaptive curve fitting procedures to this
set of artificial $R$ as we would for the actual Firth River MXD data. Note that our artificial test data also have the same sample depth as the actual data. We can then evaluate the difference between the estimated ($\hat{C}$) and known ($C$) pseudoclimate signals, which we use to assess the possible consequences of our choice of detrending methods on low-frequency variability and trend distortion in the real data, where the answer—the actual transient climate evolution of the last millennium—is naturally unknown. Such a simple experiment may not characterize all the biases potentially presented by the actual proxy data; however, pathological behavior of a particular detrending approach under the simplest of conditions is unlikely to inspire confidence in the more complicated real data. For our pseudoproxy experiments, we evaluate three pseudoclimate signals: the mean annual global and the summer local temperature derived from the Climate System Model (CSM1.4) forced last millennium simulation (Ammann et al. 2007) and a random Gaussian $[0, 1]$ series. These three experiments provide a useful spectral range of variability against which to evaluate potential biases in detrending methods. In our experiments, $\epsilon$ is specified as Gaussian noise giving a (pseudoclimate) signal-to-noise ratio of 0.5, which yields correlations between the simulated tree-ring data and the pseudoclimate signal close to the observed [Smerdon (2011), and see Fig. 2]. We conducted 1000 experiments with different random noise draws in order to develop confidence intervals around the pseudoproxy series. Transformation of our pseudoproxy chronologies into estimates of the embedded pseudoclimate signal used classical calibration, as described below.

The problem of regression dilution is a well-recognized consequence of statistical calibration with noisy predictors (e.g., Osborne 1991; Ammann et al. 2010; Christiansen 2011; Tingley et al. 2012). “Inverse” calibration (cf. Birks 1995), where the physical sense of the regression relationship is reversed to make the climate

\[ 3004 \text{ JOURNAL OF CLIMATE VOLUME 26} \]

**FIG. 3.** (a) Time series and (b) scatterplot comparisons between GISTEMP July–August mean temperature (Hansen et al. 2010) and the Firth River MXD chronology during the instrumental period. The GISTEMP data are expressed as anomalies from the mean of the period 1951–80. In (b), colors indicate the year of observation.

**FIG. 4.** Spatial correlation fields for North American treeline MXD chronologies [(a) Firth River from this study, and (b) Coppermine and (c) Thelon from D’Arrigo et al. (2009)] and the July–August GISTEMP 1200-km gridded temperature field (Hansen et al. 2010) for the period 1897–2002. The locations of each chronology are shown by stars and the respective time series are shown in Fig. 8.
a function of the proxy, has the benefit of minimizing the root-mean-squared error (RMSE) of the estimate, but the slope of the relationship between proxy and climate attenuates toward zero (Ammann et al. 2010; Tingley and Huybers 2010; Tingley et al. 2012). For temperature reconstructions the consequence can be a preindustrial reconstruction that is biased too warm (Tingley et al. 2012). Here, we utilize ordinary least squares for classical calibration, where the regression coefficients are calibrated from the causal relationship between proxy predictand and climate predictor and then inverted to estimate the temperature from MXD. The consequence is a set of reconstructed values that do not minimize the RMSE between proxy and climate and that inflate the reconstructed interannual variance, but that do minimize regression dilution and can potentially more accurately reflect the amplitude of the past multidecadal and centennial temperature variability (cf. Christiansen et al. 2009). The latter is the difference between the means of the observed and reconstructed temperatures during the validation period and is therefore an estimate of the error in the lower-frequency, multidecadal-scale reconstructed temperatures. The former is a standard measure of the high-frequency fidelity between reconstruction and observed temperatures. We also use a Monte Carlo approach to benchmark both of these statistics against the performance of “nonsense” predictors consisting of autoregressive (AR1) noise conditioned to reflect the spectral characteristics of the actual proxy data (Schneider and Neumaier 2001; Bürger 2007). Following Frank et al. (2010), we perform an ensemble of reconstructions, evaluating all possible consecutive 53-yr calibration periods during the full instrumental period (1897–2002) and validating against the remaining, withheld 53 yr. This yields 54 separate reconstructions, which allows us to assess the influence of the calibration period choice and changes in instrumental data availability and quality on the magnitude of inferred past temperature variability, as well as to quantify an additional source of uncertainty. Because of this ensemble approach, the distributions of the calibration and validation $R^2$ statistic are the same.

