Dendrochronologia 48 (2018) 52–73

Contents lists available at ScienceDirect

Dendrochronologia

journal homepage: www.elsevier.com/locate/dendro

A comparison of some simple methods used to detect unstable temperature responses in tree-ring chronologies

K.J. Allen[a,*], R. Villalba[b], A. Lavergne[c], J.G. Palmer[d], E.C. Cook[e], P. Fenwick[f], D.M. Drew[g], C.S.M. Turney[d], P.J. Baker[a]

[a] School of Ecosystem and Forest Sciences, University of Melbourne, 500 Yarra Boulevard, Richmond, Victoria, 3121, Australia
[b] Instituto Argentino de Nivología, Glaciología y Ciencias Ambientales, CONICET, CCT-Mendoza, Casilla de Correo 330, 5500 Mendoza, Argentina
[c] Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, SL5 7PY, United Kingdom
[d] Climate Change Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia
[e] Lamont-Doherty Earth Observatory, Palisades New York, 10964, USA
[f] Gondwana Tree-Ring Laboratory, Little River, Canterbury, New Zealand
[g] Department of Forest and Wood Science, Stellenbosch University, 7602, Matieland, South Africa

ARTICLE INFO

Keywords:
Divergence
Temporally unstable relationships
Dendrochronology
Dendroclimatology
Reconstruction

ABSTRACT

Temporal stability of the relationship between a potential proxy climate record and the climate record itself is the foundation of palaeoproxy reconstructions of past climate variability. Dendroclimatologists have spent considerable effort exploring the issue of temporal instability of temperature records at high-latitude and −altitude Northern Hemisphere sites. Much of this work has focused on the Divergence Problem in which the modern ends of tree-ring chronologies exhibit pronounced departures from the climate-proxy relationships of preceding decades. However, there has been little scrutiny of how different methods might influence determinations of temporal instability at either the local scale or across broader spatial domains. Here we use four sets of Southern Hemisphere (SH) chronologies and three sets of synthetic data with known interventions to compare four methodologies that have been widely used to assess the temporal stability of relationships between tree-ring series and climate. Our analyses demonstrate that a determination of temporal instability may be partially dependent on method used to examine data, that some methods are more sensitive to standardisation choice than others, and that all methods are better at detecting high- rather than low-frequency instability. In all cases, the relatively modest strength of the relationships between the selected SH ring-width chronologies and temperature is likely to be an issue, especially if changes in trends are of interest. We recommend that robust assessment of temporal instability between tree-ring chronologies and observational climate data should use a range of methods and that unstable temporal relationships across space be carefully considered in the context of large climate field reconstructions.

1. Introduction

Uniformitarianism is one of the bedrock principles of palaeoclimatology. It states that relationships between the climate and climate-recording proxies are stable over time. Temporal instability in proxy-climate relationships would undermine long-term extrapolation of climate from proxy records to pre-observation periods. This applies not only to the proxy-climate relationships at the local level, but also to broader regional reconstructions. It is, therefore, a concern that in recent years a number of tree-ring studies have reported temporal instability in the relationship between tree growth and temperature. Some of this reported instability has been short-lived and occurred early in the 20th Century (e.g., Schneider et al., 2014). However, the overwhelming majority of these studies indicate an apparent divergence between climate records and tree-ring chronologies in recent decades at high-altitude or high-latitude locations. This 'modern-end' phenomenon has largely become encapsulated in the "Divergence Problem" (DP) debate (e.g., Jacoby and D'Arrigo, 1995; Briffa et al., 2004; Wilson and Luckman, 2002; Carrer and Urbinati, 2006; Frank et al., 2007; Wilson et al., 2007; D'Arrigo et al., 2008; Esper and Frank, 2009; Esper et al., 2010; Büntgen et al., 2012). In particular, these studies have commonly shown that observed temperatures appear to be increasing faster than changes in measured tree-ring parameters (i.e., ring width or density). Two types of divergence have been described in

* Corresponding author.
E-mail address: Kathryn.Allen@unimelb.edu.au (K.J. Allen).

https://doi.org/10.1016/j.dendro.2018.02.002
Received 7 April 2017; Received in revised form 10 February 2018; Accepted 15 February 2018
Available online 01 March 2018
1125-7865/

the literature. A sufficiently large and significant divergence in trend (i.e., low frequency divergence) in the calibration period can result in a reconstruction model that either under- or over-estimates the relationship between the climate target and the climate proxy outside the calibration period. Less discussed in the literature is divergence at the interannual scale (i.e., high-frequency divergence), which may prevent verification of a calibrated model (D'Arrigo et al., 2008). Many studies of the DP do not explicitly identify whether any observed occurrence of divergence is low or high frequency in nature, although low-frequency is often implied (Appendix A).

Decoupling between a chronology and climate was first discussed in relation to forest decline in North America in the 1970s and 80 s (Visser, 1986; Cook et al., 1987, Downing and McLaughlin, 1987; Cook and Johnson, 1989; Van Deusen, 1990). However, subsequent concerns about its impact on temperature reconstructions initiated extensive interrogation of key millennial-length tree-ring chronologies in the Northern Hemisphere (NH) with strong temperature signals (often based on maximum density rather than ring-width chronologies) for its occurrence (e.g., Briffa et al., 2004; Wilson et al., 2007; Esper et al., 2010; Anchukaitis et al., 2013). Discussions of the potential drivers of temporal instability have focussed on the modern end phenomenon. Proposed explanations have included inappropriate standardisation, global dimming, changes in growth-limiting factors, differential responses to maximum and minimum temperature when mean temperature is the target, local pollution, and changes in ozone concentration (D'Arrigo et al., 2008 and references therein). Changes in limiting factors (e.g., from temperature to precipitation or drought stress) or threshold effects have been widely suggested as the primary reason for recent divergence (Appendix A). Another possibility, and one that is not necessarily limited to the modern end of series, lies with problems in the climate data such as sparse coverage, fewer stations and poorer quality data back in time, or station inhomogeneities (Büntgen et al., 2006a, ; Allen et al., 2014).

Numerous studies have also pointed to the role of inappropriate standardisation of tree-ring chronologies as an important contributing factor to the DP (Esper and Frank, 2009; Esper et al., 2010; Büntgen et al., 2008; Linderholm et al., 2010; Andreu-Hayles et al., 2011a; Anchukaitis et al., 2013; Briffa et al., 2013). While a variety of standardisation approaches have been used in studies that identified divergence as an issue, few have explored the role of the standardisation choice on the presence divergence (Appendix A). Of these studies, three have specifically undertaken rigorous examination of how different standardisation methods may bias chronologies. Anchukaitis et al. (2013) implemented different standardisation choices on a set of pseudo-proxy data and both Briffa and Melvin (2011) and Briffa et al. (2013) examined the impact of different applications of regional curve standardisation on resultant chronologies. All three studies found important differences in resultant chronologies due to standardisation. Despite the attention given to standardisation choice, both Frank et al. (2007) and Esper et al. (2010) remarked that this alone could not fully account for the presence of the DP. Other methodological choices such as different standardisation of sample subgroups (Esper and Frank 2009, Briffa et al., 2013), the influence of using pith offset (or not), or power transformation of residuals together with standardisation methodology have also been considered (e.g., Büntgen et al., 2012). Signal-free standardisation methodology has further reduced end-effects that can occur with more traditional standardisation approaches (see Melvin et al., 2007; Melvin and Briffa, 2008; Briffa et al., 2013; Melvin et al., 2013), although it may also introduce some biases into chronologies (Anchukaitis et al., 2013).

Detection of divergence has entailed a variety of approaches, although many studies have used a single approach (some important recent exceptions include Àlvarez et al., 2015; Galvàn et al., 2015; Lavergne et al., 2015; see also Appendix A). Commonly used methods include: 1) visual comparison of tree-ring chronologies and climate target on the same plot (e.g., Jacoby and D'Arrigo, 1995; D'Arrigo et al., 2009; Appendix A); 2) comparison of response/correlation functions for two or more separate periods (e.g., Leal et al., 2008; D'Arrigo et al., 2009; Andreu-Hayles et al., 2011a); 3) moving correlations (e.g., Büntgen et al., 2006a,b; Naulier et al., 2015; Lavergne et al., 2015); 4) process-based modelling of tree-ring growth (e.g., the Vaganov-Shashkin-Lite model; Tolwinski-Ward et al., 2011; Lavergne et al., 2015; Sànchez-Salguero Camarero et al., 2017; Tumajer et al., 2017); and 5) application of the Kalman Filter (KF) to estimate regression models with time-varying coefficients (Visser and Molenaar, 1988; Jacoby and D'Arrigo, 1995; Wilson et al., 2013; Cook et al., 2013). Different approaches, however, may impact conclusions regarding the presence or absence of divergence (or temporal instability more generally) and each will have its own strengths and weaknesses. They may also be differentially sensitive to standardisation method used and some may be more suitable than others for the detection of low or high frequency instability—neither of these issues has previously been examined.

To date, there has also been a greater emphasis on checking chronology-climate relationships for stability using single climate records nearest the field site or composites of local records, rather than across a gridded data set (e.g., Büntgen et al., 2008; Tardif et al., 2003; Wilson et al., 2007; Grudd, 2008; Zhang et al., 2009; Allen et al., 2014). However, climate field reconstructions (CFR) implicitly assume that relationships across an often pre-defined spatial field are sufficiently stable over time, as do reconstructions relying on broadscale composites of temperature. With a growing interest in spatial field reconstructions of Southern Hemisphere (SH) temperature and precipitation using gridded climate data (e.g., Neukom et al., 2011; Neukom et al., 2013), and a rapidly increasing pool of available predictors, it will be increasingly important to scrutinise temporal stability of relationships across various spatial scales and domains. Although the underlying causes for temporal instability at broad scales may differ from those operating on local site reconstructions, instability in relationships across broad scales may have ramifications for our understanding of regional and global temperature dynamics.

In this study, we compare several simple and commonly used pre-reconstruction techniques to examine data series for temporal instability at both the local and broader regional levels in four long SH tree-ring chronologies. Our primary question is whether any of these techniques are more effective than others in detecting temporal instability. A secondary question is whether some methodologies for detecting divergence appear more sensitive to standardisation choice than others. We also extend two of the methods to examine temporal stability over broader spatial domains. Critically, three of the four SH chronologies used here, like many of the SH ring-width chronologies thus far developed, contain only a moderately strong relationship with temperature (typically $\sim 0.35 < |r| < 0.5$; e.g., Villalba, 1990; Allen et al., 2011; Anon (2014); Lavergne et al., 2015; Lavergne et al., in press) which is likely to make detection of an actual change in the relationship with climate more challenging (Esper and Frank, 2009). These weaker relationships may in turn have implications for the effectiveness of different methods used to detect instability at both local and regional scales. Due to the scarcity of highly resolved multi-century proxies in the SH (e.g., Fig. 1a PAGES2k Consortium, 2017), exclusion of SH ring-width chronologies with only moderate correlations with climate from multi-proxy reconstructions would effectively prohibit inclusion of a large number of SH tree-ring chronologies.

This study is not intended as an exhaustive review of the SH chronologies for temporally unstable relationships with temperature. Although we do briefly consider impacts of different standardisation methodologies on the results, our focus is on the methods used for detecting temporal instability, particularly in situations where the relationship between chronologies and climate is of moderate strength only. We also examine the relationships between synthetic temperature and tree-ring series that have had known interventions. In what follows, we use the term temporal instability to refer to a separation in trend or

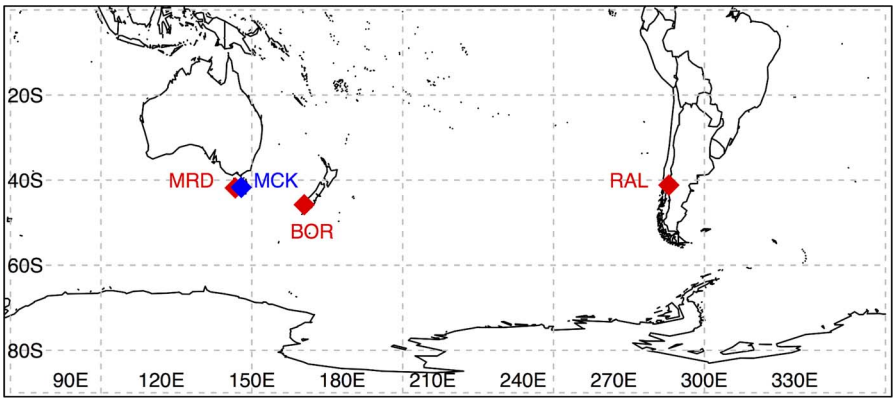

Fig. 1. Location of each of the four chronology suites used in this study. Red diamonds indicate the location of ring-width chronologies, blue indicates location of the cell wall thickness chronology. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to dissonance in the interannual relationship between two series that is sufficiently strong – or exists for a sufficiently long period of time – to be identifiable by a particular method. We do not, therefore, apply a lower threshold on the amount of time any stability must exist. Also, to be consistent with the DP literature, we use the term 'divergence' to describe modern end decoupling.