3. Results and discussion

As discussed by Kutzbach et al. (2010), some traditional metrics used in dendroclimatology for estimating reconstruction skill—including the reduction of error (RE) and coefficient of efficiency (CE)—may not be applicable here because classical calibration, while preventing regression dilution, does not minimize the RMSE and artificially inflates the interannual variance. We therefore evaluate the out-of-sample skill of our reconstruction using the validation-period coefficient of determination ($R^2$) and the bias (cf. Christiansen et al. 2009). The latter is the difference between the means of the observed and reconstructed temperatures during the validation period and is therefore an estimate of the error in the lower-frequency, multidecadal-scale reconstructed temperatures. The former is a standard measure of the high-frequency fidelity between reconstruction and observed temperatures. We also use a Monte Carlo approach to benchmark both of these statistics against the performance of “nonsense” predictors consisting of autoregressive (AR1) noise conditioned to reflect the spectral characteristics of the actual proxy data (Schneider and Neumaier 2001; Bürger 2007). Following Frank et al. (2010), we perform an ensemble of reconstructions, evaluating all possible consecutive 53-yr calibration periods during the full instrumental period (1897–2002) and validating against the remaining, withheld 53 yr. This yields 54 separate reconstructions, which allows us to assess the influence of the calibration period choice and changes in instrumental data availability and quality on the magnitude of inferred past temperature variability, as well as to quantify an additional source of uncertainty. Because of this ensemble approach, the distributions of the calibration and validation $R^2$ statistic are the same.

3. Results and discussion

As discussed by Kutzbach et al. (2010), some traditional metrics used in dendroclimatology for estimating reconstruction skill—including the reduction of error (RE) and coefficient of efficiency (CE)—may not be applicable here because classical calibration, while preventing regression dilution, does not minimize the RMSE and artificially inflates the interannual variance. We therefore evaluate the out-of-sample skill of our reconstruction using the validation-period coefficient of determination ($R^2$) and the bias (cf. Christiansen et al. 2009). The latter is the difference between the means of the observed and reconstructed temperatures during the validation period and is therefore an estimate of the error in the lower-frequency, multidecadal-scale reconstructed temperatures. The former is a standard measure of the high-frequency fidelity between reconstruction and observed temperatures. We also use a Monte Carlo approach to benchmark both of these statistics against the performance of “nonsense” predictors consisting of autoregressive (AR1) noise conditioned to reflect the spectral characteristics of the actual proxy data (Schneider and Neumaier 2001; Bürger 2007). Following Frank et al. (2010), we perform an ensemble of reconstructions, evaluating all possible consecutive 53-yr calibration periods during the full instrumental period (1897–2002) and validating against the remaining, withheld 53 yr. This yields 54 separate reconstructions, which allows us to assess the influence of the calibration period choice and changes in instrumental data availability and quality on the magnitude of inferred past temperature variability, as well as to quantify an additional source of uncertainty. Because of this ensemble approach, the distributions of the calibration and validation $R^2$ statistic are the same.

3. Results and discussion

The Firth River MXD chronology (Fig. 5a) has a mean pairwise interseries correlation of $r_{tot} = 0.33$ and a between-tree mean correlation ($r_{bt}$) of 0.40 (Cook 1985; Cook and Kairiukstis 1990). The crossdating software COFECHA, which computes the correlation between each core series and an average chronology based on all the other cores, reports a mean interseries correlation of 0.64. The chronology demonstrates a common high-frequency signal is therefore present, but the low sample depth
prior to \( \sim 1220 \) CE (Fig. 5d) requires substantial caution when interpreting the record during that time.

The low-frequency component of the reconstruction is dependent on the choice of detrending technique. Pseudoproxy experiments (Figs. 6a–c) show that RCS consistently imparted the “contemporaneous growth rate” bias identified by Briffa and Melvin (2011), irrespective of the pseudoclimate time series. This resulted in a positive bias from \( \sim 1250 \) to 1600 CE, and a negative bias between \( \sim 1100 \) and 1250 CE. SF detrending appears to be less biased for the series length, sample size, and growth trajectory of these data, although RCS and SF are quite similar if considered over only the most recent centuries. “Standard” detrending using data-adaptive curves fit to each individual series here unavoidably falls prey to the segment length curse when substantial low-frequency variability is present. These features also appear to be present when the different detrending techniques are applied to the actual MXD data (Fig. 6d). Because our experiments suggest SF is the least biased of the detrending choices for these data, we applied the technique with linear growth curves to the Firth River MXD data to develop the final master chronology (Fig. 5a).