2. Data and methods

We use several common methods to check for temporal instability in two sets of data. The first of these is the selected tree-ring chronologies and relevant climate data described in Sections 2.1–2.3, and the second includes synthetic ring width and climate series, as described in Section 2.4. In Section 2.5 we outline the methods used to check for temporal instability at the local and broader spatial scales.

2.1. The chronologies

Tree-ring width chronologies were developed for one site from each of Argentina, Australia and New Zealand and cell-wall thickness chronologies developed for a single Australian site (Fig. 1; Table 1). The chronology sets include: Borland (BOR; *Nothofagus menziesii*; New Zealand) extending from 1389 to 2007 CE; Rio Alerce (RAL; *Fitzroya cupressoides*; Argentina) extending from 836–2011 CE and Mt Read (MRD; *Lagarostrobos franklinii*; Australia) that extends from 2145 BCE to 2009 CE (Table 1). The set of cell wall thickness chronologies from Mickey Creek (MCK; *Athrotaxis cupressoides*, Australia) extends from 1354 to 2009 CE. RAL and MRD were selected because both chronologies exceed 1000 years in length and have previously been used in temperature reconstructions (Villalba, 1990; Lara and Villalba, 1993; Salinger et al., 1994; Cook et al., 2000; Lavergne et al., in press). Of the New Zealand sites that extend beyond 1980, BOR has one of the strongest relationships with mean temperature, exceeds 500 years in length, and is located close to the southern limit of the species distribution. The MCK mean cell wall thickness chronology has one of the strongest correlations with mean temperature in Australasia and also exceeds 500 years in length. Although these chronologies can be considered only moderately sensitive to temperature (RAL $r \sim -0.4$ prior December–March, Villalba 1990, Lavergne et al., in press; MRD ~ 0.6 November–April Cook et al., 1991; BOR January–April; $r \sim 0.4$; MCK $r \sim -0.6$ January), they are amongst the longest and most temperature-sensitive ring-width chronologies in the SH mid-latitudes. Most of the long NH maximum density chronologies investigated for divergence are correlated with temperature at $r \geq 0.6$, and often $r \geq 0.7$ (e.g., north of 50°N, Briffa et al., 2004; Sweden, Grudd, 2008; Polar Urals, Briffa et al., 2013; northern Canada, D'Arrigo et al., 2009). The mid-latitude location of our four sites, and the lack of high elevation mountains in Tasmania are partly responsible for modest temperature signals in many ring-width chronologies from this region. The reason for negative correlations at MCK is unclear, but strongly negative relationships have also been found for several other Tasmanian wood-properties chronologies (Allen et al., 2013; Drew et al., 2013). The moderate and negative relationship between RAL and temperature is observed for the prior growing season.

2.2. The standardisation methods

We used four different standardisation methods for all chronologies. A 67% cubic smoothing spline (67%nSpl) was used to help to retain lower frequency variation in the data (dependent on length of sample) while still removing much non-climatically related variance (Cook, 1985). The median series length spline (MedSpl) was used as a compromise between the 67%nSpl and less flexible methods. Median series lengths at the different sites were: 444 years (RAL), 225 years (BOR), 470 years (MRD) and 213 years (MCK). The extent of resolvable low frequency variance from BOR and MCK in particular will therefore be limited. We also used the deterministic negative exponential curve or regression line (any slope; Nex/reg) that is commonly used in an effort to preserve as much low frequency variance as possible. Finally, we employed an age dependent smoothing spline (ADspl; Melvin et al., 2007) which retains a large proportion of low frequency variance. The flexibility of the curve is greatest when trees are young and decreases with increasing tree age (Melvin et al., 2007). Although Regional Curve Standardisation (RCS) is often used to preserve low frequency variability (Briffa and Melvin, 2011), we have not used it here due to relatively low sample depth in both BOR and MCK. Our use of a number of more flexible detrending methods is also directly linked to the mesic forest environment of our sites in which non-synchronous endogenous disturbances are known to occur.

Both signal-free standardisation and the calculation of chronologies based on residuals (i.e. absolute differences from the standardisation curve) have been shown to reduce end biases in many chronologies (Cook and Peters 1997; Melvin and Briffa 2008). We compare chronologies standardised and computed using signal-free methodology (Melvin and Briffa, 2008) and those that have simply been standardised. For the purposes of this study we focus on residual chronologies, although the results for chronologies based on ratios are similar and are provided in the Supplementary material.

2.3. Climate data

The type of data used to represent local temperatures differed across the three continents. The longest available temperature series for the New Zealand site is the Climate Research Unit (CRU) data starting in 1901 (Mitchell and Jones 2005). There are nine climate stations within the 1200 km decay distance used to compile the CRU data (Mitchell and

Table 1
Chronology and climate data characteristics. N is number of trees and n the number of samples.

Site	MRD	MCK	BOR	RAL
Country	Australia	Australia	New Zealand	Argentina
Chronologies				
Latitude	41.84° S	41.75° S	45.78° S	41.17° S
Longitude	145.54° E	146.7° E	167.37° E	71.77° W
Elevation (m ASL)	900	1200	910	980–1060
Aspect	S	S	N	S
N (n)	150 (283)	16 (30)	31 (65)	99 (177)
Total chronology period (CE)	~ 2145 to 2009	1354–2009	1389–2007	836–2011
Climate data				
Region from which temperature stations contribute to grid/location local temperature station	~ 40°–44° S, 145°–148.5°E	~ 40°–44° S, 145°–148.5°E	~ 39.5°–46.5° S, 168°–177°E	NA
Grid point used for local temperature	41.85°S, 145.55°E	41.8°S, 146.7°E	45.75°S, 167.25°E	NA
Temperature station elevation	All stations > 40 years at elevations much lower than tree-ring sites	All stations > 40 years at elevations much lower than tree-ring sites	All stations < 100 m ASL	790
Number temperature stations covering whole period used	0	0	5	1
Comments	22 Tasmanian stations > 12 years of data western-central Tasmania. These stations would have been more heavily weighted than stations elsewhere in Tasmania.	Stations form the central north would have received higher weighting the calculation of the gridbox temperatures for MCK than they would have for MRD due to their relative proximity to the site.	9 stations within the 1200 km decay distance (Mitchel and Jones 2005)	Single meteorological station

Jones 2005) that would have contributed to the temperature data for the grid square over the BOR site. Five of these cover the entire period used and are all at low elevation (Table 1). Bariloche, at 840 m ASL, is the closest meteorological station to the RAL site on the eastern flank of the Argentinian Andes and the record extends from 1915 to 2011 CE (41.09°S–71.10°W; Servicio Meteorológico Nacional). As there is no long-term meteorological station at a similar elevation to either the Mt Read (MRD) or Mickey Creek (MCK) sites in the mountainous western and central parts of Tasmania, we have used the appropriate grid squares (see Table 1) from the 0.05° × 0.05° gridded Australian Water Availability Project data (AWAP; Jones et al., 2009). In this dataset, all stations containing 12 or more consecutive years of data for individual months have been incorporated in single grid square time series (although this constraint was relaxed to four for some topographically variable areas; Jones et al., 2009). The data for any grid box is based on a weighting function in which the weight depends on the distance between a station and the grid box of interest (Jones and Trewin, 2000). Of the six stations that were operating by 1926 in the more heavily weighted central and western regions, three have records that extend beyond 2000, but these are discontinuous, with some large gaps. Of the 157 Tasmanian temperature records used to generate the Tasmanian AWAP data, no single site recorded temperatures over the entire period (http://www.bom.gov.au/climate/data/lists_by_element/alphaTAS_36.txt; Table 1). For the Australian and New Zealand gridded data, we arbitrarily limited our analyses to begin in the year in which the number of stations from the broader region around the relevant grid point was at least 50% of the number of stations operating in 2000. This was done because correlation between a chronology and the gridded climate data for a region is more likely to reflect the true relationship as the number of climate stations used in the gridded data increases (Anchukaitis et al., 2013). This resulted in the use of the periods 1901–2011(New Zealand) and 1926–2009 (Australia).

To check spatio-temporal stability across broader areas we used the gridded CRU temperature data in all cases because more highly resolved data are not yet available for all of South America and New Zealand over the 20th Century (Bianchi et al., 2016; pers. comm. D. Lorrey, National Institute of Water and Atmospheric Research (NIWA), New Zealand). We recognise that integrity of the climate data for highly topographically heterogeneous areas will be compromised by the use of the relatively coarse CRU data, hence limiting inference. The same criterion described in the above paragraph was used to select the periods examined, resulting in a change only for RAL: now 1954–2010. The selection of the region covered by the broader domains was determined by a number of factors. For southern South America and Australia, temperature reconstructions have previously been developed for these areas or are related to much of them (South America; Neukom et al., 2011, 2010 and southeastern Australia; Allen et al., in revision). Additionally, both regions are subject to multiple interacting ocean-atmosphere processes that have non-stationary relationships (Risbey et al., 2009; Gallant et al., 2013) such as the Southern Annular Mode (SAM), El Niño Southern Oscillation (ENSO) and fluctuations in the strength of the Hadley circulation (Evans and Kaplan, 2004; Ambrizzi et al., 2004). They also include a wide variety of bioclimatic environments. The New Zealand domain simply encompasses the two main islands of the country.

2.4. Synthetic data

As a basis for assessing the different methods used, we generated a synthetic temperature series loosely based on the Berkeley temperature anomaly data for Tasmania (December-February), embedding an increasing trend over the last 30 years. To generate the tree-ring data we used an amalgam of three *L. franklinii* sites of measured cell diameter data (TRD), a parameter typically strongly associated with temperature ($r < -0.5$). We then inverted each series (TRD in *L. franklinii* is negatively associated with temperature) and scaled it to the temperature

data. We produced two sets of tree-ring series from this. To the first we added a normally distributed random error term with relatively low variability (final chronologies correlated with temperature at r ~ 0.8), but for the second, we allowed the error term greater variability (final chronologies correlated with temperature at r ~0.42; cf. our actual tree-ring series) to result in series strongly (strong) and moderately (moderate) coupled with temperature, respectively. Each set of synthetic tree-ring series were then perturbed to produce a set of series in which the interannual relationship at time 80 (‘1980¢) decoupled and a set of series in which the trend from time 80 decreased relative to the original series. This produced six different sets of synthetic data: the two original series, the two series (strong and moderate) including the interannual intervention and the two series (strong and moderate) containing the trend intervention. Each set of series was then subjected to the same standardisation choices as the actual data.

2.5. Checking for temporal instability

We used a variety of approaches (described below) to check for temporal instability at the local scale, while at the broader spatial scale we applied two-period comparisons of correlations and the Kalman Filter.

2.6. Visual comparison

Chronologies and temperature were compared on the same plot (e.g., Jacoby and D’Arrigo, 1995; D’Arrigo et al., 2009). In these cases, temporal instability was subjectively identified as a substantial departure of trends or interannual variations in the tree-ring series and temperature.

2.6.1. Running Pearson correlation coefficients

Running correlations were generated for successive 30-year windows (e.g., Büntgen et al., 2006a,b; Naulier et al., 2015) and significant (p £ 0.05) low frequency modulation was tested for using the g_test function in R package Treeclim (Zang and Biondi, 2013, 2015). This test is a multivariate extension of the test originally proposed by Gershunov et al. (2001) that has been applied elsewhere (e.g., Franceshini et al., 2012). It produces 1000 sets of simulated climate data that are Gaussian noise and a simulated proxy based on the original correlation function that includes an error component. The variance of this error component is the same as the variance not explained by climate in the real data. For each window, the correlations between the actual chronology and climate predictor are compared to the distribution of bootstrapped standard deviations of the simulated data to identify if higher or lower than expected low-frequency modulation occurs.