The ensemble reconstructed calibration and validation mean \( R^2_{cv} \) is 0.36 with a range of 0.21–0.44, which is larger than and statistically different (\( p < 0.01 \)) from the “nonsense” red noise predictors, which showed a mean of 0.02 and ranged from 0.00 to 0.19. The mean bias during the validation period is \(-0.18^\circ C\) and the ensemble range is from \(-0.38^\circ C\) to \(0.26^\circ C\), reflecting the size of the mean multidecadal error. Bias for nonsense predictors span zero but have a 95% two-tailed range of \(-13.1^\circ C\) to \(14.3^\circ C\) with an absolute median bias of \(1.2^\circ C\). We interpret these results to indicate that the reconstruction has greater skill than expected by chance when evaluated against red noise, but depending on the choice of calibration and validation periods may have an error at multidecadal time scales of several tenths of a degree Celsius. The ensemble members include results that are biased both too warm and too cold, depending on the calibration and validation period. The ensemble mean 30-yr smoothed standard error of prediction over the instrumental period is estimated to be \(0.33^\circ C\) with an ensemble range of \(0.27^\circ C\)–\(0.48^\circ C\).

Based on our reconstruction (Fig. 7a), the warmest epoch in the last nine centuries occurs during the twentieth century, with average summer temperatures over the last 30 yr of our reconstruction (1973–2002 CE) estimated to be approximately \(1.3^\circ C\pm 0.4^\circ C\) warmer than the long-term preindustrial mean (1100–1850 CE). These \(1\sigma\) ranges account for the errors in the calculation of epochal differences due to the estimated uncertainties in the reconstruction. The range of preindustrial to modern summer temperature change in our reconstruction ensemble members is approximately \(1.0^\circ C\)–\(2.0^\circ C\). The coldest sustained period occurs at the same time as the Maunder Minimum (1645–1715 CE), and was approximately \(1.8^\circ C\pm 0.4^\circ C\) colder than the last three decades of the reconstruction (1973–2002 CE). Our reconstruction ensemble reveals a range of Maunder Minimum to modern summer temperature change from \(1.3^\circ C\) to \(2.3^\circ C\). One of the coldest individual years in the Firth River chronology...
is 1783 CE, corresponding to the eruption of Laki (Jacoby et al. 1999) in Iceland [but see D’Arrigo et al. (2011)].

To evaluate common broad-scale northwestern North American summer temperature trends over the last millennium, we can compare our estimates to those from other proxy records from the region. Our temperature reconstruction has features similar to those observed in the lake sediment varve thickness record by Bird et al. (2009) from Blue Lake near Atigun Pass in Alaska and the multiproxy Arctic temperature reconstruction from Kaufman et al. (2009) (Figs. 7b,c). All three of these records indicate a shift toward colder conditions in the seventeenth to nineteenth centuries, although there is a period of relative warmth in the Firth River record in the midst of the eighteenth century. The three records also show that some of the coldest temperatures of the last millennium occur at the time of the Maunder Minimum, and a sustained rise in temperature from the midnineteenth century to the present, when summer temperatures appear to be higher than the preceding nine centuries, although the varve record from Blue Lake does suggest a comparable warm epoch near 800 CE (Bird et al. 2009). In southern Alaska at Hallet Lake (Fig. 7d), McKay et al. (2008) found the coldest summer temperatures of the Common Era between 1750 and 1900 CE, and estimated a $\sim 2.5^\circ \pm 0.8^\circ$C rise in temperatures over the last 130 yr, with the last two decades as the warmest of the prior two millennia. There are also similarities between the Firth River and the hemisphere-scale multivariate reconstruction by D’Arrigo et al. (2006; Fig. 7e), including local minima at the time of the Maunder Minimum and the early nineteenth century, an upward temperature trend beginning in the middle of the nineteenth century, and an earlier twentieth century warm period. There are notable differences, however, including an earlier cold epoch in the fifteenth century and a deeper cooling during the Maunder Minimum at the Firth River.