2.6.2. A two-period comparison of correlations

We compared the relationship between each chronology and mean temperature for ‘early’ and ‘late’ periods at both the local (e.g., Andreu-Hayles et al., 2011a) and regional (e.g., D’Arrigo et al., 2009) scales. Each ‘early’ and ‘late’ period contained half of the available overlap between temperature and chronology series. Although this two-period scheme is only one (and a relatively naïve one perhaps) of several that could be used (e.g., > 2 periods, periods of different length), we have used a 50/50 split because this is a common procedure for validating the quality of climatic reconstructions. Clearly, however, the use of alternative periods or a greater number of periods could influence results. The following early and late periods were used for the ‘local’ results: BOR: 1901–1953 and 1954–2007; RAL 1915–1962 and 1963–2010; MRD and MCK 1926–1967 and 1968–2009. With the exception of RAL (1954–1981 and 1982–2010), the same periods were used for the spatial analyses. The significance of the difference between the early and late period r-values was assessed after stabilising the variance and transforming the r scores to z-values using Fisher’s transformation (Fisher, 1915).

2.6.3. The Kalman filter (KF)

The Kalman Filter (KF) estimates the dynamic relationship between two variables over time (Visser and Molenaar, 1988; Jacoby and D’Arrigo, 1995; Cook et al., 2013) and compares it with a linear model. Here, the linear (constant) model can simply be described as an observation equation:

$$y_t = z_t\alpha + \xi_t \quad (1)$$

where y_t is the tree-ring series, z_t is the temperature series, and ξ_t are the randomly distributed error terms. The dynamic model in which the α_t are allowed to vary over time as a random walk can be described as:

$$y_t = z_t\alpha_t + \xi_t \quad (2)$$

$$\alpha_t = \alpha_{t-1} + \eta_t \quad (3)$$

Eq. (3) is known as the transition equation. Both ξ_t and η_t are assumed to be normally distributed and independent ($\xi_t \sim$ iid N(0,σ^2) and $\eta_t \sim$ iid N(0,σ^2Q); see Visser and Molenaar, 1988 or Harvey, 1989 for a fuller treatment). The best model—either (1) or (2)—is selected using the bias-corrected Akaike Information Criteria (Van Deusen, 1990; AIC_t for dynamic model and AIC_c for the constant coefficient model). The model with the lower AIC is considered a better fit to the data if $AIC_c - AIC_t < -2$ (the penalty on the AIC to account for the inclusion of additional parameters). In addition to using this AIC comparison, we also consider the 95% intervals around the estimate (point estimate ± 2s), based on the Q estimates (variance associated with the error term in the transition equation). If there is no variation in the relationship over time, then the Q values will be 0. If the entire interval lies above (below) zero, there is a significant positive (negative) relationship between the chronology and temperature series. In applying the KF here, only those dynamical relationships in which the $AIC_c - AIC_{t<} -2$ and a change of sign in the relationship occurred were noted as ‘significant’, and thus considered to indicate temporal instability in the relationship (not necessarily just at the modern ends of series). A change in the sign of the coefficient interval from positive (negative) to negative (positive) or from positive/negative (indistinct from zero) to indistinct from zero (positive/negative) is likely to be more problematic for climate reconstruction than a relationship in which the sign does not change. Our criteria for identifying temporal instability are therefore conservative.

3. Results

Below, we firstly report the results for the synthetic data (Figs. 2–4) and then the results for the different sites (Figs. 5–11). For the site results, a single figure is used to present results based on a particular technique across all sites. The information presented in each figure is as follows: Fig. 5, the smoothed chronologies standardised in different ways; Fig. 6, a visual comparison of the chronology sets against mean temperature; Fig. 7, running correlations; Fig. 8, comparison of two-period correlations; Fig. 9, Kalman Filter site results. Fig. 10 shows the comparison of two-period spatial correlations across a broader spatial scale and results of the spatial application of the Kalman Filter are shown in Fig. 11. In the Kalman Filter results, a solid horizontal line indicates a constant relationship. If the 95% interval, shown by the dotted lines, is entirely above (below) the zero line, then the relationship between the chronology and temperature is significantly positive (negative). A non-horizontal line indicates a non-constant relationship. Where the 95% interval changes from positive (negative) to negative (positive), or from positive/negative (no different from zero) or vice versa, this indicates a change of sign in the relationship. The same criteria (see Methods) were used to define instability for the local- and regional-scale, but for the regional-scale results, shading on Fig. 11 shows the decade in which the instability began, or indicates instances of multiple sign changes. Only results for the residual chronologies (local data) are shown in the main body of this study. Results for

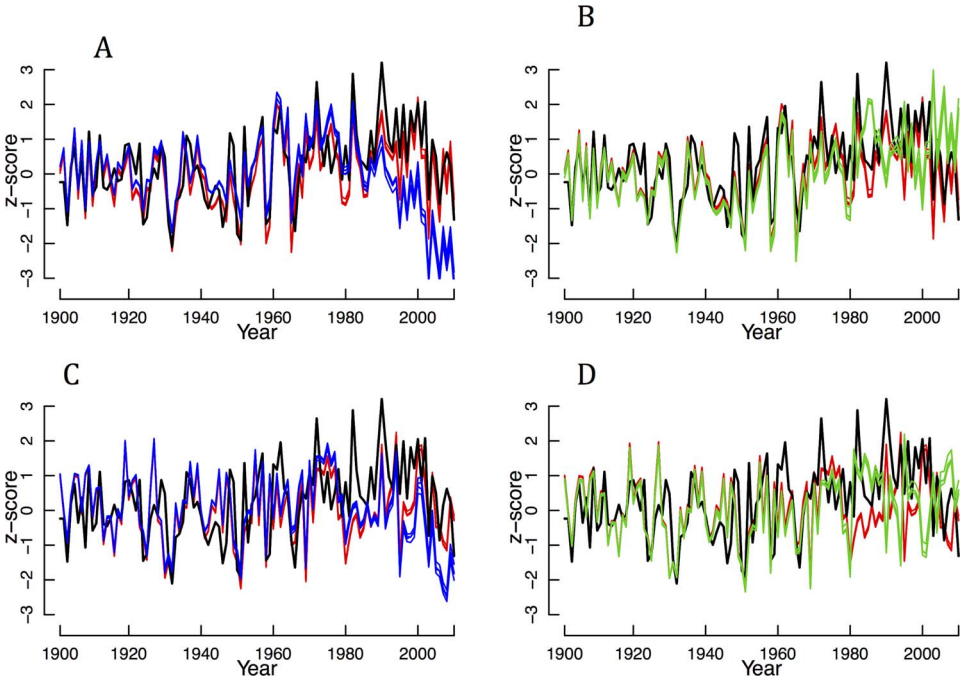

Fig. 2. Synthetic temperature (black) and tree-ring data. A. Synthetic data that is strongly associated with temperature (r ~ 0.8); B. Synthetic data that is only moderately associated with associated with temperature (r ~ 0.42). Note that in each panel, each differently standardised chronology will be identical before the interventions at time 80. Multiple different lines of one colour reflect the different standardisations applied. Red series are unperturbed, blue series have a decreasing trend relative to the temperature series embedded, while green series have been inverted from time 80. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

chronologies based on ratios are shown and briefly discussed in the Supplementary Material. Similarly, we present spatial results only for one standardisation option (Spl67%n) in the case of two-period comparisons, presenting the rest in the Supplementary Material (residual chronology results only).

3.1. Synthetic data

Visually, the change in trend is more apparent for the strongly (Fig. 2a) associated series (strong) than the moderately associated series (moderate; Fig. 2c). The changed relationship due to the interannual perturbation from time 80 (‘1980¢) is apparent for both the strong and moderate series (Fig. 2b, d). Only small differences due to standardisation occur, but are largest at the series’ modern ends.

Running correlations (Fig. 3a–b) were effective in detecting the interannual intervention from time 80 (‘1980¢) for both the strong and moderate series, and the *g*_test indicates that low frequency modulation is significant (Table 2) in all cases. The visually apparent change in trend (Fig. 2) was significant for both the moderate and strong series (Table 2). More unexpected was significant low frequency modulation in the unperturbed strong series (Table 2). Pertinently, differences due to standardisation method are relatively small in the running correlation results, but are clearer for the trend than the interannual intervention (Fig. 3a–b).

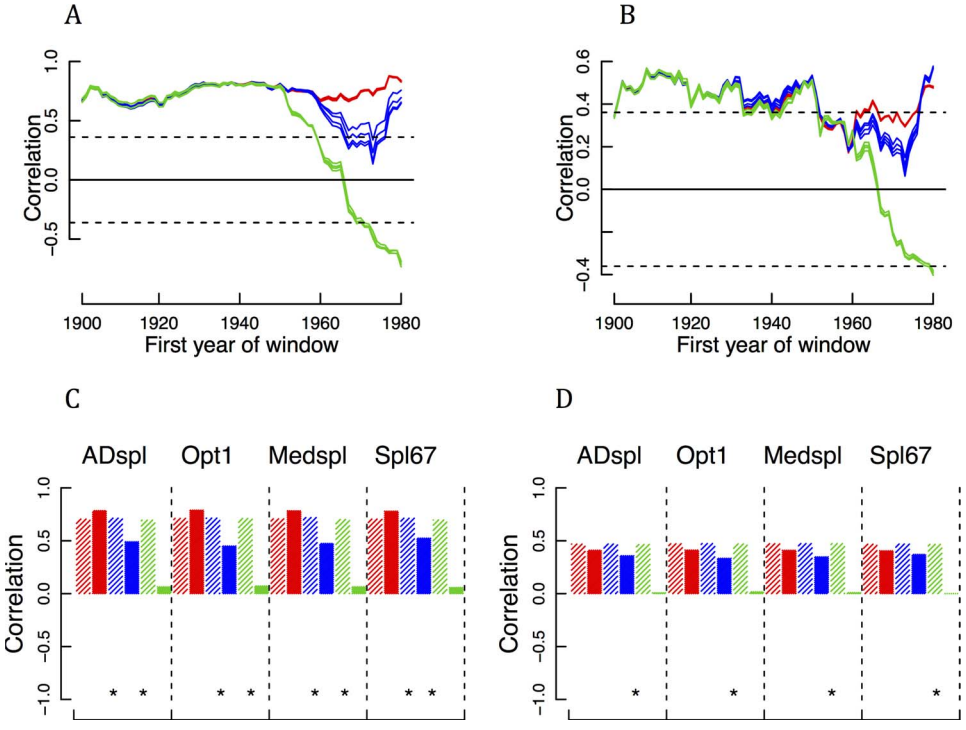

Fig. 3. Top: Running correlations between synthetic tree-ring data (residual chronologies only) and temperature. A Synthetic data that are strongly associated with temperature. B Synthetic data that are moderately associated with temperature. Bottom: Two period correlations for synthetic data. C. Strongly associated chronologies, D. moderately associated chronologies. Red series are unperturbed, blue series represent the chronologies with a trend intervention from time 80 and green series have an interannual perturbation added from time 80. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Results of g_test for low frequency modulation of running correlations. An H (L) indicates higher (lower) low frequency modulation than expected ($p < 0.05$). Black is for chronologies calculated using residuals and red for chronologies calculated using ratios data. A black (red) * indicates that residual (ratio) series were not normally distributed and therefore results of g_test may not be reliable.

Chronology:	Low frequency modulation	Chronology	Low frequency modulation	Chronology	Low frequency modulation	Chronology	Low frequency modulation
MRD_Nex/reg_residual_std		RAL_Nex/reg_residual_std	* H*	BOR_Nex/reg_residual_std	H	MCK_Nex/reg_residual_std	
MRD_Nex/reg_residual_ssf		RAL_Nex/reg_residual_ssf	* *	BOR_Nex/reg_residual_ssf	H	MCK_Nex/reg_residual_ssf	
MRD_ADSpl_residual_std		RAL_ADSpl_residual_std	* H	BOR_ADSpl_residual_std	H	MCK_ADSpl_residual_std	
MRD_ADSpl_residual_ssf	*	RAL_ADSpl_residual_ssf	* H	BOR_ADSpl_residual_ssf	H	MCK_ADSpl_residual_ssf	
MRD_Medspl_residual_std		RAL_Medspl_residual_std	* H	MCK_Medspl_residual_std	H	MCK_Medspl_residual_std	
MRD_Medspl_residual_ssf		RAL_Medspl_residual_ssf	* *	MCK_Medspl_residual_ssf	H	MCK_Medspl_residual_ssf	
MRD_67nsplp_residual_std		RAL_67nsplp_residual_std	* H	MCK_67nsplp_residual_std	H	MCK_67nsplp_residual_std	
MRD_67nsplp_residual_ssf	*	RAL_67nsplp_residual_ssf	H* H*	MCK_67nsplp_residual_ssf	H	MCK_67nsplp_residual_ssf	
Strong_Nex/reg_residual_ssf	H	Strongtrend_Nex/reg_residual_ssf	H	Stronginterann_Nex/reg_residual_ssf	H		
Strong_ADSpl_residual_ssf	H	Strongtrend_ADSpl_residual_ssf	H*	Stronginterann_ADSpl_residual_ssf	H		
Strong_Medspl_residual_ssf	H	Strongtrend_Medspl_residual_ssf	H	Stronginterann_Medspl_residual_ssf	H		
Strong_67nsplp_residual_ssf	H	Strongtrend_67nsplp_residual_ssf	H	Stronginterann_67nsplp_residual_ssf	H		
Weak_Nex/reg_residual_ssf		Weak_trend_Nex/reg_residual_ssf	H	Weakintreann_Nex/reg_residual_ssf	H		
Weak_ADSpl_residual_ssf		Weaktrend_ADSpl_residual_ssf	H	Weakinterann_ADSpl_residual_ssf	H		
Weak_Medspl_residual_ssf		Weaktrend_Medspl_residual_ssf	H	Weakinterann_Medspl_residual_ssf	H		
Weak_67nsplp_residual_ssf		Weaktrend_67nsplp_residual_ssf	H	Weakinterann_67nsplp_residual_ssf	H		

The two period correlations for the synthetic data suggest their effectiveness in detecting changes in interannual relationships between both moderately and strongly associated series (Fig. 3c–d). Significant differences in correlation between the two periods was detected for the trend intervention in the strong series for all standardisations but was not detected in any case for the moderate series.

The Kalman Filter results (Fig. 4) do not suggest a change in the relationship with temperature for either the unperturbed strong or moderate series. The trend intervention in the strong series resulted in a change of sign of the relationship for one standardisation method (Nex/reg) only, whereas the interannual intervention has had a clear impact on both the moderate and strong series. In summary, the interannual intervention is apparent when using any of the correlation or KF methods, but the impact of the trend intervention is more ambiguous, particularly for the moderate series.

3.2. Site results

3.2.1. Comparison of signal-free and non-signal-free standardisation

Little difference between a standardised chronology and its companion signal-free chronology indicates standardisation has introduced minimal distortion whereas large differences suggest considerable distortion has been introduced (Melvin and Briffa, 2008). Largest visual differences between signal-free and non-signal-free chronologies occurred for ADspl (RAL) and Nex/reg (MRD; Fig. 5). Most pertinently, most of these differences occurred at the modern ends of BOR (ADspl, Nex/reg; 67%nSpl), RAL (all options) and MRD chronologies (ADSpl, Medspl, 67%nSpl; Fig. 5) and underscore previous observations concerning the importance of appropriate standardisation. The question of interest in this study is whether the methods used here are sensitive to differences in standardisation.

3.2.2. BOR

Different standardisation options have had some impact on the early and modern ends of the resultant BOR chronology (Figs. 5; S1 – especially signal-free). Visually, there does not appear to be any temporal instability (Figs. 6, S2), although running correlations (Figs. 7, S3) indicate a weakening in the relationship ~1940–1955 (cf. similar feature in Schneider et al., 2014), or perhaps weaker external forcing. Low frequency modulation in correlations is significant (Table 2). However, the difference between correlations for the early and late periods is not significant (Figs. 8; S4) and the Kalman filter results indicate stability at the local level (Figs. 9; S5). Spatially, results are very similar across standardisation options (Figs. S6–8), there being almost no change in the number of grid cells for which a statistically significant difference in correlations across the two periods exists (Fig. 10). The spatial application of the KF also indicates stable relationships across all standardisation options. There are, however, fewer grid cells with significant relationships for Nex/reg than for the other standardisation methods (Fig. 11). With the exception of the running correlations, results for BOR suggest a relatively stable relationship with temperature over time.

3.2.3. RAL

There are visible differences at the modern ends of the RAL chronologies (after 1800; std and ssf), especially for ADspl (Figs. 5, S1). Medspl and 67%n produced very similar chronologies for both the signal-free and non-signal-free approaches (Figs. 5, S1). With the exception of the ADspl chronology, all the RAL chronologies have had their modern ends pulled up by signal-free standardisation suggesting a larger end-effect bias for other standardisations (Figs. 5; S1). Visually, the inverted RAL chronology has progressively decoupled from temperature since ~1980 (Figs. 6, S2), this being greatest for ADSpl and the deterministic Nex/reg standardisations, both of which aim to retain medium-low frequency variation. This same pattern is also evident in the running correlations after ~1960 (Figs. 7, S3). Unfortunately, the

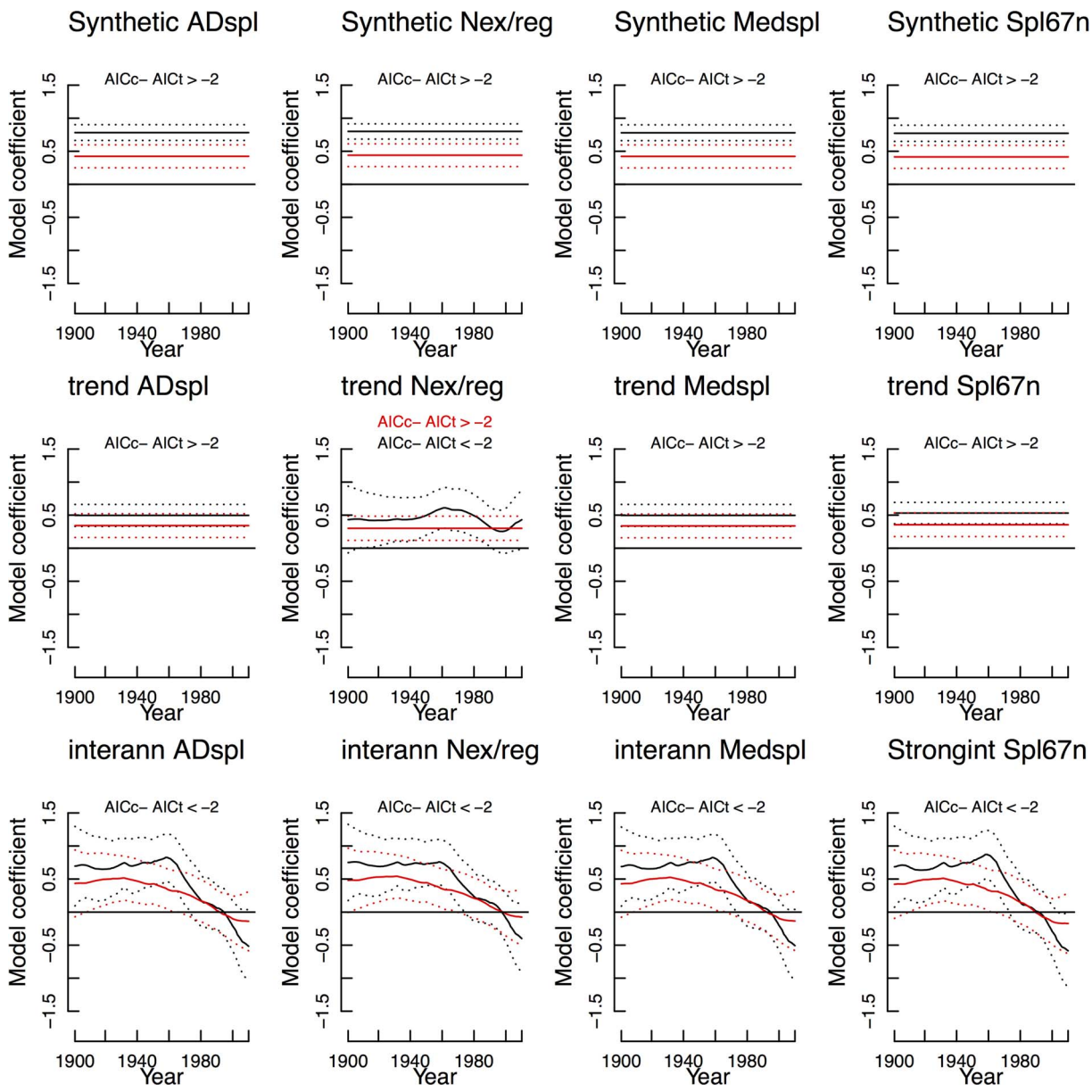

Fig. 4. Kalman Filter traces for the synthetic data. Black is for strongly associated data, and red for moderately associated data. Where $AIC_c - AIC_t$ is the same (> -2 or < -2), for both sets of results only one value is added to plot. Top row is for unperturbed synthetic data, second row shows synthetic data with trend intervention and third row shows results for synthetic data with the interannual intervention. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

g test relies on normally distributed data, which was not the case for several instances of the chronology, making the results of the test unreliable (Table 2). Based on Fig. 6, it appears that the modern-end decoupling is low frequency in nature and Figs. 7 & S3 also indicate weaker correlations towards the start of the temperature record. Correlation between temperature and the chronology for the early period is considerably stronger ($r \leq -0.5$) than for the later period ($-0.43 \leq r \leq -0.15$) in all cases (Figs. 8; S4). However, only the difference for (ADspl) is statistically significant ($p < 0.05$). The local KF results (Figs. 9, S5) show a significantly temporally unstable relationship for some standardisation options (e.g., Nex/reg, ADspl (ssf) and ADspl (std, ssf), Medspl(ssf), 67%nSpl (std, ssf)), but not others.

The two-period comparisons for the broader spatial region show dramatic changes (Figs. 10; S6–S8), but the very short periods used (28 years), as well as the timing of the split at 1982 exert important effects here. Note that the two periods used for the spatial comparison differ from those used for the local comparison (local meteorological station) and so the two sets of results are not directly comparable. Strongest correlations occur in the central part of the domain in the early period. The greatest extent of significant changes in correlation varies by standardisation method, but generally occurs for the eastern and southern parts of the domain. Similarly, the areal extent of multiple changes in sign – as diagnosed by the KF – differs depending on standardisation method used (Fig. 11), with greatest extent occurring for Medspl and Spl67%n. Overall, results point to temporal instability of the relationship between RAL and temperature that most likely began around 1980.