Other tree-ring density records from the North American treeline suggest broad-scale common multidecadal and centennial climate variability during the last several centuries (Fig. 8). D’Arrigo et al. (2009) developed two MXD records from northern Canada that also do not appear to show divergence from the instrumental temperature record (Figs. 4 and 8). All three records suggest local summer temperature minima in the early nineteenth century, at the same time as the solar Dalton Minimum and several strong volcanic events (Crowley 2000; Wagner and Zorita 2005), although the onset of overall cooler conditions at the Firth River appears to begin earlier, in the 1780s. All three records suggest increases in temperature beginning in the late the nineteenth century, an early twentieth century period of warmer temperatures, and the highest estimated temperature at the end of the twentieth century. Thus, although quite widely separated in space, these three MXD chronologies have common features at lower frequencies as well as evidence for similar temperature responses over the instrumental era (Fig. 4).

Episodic sustained colder periods in the past tend to be associated with major minima in insolation (Figs. 9b,c). Multidecadal-scale cold anomalies occur concomitantly

**Fig. 7.** Summer temperature reconstructions from the last millennium. (a) Summer (July–August) weighted mean ensemble Firth River temperature reconstruction (black), smoothed with a 30-yr Gaussian filter. Values are anomalies from the 1951–80 mean (Hansen et al. 2010), and the ensemble mean is developed by weighting the ensemble methods by their validation period bias such that more skillful reconstructions receive a larger weight for the ensemble mean. Light blue shading indicates ±2σ uncertainties estimated as the standard error of prediction (Wilks 1995). The range of all ensemble members is shown by the dark blue shading. (b) Decadal average standardized varve thickness, inferred to be a summer temperature proxy, from Blue Lake (Bird et al. 2009) near Atigun Pass, as calculated by Kaufman et al. (2009) (see Fig. 1). (c) Reconstructed Arctic summer temperature from Kaufman et al. (2009). (d) Biogenic silica (BSi) concentrations from Hallet Lake (McKay et al. 2008). The data have been log transformed. (e) Reconstructed (gray) and instrumental (black) Northern Hemisphere (20°–90°N) land-only annual temperature from D’Arrigo et al. (2006). The reconstructed climate in gray reflects the bounds from the standard negative-exponential and regional curve standardized time series.
with the Spörer, Maunder, and Dalton Minima, although there is no sign in the Firth River reconstruction of a climate response to the earlier Wolf Minima. Similarly though, Wiles et al. (2004) documented Alaskan glacier advances during the Dalton, Spörer, and Maunder Minima, but not the Wolf. Several large volcanic eruptions and a period of sustained negative radiative forcing also occurred during the Spörer and Dalton Minima (Fig. 9a; Crowley 2000; Schmidt et al. 2011). Rapid warming is observed in our reconstruction from the middle of the nineteenth century toward the present, concurrent with an increase in both insolation and well-mixed greenhouse gases, although the increased radiative forcing due to greenhouse gases is an order of magnitude larger over that time scale. Consistent with instrumental records, summer warming has continued throughout the twentieth century, as have increases in greenhouse gas concentrations (Fig. 9d), while solar forcing has reached a plateau. Gillett et al. (2008) attributed overall twentieth-century temperature trends in the Arctic to anthropogenic forcing, and showed that only climate models that included anthropogenic forcing reproduced the twentieth-century temperature trends over Alaska.

There is little sign of a distinct Medieval Warm Period sensu lato (cf. Seager et al. 2008) in the Firth River reconstruction, but the chronology does not span the entire “High” Medieval Warm Period sensu stricto (Bradley et al. 2001), nor does it contain the peak in temperatures near 1000 CE seen in most hemisphere-scale reconstructions (cf. Jansen et al. 2007; Kaufman et al. 2009; Frank et al. 2010); rather, the Little Ice Age conditions after the seventeenth century are perhaps the most distinct of the preindustrial centennial features in the record. The proxy reconstructions in Figs. 7 and 8 all do reflect a period of previously observed warmer temperatures in the early half of the twentieth century (Bengtsson et al. 2004; Overland et al. 2004; Wood and Overland 2010). Collectively, the coherence between proxy records at decadal time scales and longer suggests that large-scale regional climate variability is being captured by multiple proxy indicators at a number of Arctic locations.