3.2.4. MRD

Standardisation and chronology construction methods also impacted the modern end of MRD (Figs. 5; S1). Signal-free standardisation has led to a greater difference between temperature and most residual chronologies for MRD at their modern ends compared to the non-signal-free chronologies (exception of Nex/reg; Figs. 6 & S2). The largest increase due to signal-free standardisation occurred for Spl67%n and Medspl (Fig. 5), suggesting that these standardisation methods resulted

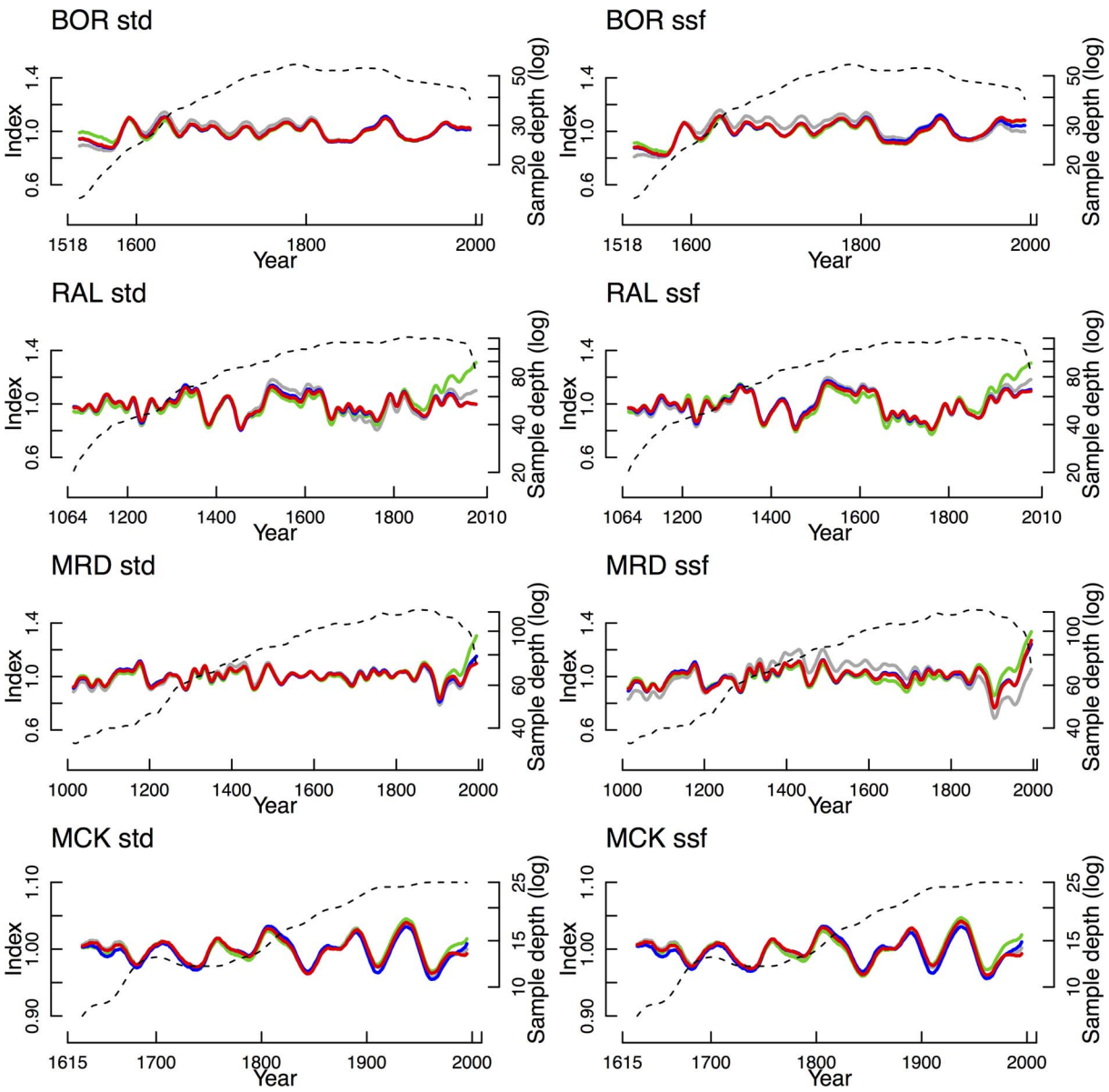

Fig. 5. The four smoothed Southern Hemisphere residual chronologies standardised in four different ways (see text). Only portions of chronologies for which EPS > 0.85 and n > 5 have been included. Note that although MRD is shown only since 1000CE, EPS > 0.85 back to 400BCE. Sample depth (log scale) of each chronology is shown on the right axis. Grey is Nex/reg, green is ADspl, blue is Medspl and red is 67%nSpl. Std indicates non-signal-free chronologies while ssf indicates series have been standardised within the signal-free framework. Time-scale shown on x-axes differs across sites. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in greater bias at the modern ends of series. The modern ends of the Medspl and Spl76%n chronologies were more similar to the ADspl after being processed via the signal-free framework. However, the signal-free Nex/reg standardisation option has resulted in the modern ends (since ~1900) of many series being pulled downwards while earlier sections have been pushed higher (see Fig. 5), thus influencing the shape of the final chronology. The signal-free standardisation has emphasised what appear to be common characteristics across many series in the chronology, but are instead a product of Nex/reg standardisation. This demonstrates that application of signal-free standardisation alone is not a "silver bullet" (Melvin and Briffa, 2008; Briffa et al., 2013) to resolving issues that may occur due to standardisation method.

Running correlations for MRD are generally statistically significant, excepting the periods prior to ~1940 (some standardisations) and ~1970 ($p < 0.05$; Figs. 7, S3), but in no case do the g_test results support the presence of significant low frequency modulation (Table 2). There is little difference between correlations in the early and late periods for any standardisation (Figs. Fig. 8; S4). The local KF results indicate a weakening mid-century relationship, but it remains positive despite AIC_c-$AIC_{t<}$ −2 in three cases (Figs. 9; S5). Although the spatial extent of strong correlations is greater in the later period (Figs. 10; S6–S8), significant difference between the two periods occurs only over the central north and far northeast of the domain. The spatial application of the KF suggests that issues may exist over much of the domain for which significant relationships with temperature exist (Fig. 11). The extent of the area for which there are multiple changes in the sign of the relationship varies depending on the standardisation approach used. For example, the Nex/reg chronology suggests growth at Mt Read has accelerated relative to temperature since the 1990s across much of the northwestern part of the domain whereas multiple changes in the sign of the relationship across much of the region is indicated for the ADspl chronology. Overall, the results for MRD suggest that, while local scale results are not problematic, relationships across the broader domain may be.

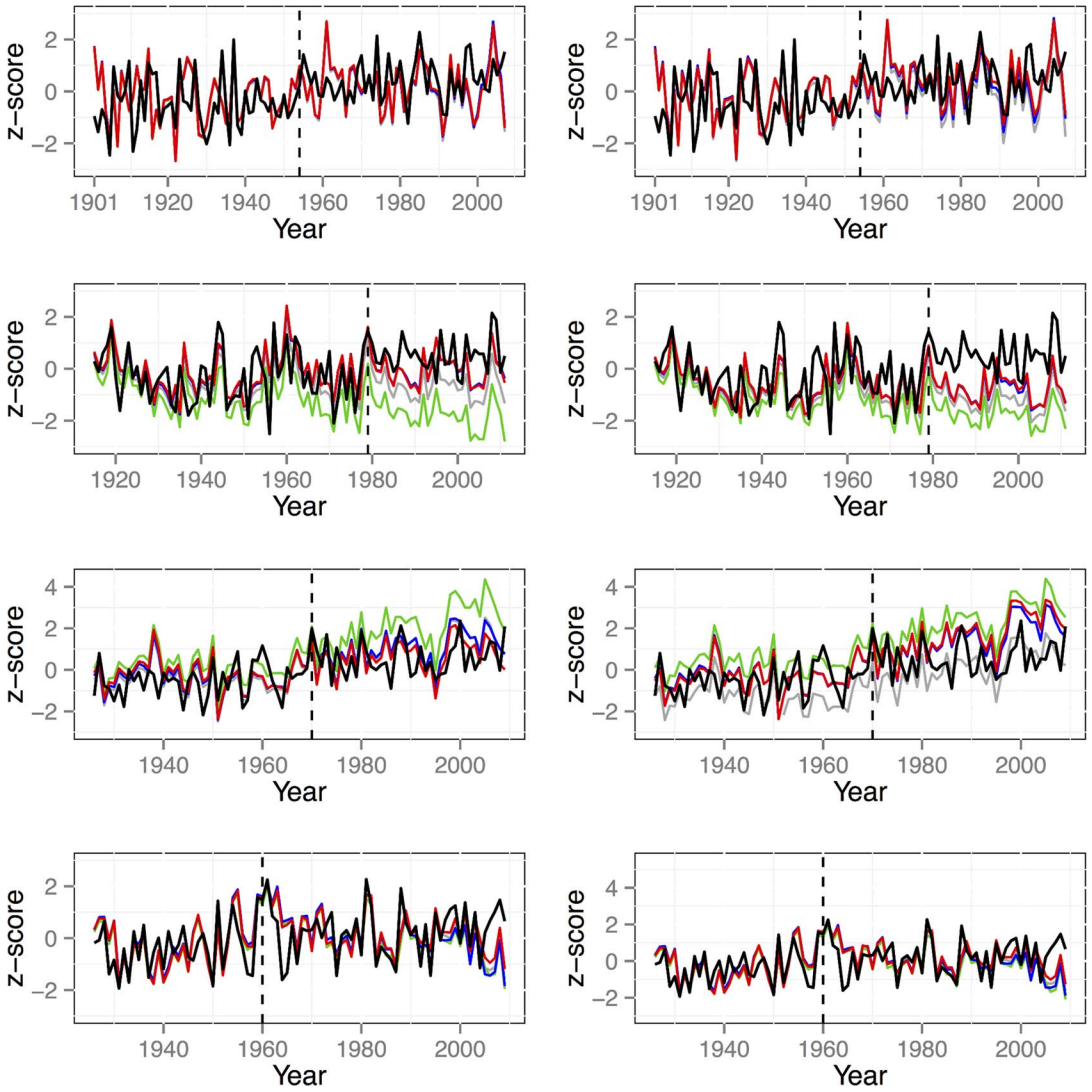

Fig. 6. Comparison of interannual temperatures and residual chronologies. For each plot, the thick black line is the temperature series and the remaining thin lines show the variously detrended chronologies. Grey is Nex/reg, green is ADspl, blue is Medspl and red is 67%nSpl. Note that the RAL and MCK chronologies have been inverted for clarity − the relationships are actually negative. Std indicates non-signal-free chronologies while ssf indicates series have been standardised within the signal-free framework. Time-scale shown on x-axes differs across sites. Dashed lines on plots show breakpoints identified by homogeneity tests on the climate data (breakpoints occur at BOR: 1954/5; RAL 1978/9; MRD 1969/70; MCK 1959/60). Only that for MCK was not considered significant (see Table S1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.5. MCK

Although standardisation option had relatively little impact on the MCK chronology, there are small differences, mainly at the modern end of the chronology (Fig. 5, but see S2). The relatively low variability in raw measurements, compared to the ring width chronologies, may be responsible for this. Visual comparison with temperature and two-period local correlations (ADspl and Medspl) both suggest local-level issues with MCK (Figs. 6; S2). Running correlations also show a weakening of the association with temperature at the modern end of the non-signal-free chronologies, although the correlation does remain significantly negative (Figs. 7; S3) and there is no significant low frequency modulation (Table 2). O'Donnell et al. (2016) also noted an issue with the last half decade of the density chronology from this same site and found that excluding these years from the model improved the reconstruction. The pattern of difference between early and late period correlations is the same across different standardisations but only differences for ADspl and Medspl are significant (Figs. 8, S4). The KF results show a weakening of the relationship over time, and the length of time for which non-significant coefficients occurred varies by standardisation method (cf. Nex/reg and 67%nSpl).

Spatial correlations (Figs. 10, S6–S8) indicate significant changes ($p < 0.05$) in the relationship over Tasmania and to the far west of the domain. The spatial KF results suggest multiple changes in the sign of the relationship since 1926 over much of western Tasmania and the southern Australian mainland in two cases (Fig. 11; ADspl, Medspl), but the 67%n and Nex/reg series indicate a significant change in relationship sign towards the modern end of the series for parts of the same area. Most of the results for MCK point to a change in the relationship with temperature that likely occurred at the modern end of the series.

4. Discussion

4.1. Is there a 'best' method to detect temporal instability?

Ideally, any method used to test temporal stability of the relationship between two series would identify known issues between two series. Additionally, given overwhelming evidence (e.g., Büntgen et al., 2008; Esper et al., 2010) that standardisation is a critical aspect of temporal instability at the modern ends of series (i.e., the DP), any method used should be sensitive to standardisation method. A summary

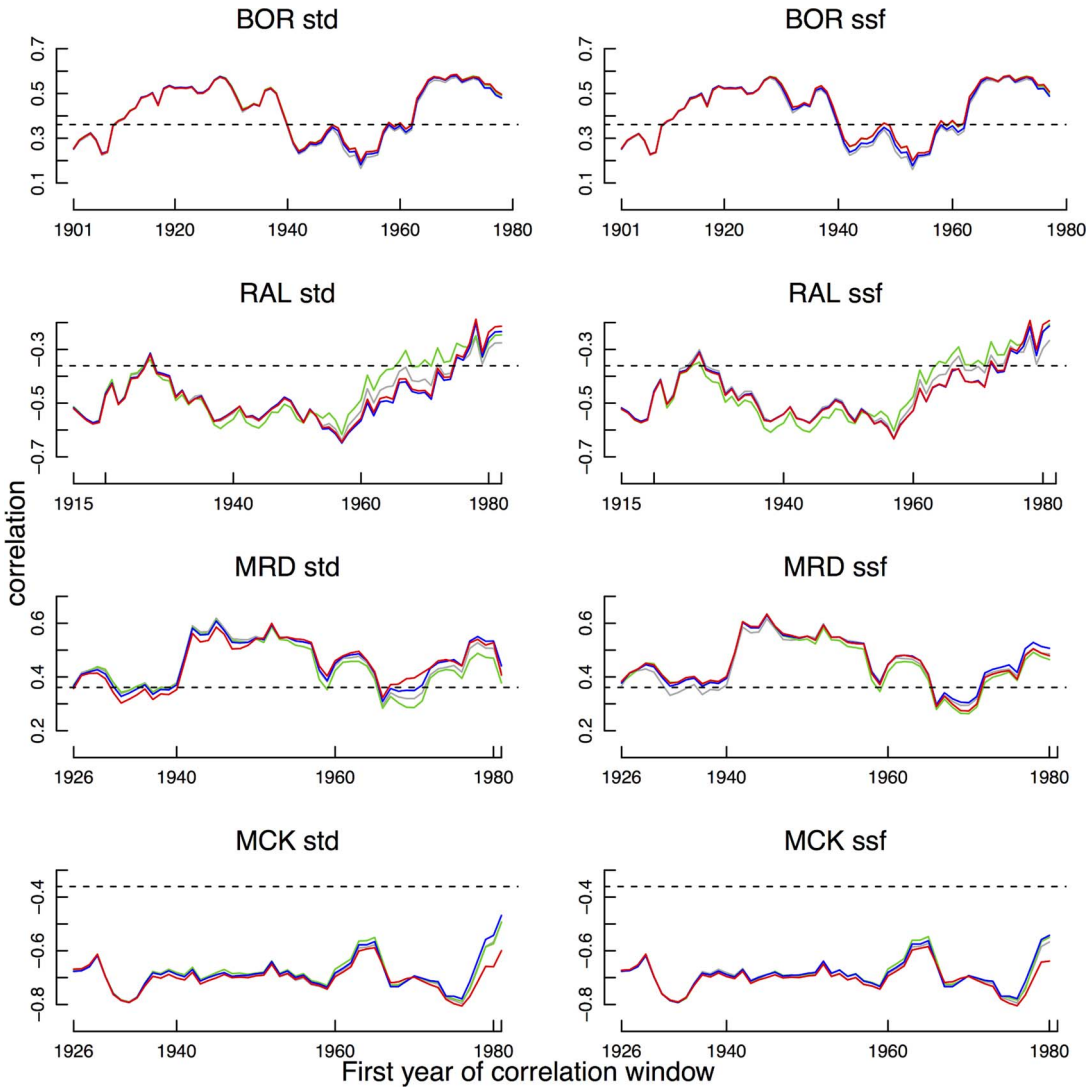

Fig. 7. Running correlations for chronologies based on residuals. Grey is Nex/reg, green is ADspl, blue is Medspl and red is 67%nSpl. Black dashed lines indicate $p \leq 0.05$. Window width is 30 years. Horizontal axes reflect starting date of window, and time-scales differ for the different chronologies. Std denotes standardised but not signal-free chronology, ssf indicates signal-free chronology. Correlations have not been adjusted for serial correlation. Analyses suggested adjustments would differ across sites and standardisations, but in general would be minimal (data not shown). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the results for each method by site (Table 3) illustrates that no one method is consistently sensitive to standardisation – at least for the data used here. Visually, different standardisations made little difference to the resultant series for the synthetic data (Fig. 2), but for the longer series from actual sites, there are clear differences amongst series, particularly at the modern ends (Fig. 6; Table 3). Notably, running correlations appear relatively insensitive to standardisation for both the synthetic and the actual site data although some differences at the modern end were apparent for RAL and the synthetic series (trend intervention; Figs. 3a&b, 7). Overall, it also appears that two-period correlations are not particularly sensitive to standardisation choice (Figs. 3c&d, 8, S5), although this apparent insensitivity may partially reflect the limited options examined here. In contrast, the KF results for the site data suggest some sensitivity to standardisation method used, including whether standardisation was performed within the signal-free framework or not (e.g., Fig. 9, Table 3: RAL Nex/reg, Medspl, MRD Nex/reg, Medspl, BOR Nex/reg and Fig. S5:RAL Medspl, 67%nSpl, MRD ADspl, Medspl, 67%nSpl).

Essentially, all methods used here identified changed relationships due to a high frequency, or interannual, intervention (Figs. 2–4; Table 3) in the synthetic data, but results for the trend intervention were more ambiguous, depending both on method and whether the data were strongly or moderately associated (Figs. 2–4; Table 3). The KF, for example, was highly sensitive to the interannual intervention but failed in all but one case (Nex/reg, strong) to detect the trend intervention. *A priori*, one would expect correlation-based techniques to be more sensitive to changes in the high- rather than the low-frequency relationship (cf. Esper et al., 2010) that is more likely to be attributable to different standardisation approaches. The results for the synthetic data, however, indicate that either running correlations or a comparison of two-period correlations may identify changes in either the low or high frequency relationship between two series (Figs. 3 and 7, Table 3). The correlation techniques were somewhat effective in identifying some change in the low-frequency relationship (Figs. 6, S2) between temperature and actual chronologies (MRD, RAL and MCK), although standardisation method influenced the extent of the change, and hence whether it was significant or not ($p < 0.05$; Figs. 6, S2). Most methods used detected some instability for RAL and MCK (Table 3; Figs. 6–9).

The trend intervention was more readily detected in the strongly, rather than the moderately associated, synthetic series (Figs. 2–4; Table 3). This is consistent with one of the pitfalls discussed in Esper and Frank (2009), namely that reliable detection of a changed relationship between two moderately related series will be relatively

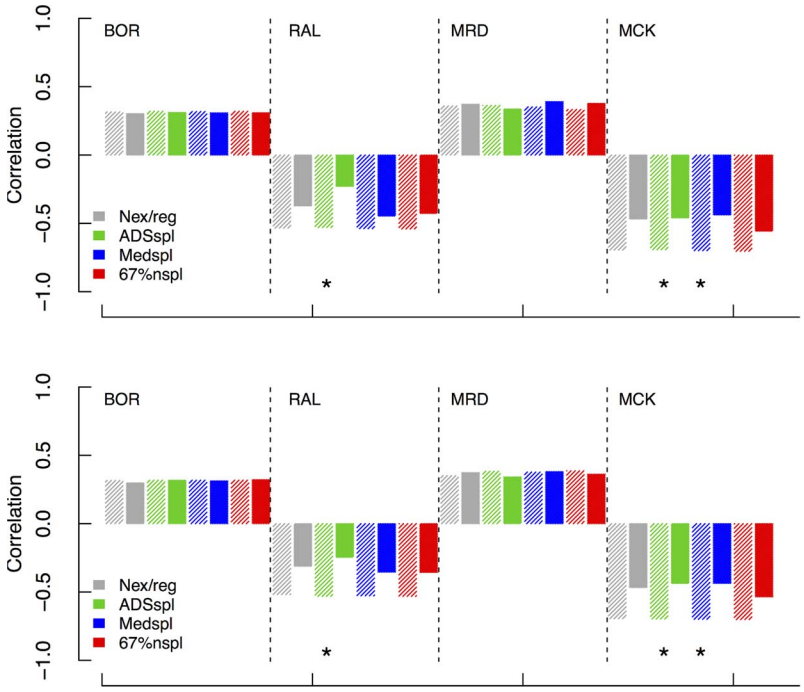

Fig. 8. Two period correlations for chronologies computed using residuals. Periods used were: RAL 1915–1962CE and 1963–2011CE; MRD 1926–1967CE and 1968–2009; BOR 1901–1955CE and 1956–2010CE; MCK 1926–1967CE and 1968–2009CE. Top panel is for non-signal free chronologies, and bottom panel shows signal-free chronologies. Grey represents Nex/reg, green is ADSpl, blue Medspl and red 67%nSpl. Significant differences ($p \leq 0.05$) between early and late correlations are shown as * under early period correlations. See Figure S4 for ratio chronology results. Correlations have not been adjusted for serial correlation. Analyses suggested adjustments would differ slightly across standardisations for any one site, but in general these would be minimal (data not shown). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

challenging. However, the low-frequency modulation found in correlations for the unperturbed strongly associated synthetic series (Fig. 2) suggests the need for some caution in interpreting results of the *g*_test (Table 2). It is similarly possible that some of the apparent sensitivity of the KF to changes in relationships for moderately associated series (Fig. 9; Table 3) may reflect its vulnerability to noisy relationships when those relationships are not strongly coupled (c.f. Esper and Frank, 2009).

While visual assessment of series is highly subjective, it does serve as a means of analytical triage—a first step in identifying whether temporal instability may exist in the relationship between a tree-ring chronology and climate variable (e.g., RAL, MRD, MCK; Fig. 6; Table 3). It may also help identify whether any potential instability is high- or low-frequency (Figs. 2 and 5); however, it does not provide information on the significance or otherwise of any instability. Running correlations offer the potential to identify when a change in the relationship has occurred. Critically, testing for greater (lower) than expected low-frequency modulation is an important part of their application because an apparent instability, especially in moderately associated series, may simply imply a loss of signal for a short period, rather than temporal instability (Gershunov et al., 2001). As applied here, two-period comparisons are inherently constrained by the periods chosen for comparison, making it difficult to detect the timing of any change, or whether multiple changes in the relationship have occurred. Further, the use of very short periods, due to, for example, short instrumental records, may also result in large differences in correlations between two periods. Our results would almost certainly differ if alternative periods were used, all the more so if break points were aligned with any visually apparent changes in relationships. In contrast, an attractive feature of the KF is the simple coefficient traces and a comparison of AIC_c and AIC_t that can highlight what changes occurred in relationships and when they occurred (Figs. 9, S5) without the researcher subjectively imposing breakpoints. These simple traces can also help avoid the 'unequal attention' pitfall (Esper and Frank, 2009), in which attention is only focussed on detecting decoupling at the modern end of series.

The limitations of methods used in this study are amplified by the moderate relationships between the SH tree-ring series and temperature, pointing to the need to carefully consider the limitations of different methods, their ability to distinguish between high and low frequency instability, and the need to apply more than a single approach when investigating many of the SH chronologies. We recognise that our results and conclusions will be constrained by a focus on commonly used approaches to detection of decoupling that does not include methods such as assessment of residuals for autocorrelation in a regression context (e.g., Wilson et al., 2007), process-based modelling (Vaganov et al., 2006), or consideration of changes in coherence over time (e.g., Prisarc et al., 2007). Our approach also focussed on pre-reconstruction methods that examine one chronology at a time. Based on our relatively simple pre-reconstruction approach, it is difficult to determine the likely impact of any apparent decoupling (regardless of whether it is due to standardisation or other factors) on a multi-proxy temperature reconstruction, even if specific corrections are applied to tree-ring data to remove non-climatic trends (see Lavergne et al., in press). This issue, however, is beyond the scope of this paper. At a very general level, most of our methods suggest a decoupling in the RAL chronology (Figs. 6, 8 and 9; Table 3). Therefore, solely using this chronology to reconstruct temperature would likely result in underestimation of past temperatures in a model based on a modern calibration period. Its impact on a multiproxy reconstruction would be mediated by the reconstruction method used and the number of other proxies included.