While our pseudoproxy assessment of the SF technique leads us to infer that it represents the least biased standardization choice available for these particular
tree-ring data, by far the largest source of methodological uncertainty in our reconstruction is the detrending technique. This is an issue common to most tree-ring-based approaches to past temperature reconstruction (cf. Frank et al. 2007) and reflects the challenge and importance of removing nonclimatic noise from the individual data series while preserving the actual low-frequency climate variability and avoiding creating statistical artifacts. Our tests of the SF methodology here give no indication that it creates any centennial-scale artifacts similar to those seen in the RCS pseudoproxies. However, the ensemble pseudoproxy tests do reveal that it is still possible for SF detrending to either overestimate or underestimate the magnitude of past temperature change, depending on the specific transient noise structure. The comparisons to other independent proxies do increase our confidence that the SF detrending is a reasonable choice and results in a long-term climate reconstruction with a similar magnitude of secular variability. Signal-free detrending is a new standardization technique (Melvin and Briffa 2008), and its use and application will continue to require investigation.

Why does the Firth River MXD chronology not show divergence from local summer temperatures? While one hypothesis might be that the use of the signal-free detrending prevents divergence (Melvin and Briffa 2008), at other sites at the North American treeline we have also observed stationary relationships between temperature and MXD (D’Arrigo et al. 2009) using standard detrending methods. Our pseudoproxy tests (Fig. 6) indicate negligible bias in accurately estimating recent temperature trends when using RCS and standard data adaptive approaches. Grudde (2008) likewise found a stable relationship between MXD and instrumental temperatures at Torneträsk, Sweden, at both interannual and lower frequencies when using RCS. Indeed, as we note in the introduction, the divergence problem is not ubiquitous and many sites with significant relationships to local climate maintain these to the present day, as the Firth River record does. Although several possible causes have been proposed to explain nonstationarity in proxy–climate relationships (e.g., Briffa et al. 1998; D’Arrigo et al. 2008; Esper and Frank 2009), at the Firth River and other sites along the North American treeline at least, there is no sign that climate controls on MXD have degraded, nor is there evidence for diverging trends at lower frequencies. Thus, while the choice of detrending is clearly important for correctly characterizing the long-term transient temperature history of the Common Era prior to the instrumental period, at the Firth River at least the detrending choice does not determine the fidelity of the interannual association between MXD and climate during the most recent decades. There is also no evidence that changes in atmospheric CO₂ directly influence MXD in the Firth River chronology. The relationship between July and August temperatures is stable and there is no indication that another factor has come to influence tree growth variability over recent decades. This observation is supported by idealized 2 × CO₂ growth experiments where the overall effect on MXD is equivocal, detected inconsistently across species, years, and site characteristics (cf. Hättenschwiler et al. 1996; Telewski et al. 1999; Kilpeläinen et al. 2003; Kostilaainen et al. 2004; Yazaki et al. 2005).

4. Summary and conclusions

By attempting to circumvent potential statistical hazards in reconstructing past temperature variability from high-latitude tree-ring proxies, we are able to estimate the magnitude of recent Arctic warming and the temperature history of northeastern Alaska over much of the last millennium. While from 1400 to 1850 CE the low-frequency reconstructed temperature variability qualitatively follows solar forcing, twentieth-century temperature increases in northwestern Alaska do appear to be the highest in at least the last nine centuries and are associated with substantial simultaneous increases in well-mixed greenhouse gases (Gillet et al. 2008). Our reconstruction provides data of relevance for comparison against last millennium forced climate model simulations (Schmidt 2010; Schmidt et al. 2011) and provides quantitative data that could be used to evaluate regional climate sensitivity (cf. Voulgarakis and Shindell 2010). The Firth River data can also be included in climate field reconstructions of the last millennium. Our study reiterates that carefully collected and analyzed tree-ring data can indeed provide quantitative information about the climate evolution of the Common Era, despite the specter of the “divergence problem.”

Acknowledgments. We thank Chris Buckley for additional field assistance. We have benefited greatly from methodological discussions organized by the European Union project MILLENNIUM (Contract 017008), particularly with Hiro Yamazaki, Anders Moberg, Eduardo Zorita, Phillip Brohan, and Myles Allen. We thank three anonymous reviewers for providing comments and critiques that helped substantially improve the manuscript. We are also grateful to the State of Alaska, the Arctic National Wildlife Refuge (ANWR), and the U.S. Fish and Wildlife Service for permission to work at the Firth River. This research was supported by the U.S. National Science Foundation (ARC-0902051 and ATM-0211583), as well as the Swiss National Science Foundation (NCCR-Climate and INTEGRAL 121859).