4.2. Temperature data issues

Detecting changed relationships between climate proxies and targets is crucial, but it is equally important to consider why such changes might occur. For the SH with generally short, and in many regions, quite sparse meteorological station coverage, the climate data may be partially responsible for an identified problem (c.f. Esper et al., 2010; Table 3). We used four measures to test for inhomogeneity in the 'local' temperature data (package iki.dataclim; Orlowsky, 2015). These measures included Pettitt's non-parametric test based on ranks (Pettitt, 1979); the standard normal homogeneity test based on ratios that is particularly sensitive to changes near the start and end of series

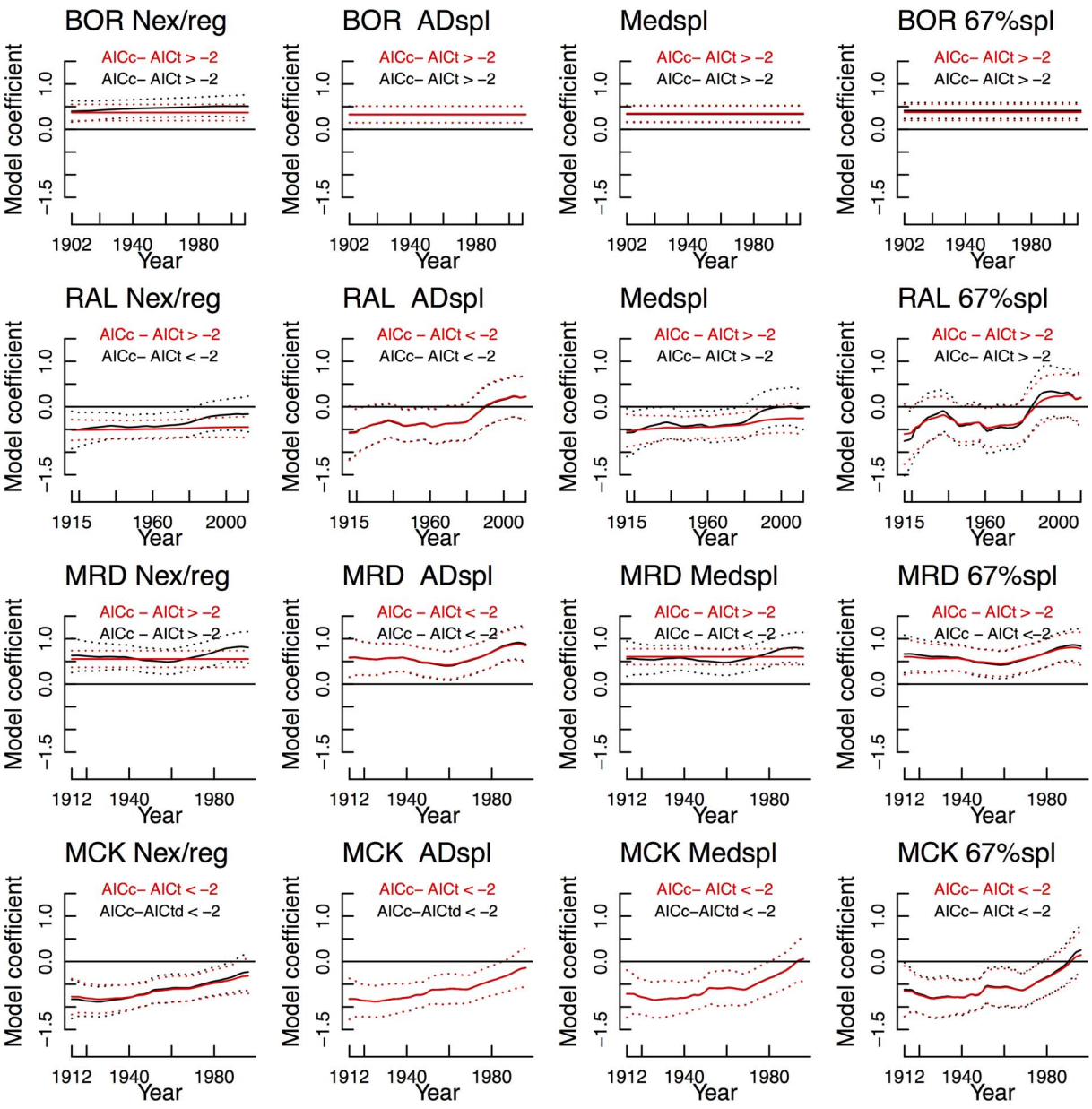

Fig. 9. Kalman Filter traces for the standardised and signal-free SH residual chronologies standardised in different ways. Solid line is KF trace of optimal model and dotted lines are the 2s limits. Red lines and text relate to non-signal-free chronologies and black lines and text relate to signal-free chronologies. Note that x-axes differ across sites. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Alexandersson, 1986); the Buishand Range test based on the distance between the mean and each value (Buishand, 1982); and the von Neumann ratio test that examines the ratio of the successive mean square and the variance (Von Neumann, 1941). Both the Buishand and von Neumann tests assume data normality. Taken together, these tests indicated that climate data for three of the four sites were suspect (Table S1). Only the series for MCK was considered 'useful'.

Reasons for these inhomogeneities are likely to vary. Approximately 21% of the 157 Tasmanian temperature stations sites began operation in the 1960s and ~17% finished operating in the 1990s. The large increase in the number of stations operating in the 1960s may be linked to the breakpoint identified for MCK in 1958 and MRD in 1968 (Figs. 6, S2): note, however that only the MRD series is deemed suspect. It is also possible that the cessation of a relatively large proportion of stations in the 1990s may have had an impact (but not always a significant one) on the modern end of MCK (Figs. 6 and 8).

The breakpoint identified for Bariloche temperatures in 1977 coincides with an abrupt temperature increase from 1976 at many temperature stations in southern South America. Villalba et al. (2003) attributed this abrupt increase to the change from a negative to positive state of the Pacific Decadal Oscillation and also noted anomalously high growth at some high elevation sites of *N. pumilio* since 1976 (see also Lavergne et al., 2015). Notably, the divergence between temperature and the RAL chronology begins to become visually apparent from this point (Figs. 6 and 7). It is also possible that the anomalously high growth at the Mt Read site in most of the different versions of the

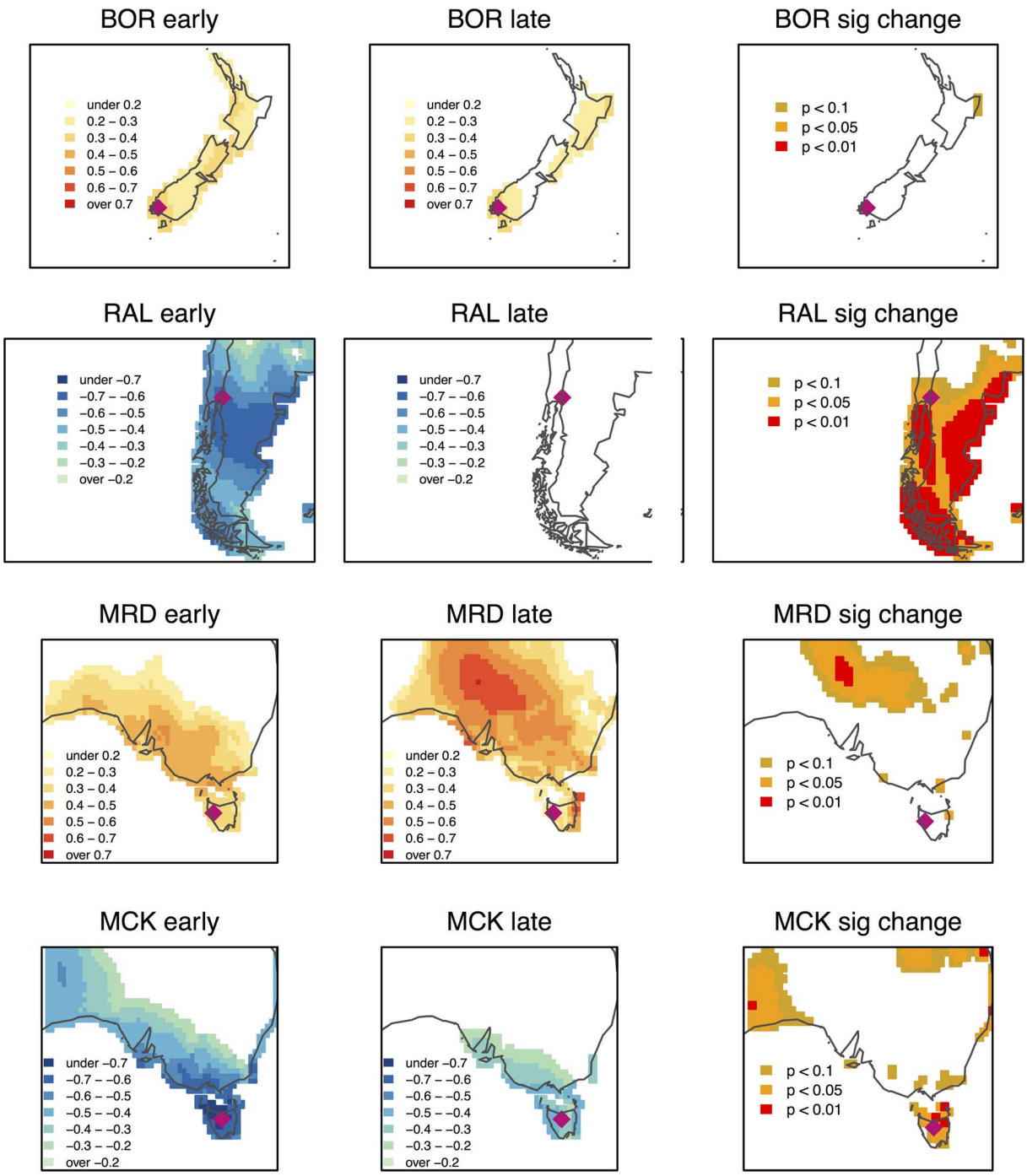

Fig. 10. Spatial correlations between 67%nSpl chronologies and CRU temperature for two periods. Left panel: early period, middle panel: late period correlations. Right panel shows significant differences in correlation between temperature and chronologies over the two periods. For periods used, see text. White indicates area for which no significant correlations were found. The RColorBrewer package (Neuwirth 2014) was used to create the colour schemes. Note that the actual values at which correlations are significant differ for the different sites due to differing length of periods used. Correlations are significant ($p \leq 0.05$) at $|r| \geq 0.254$ (MRD and MCK); $|r| \geq 0.231$ (BOR); $|r| \geq 0.367$ RAL. Also, significance of correlations (left-hand middle columns) has not been adjusted for serial correlation. Results for local data suggested adjustments would be minimal (data not shown).

chronology from ~1980 is associated with related climatic changes (Allen et al., 2014). The sink inhibition hypothesis discussed by Körner (1998) was invoked by Salzer et al. (2009) to explain accelerated growth in Bristlecone Pine at tree line relative to increasing temperatures. This hypothesis would also be consistent with the visible changes in the relationships for RAL and MRD. The visually apparent change in relationship for MCK might also be a direct response to temperature changes. According to the Australian Bureau of Meteorology (http://www.bom.gov.au/climate/change) mean summer temperatures have been ≥ 0.3° above average almost every summer since ~2000. If indeed increased temperatures have been responsible for the apparent changes in relationships, then selecting a standardisation method

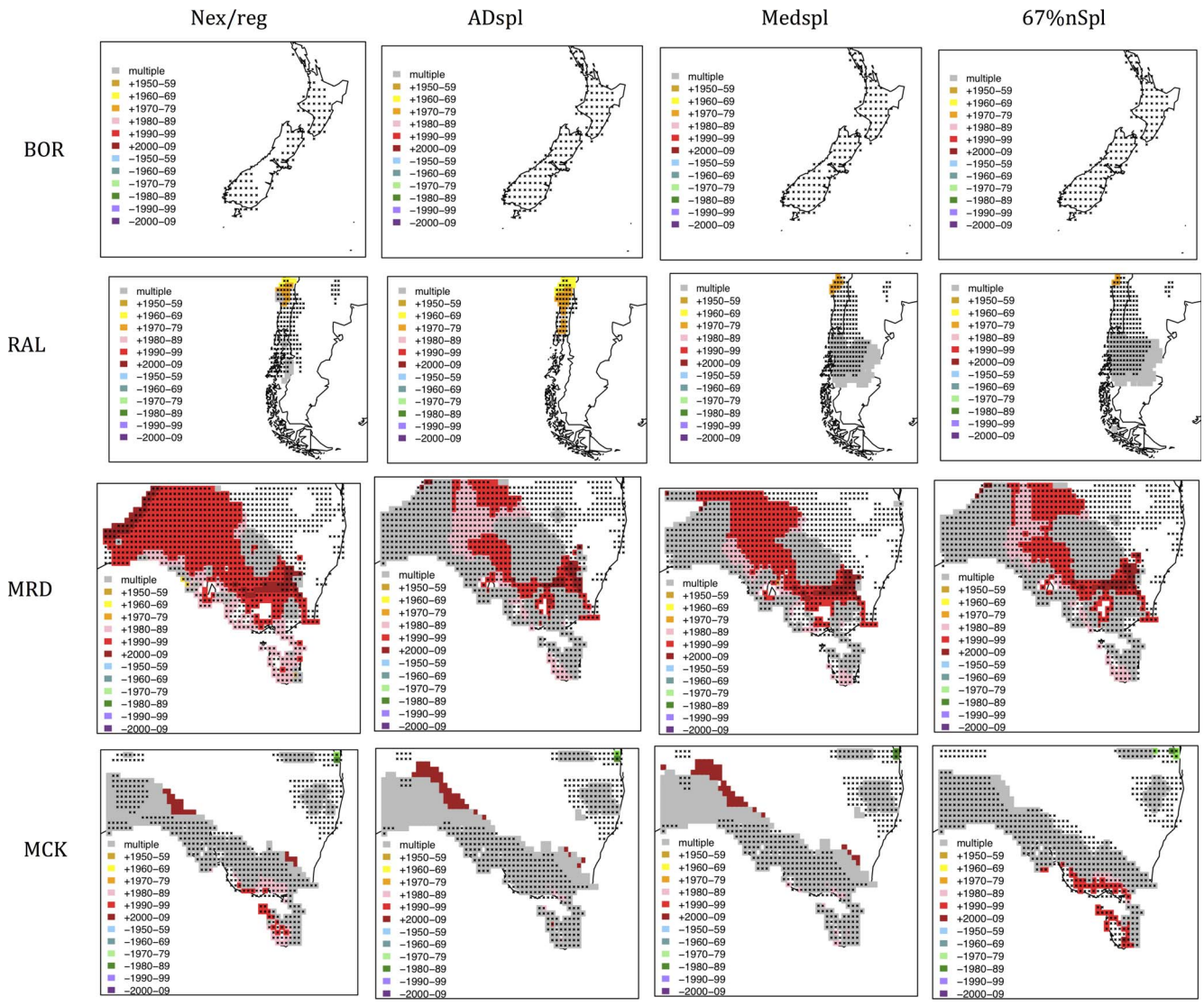

Fig. 11. Application of KF to relationships between gridded (0.5° × 0.5°) CRU3.22 mean temperature data and chronologies. Only grid squares for which relationships had $AIC_c - AIC_t < -2$ AND a significant ($p < 0.05$) change in sign are shaded. Both positive (coefficient of the relationship increases) and negative (coefficient of the relationship decreases) decoupling is shown. Different colours relate to whether decoupling is positive or negative and its timing. Instances of multiple changes in the sign of the relationship are shown in grey. A small black cross indicates that there is a significant linear relationship between temperature and the tree ring chronology over their common period for that grid square. A black cross on a white background indicates a significant linear relationship that did not change over the period examined.

purely on the basis that it minimises any apparent divergence would be inappropriate.

Possible causes of the 1953 inhomogeneity in the New Zealand data are less clear, especially because the process used to generate the CRU dataset incorporated inhomogeneity tests. Interestingly, the decrease in running correlations for BOR that begins in the 30-year window starting in 1940 and ending ~ 1960 appears connected to a few years with relatively high temperatures around 1960.

Table 3
Summary of results for individual sites by technique used. A * indicates decoupling issues common across the chronology suite; end (mid-series) indicates issue at modern end (mid-series) of period in common between temperature data and chronology, and standardisation means issues were identified for some standardisations. Note that no spatial analyses were undertaken for the synthetic data.

Detection method	Visual	Running correlation traces	Two period correlations	KF	spatial 2 period	spatial KF
BOR		* (mid-series)				
RAL	* (end)	*	* (standardisation)	* (standardisation)	*	*
MRD	* (end)					*
MCK	* (end)		* (standardisation)	* (standardisation)	*	*
Synthetic strong (no change)		*			NA	NA
Synthetic strong (trend perturbation)	*	*	*	(standardisation)	NA	NA
Synthetic strong (interann perturbation)	*	*	*	*	NA	NA
Synthetic moderate (no change)	*	*			NA	NA
Synthetic moderate (trend perturbation)	*	*			NA	NA
Synthetic moderate (interann perturbation)	*	*	*	*	NA	NA

4.3. Spatial domain instability

Temporal instability across space is also suggested for three of our four sites. While spatio-temporal instability of relationships between climate and climate proxies is not the same thing as the DP, it is just as relevant an issue to consider before embarking on a broadscale climate field reconstruction. The spatial extent of a stable relationship should help define the domain over which it is most appropriate to use a particular proxy in a reconstruction. In the context of our study, there are several reasons why spatio-temporal instability may arise. The moderate strength of the relationship with temperature means that the apparent instability (RAL, MRD, MCK) or stability (BOR) for some grid cells is likely to be a statistical artefact. Somewhat related to this is the problem of testing the same hypothesis across many grid cells simultaneously. The number of false positives will be proportional to the significance level being used (10, 5 and 1% for the two-period comparisons, and 10% for the KF). Correlation of two series should also be adjusted for the level of serial correlation in the two series. Appropriate adjustment made minimal differences in 'local level' results (see Figs. 7 and 8) but had a greater impact on the spatial results, particularly for MRD (data not shown).

The drop-off in the number of meteorological records back in time means gridded products are likely to be less reliable in their early portions. This is especially relevant to topographically diverse regions such as western Tasmania, New Zealand and southern/western South America. Further, the use of the two halves of the short 1954–2010 period for southern South America meant that the second period began in 1982, just five years after a change in the climate regime of the region (Villalba et al., 2003), almost guaranteeing a large change across the two periods. Lastly, the climate in each of these regions is subject to multiple ocean-atmosphere processes (e.g., the SAM, subtropical ridge, ENSO, the Interdecadal Pacific Oscillation) that interact in a complex and nonstationary manner (Risbey et al., 2009; Gallant et al., 2013) and will affect parts of the domains shown in Figs. 10 and 11 and S6–8 differentially and intermittently. Disentangling these possible sources of spatio-temporal instability will be a complex task, particularly when correlations between a tree-ring chronology and climate are moderate at best. Admittedly, the relative lack of long-term (pre-1950), high-quality climate data for much of the SH makes this task even more challenging. Twentieth century reanalysis products may offer some potential to address this apparent impasse, however (Compo et al., 2011).

5. Conclusions

Our review of methods commonly used to detect temporal instability in relationships between climate and tree-ring chronologies is certainly not exhaustive and is, to some degree, conditional on the particular data being examined. Our analyses identify the relative difficulty in identifying changing relationships in trends between series that are modestly associated with one another—as is typical of many of the multicentennial SH tree-ring chronologies. They also highlight that commonly used methods are generally more effective in detecting changes in interannual relationships between series than changes in trend. This is particularly relevant when it has primarily been the changed relationship between trends in the tree-ring and climate series that has been of interest and points to the need for careful consideration of methods used to examine these relationships for instability. Each of the methods we examined has its limitations, with the KF and running correlations better able to identify when changes occurred and provide an indication of the significance of these changes. In general, the correlative methods were relatively insensitive to standardisation approach, but the Kalman filter may be overly sensitive to noise in moderately correlated series. Given the variation in results across methods for some sites (BOR, MRD, MCK), it would be unwise to rely on the results of a single method to identify the presence or absence of temporal instability in relationships, particularly for moderately associated series. Although spatio-temporal instability is a problem somewhat distinct from the Divergence Problem, it is important to consider when undertaking large-scale reconstructions.

Acknowledgements

Australian tree-ring data for this study were obtained under permits issues by Parks Tasmania. Rob Evans and Michael Goddard were instrumental in obtaining the cell wall thickness data from Silviscan3. This research was funded by Australian Research Council grants DP120104320 to PJB, ERC, and JGP and FT 12010715 to PJB. David Drew received funding from the Herman-Slade Foundation (HS 09/5). Lamont-Doherty contribution no. XXXX. Data for the four sites examined here are, or will become available very shortly, at the International Tree-Ring Databank (ITRDB). We also thank the anonymous reviewers whose comments have helped to significantly improve this work.

Appendix A. A list of previous studies on the DP. The majority of studies target temperature, but some also examine responses to precipitation, drought or climate indices (e.g., Àlvarez et al., 2015). Most also focus on either ring width or density, but there is an increasing number of studies that are using isotopic or other properties to reconstruct climate, and hence, some studies also examine stability of the relationship of these parameters with climate (e.g., Barber et al., 2000; Andreu-Hayles et al., 2011a,b; Franceshini et al., 2012; Matisons et al., 2012). As also demonstrated in the table, interest in the DP extends beyond the high latitude or high elevation chronologies in the NH (e.g., Cook and Johnson 1989; Andreu-Hayles et al., 2011b; Allen et al., 2014). Interest in the implication of changing responses to climate also extends beyond those engaged in climate reconstruction (e.g., Cook et al., 1987; Lavergne et al., 2015; Sànchez-Salguero Camarero et al., 2017; Tumajer et al., 2017). The list of studies in the table is not exhaustive

Country/ region	Species	Type of tree-ring data	Target climate variable	Methods used to check for stability	Factor considered most likely responsible for any change	High or low frequency divergence identified/ discussed	Standardisation method	Impact of standardisation examined	Reference
Alaska, USA	Picea glauca	RW/ MXD	May-Aug mean temperature	Visual comparison (observed vs reconstructed temperature)	Change in factor most limiting to growth	Low	No information provided	N	Jacoby and D'Arrigo (1995)
NH	Various	MXD	Apr-Sep mean temperature	Visual comparison	Changes in stratospheric ozone concentration	Low	Age-band decomposition standardisation	N	Briffa et al. (2004)
NH	Various	RW/ MXD	Apr-Sep mean temperature	Two-period comparison, moving correlations, visual comparison	Some change occurring early in growing season (esp. MXD)	Low, possibly also high in some cases	Hugershoff, residuals	N	Briffa et al. (2002)
Italian Alps	Larix decidua	RW	Mean monthly temperature, precipitation	Moving correlations	Climate threshold effects?	NA	Negative exponential, ratio	N	Carrer and Urbinati (2006)
NH	Various	RW/ MXD		Visual comparison	Case dependent: detrending, autocorrelation, issues with climate data	Low?	Used others' reconstructions	Y	Frank et al. (2007)
NH	Various	RW/ MXD	Mean temperature	Significant autocorrelation of residuals	NA	Low (implied)	Various	N	Wilson et al. (2007)
Switzerland	Picea abies	RW	Annual/ monthly mean, min, max temperature, precipitation	Moving correlations	Change in limiting factor	Not specified	RCS	N, but mentioned	Büntgen et al. (2006b)
Switzerland/ Austria	Larix decidua	MXD	Jun-Sep mean temperature	Moving correlations	Climate data change, elevation differences between tree-ring sites and climate data	Not specified (low implied)	RCS	N	Büntgen et al. (2006a)
European Alps	Larix decidua, Picea abies	RW	Jun-Jul mean temperature	Visual comparison	Tree-age, standardisation, calibration approach	Not specified	300-yr splines, RCS	Y	Büntgen et al. (2008)
European Alps	Larix decidua	RW	Jun-Aug mean temperature	Two-period correlation comparison, moving correlations	Use of ratios in chronology building, standardisation method, climate data issues	Not specified	50%n spline, cut-off 150 − 300 yrs, negative exponential, RCS	Y	Büntgen et al. (2012)
Siberia, Russia	Various (larch, spruce, pine)	RW/ MXD	Mean temperature	Visual comparison, assessment of residual between calibrated tree-ring and target instrumental time series	Inhomogeneity in climate data, standardisation used, calibration	Not specified (low implied)	Various (RCS, Hugershoff, negative exponential, 300-yer spline	Y	Esper et al. (2010)
Chile	Nothofagus pumilio	RW	AAO, ENSO (MEI)	Moving correlations, crosswavelet analysis	Climate shift	Not specified (low implied)	Negative exponential/ linear regression	N	Àlvarez et al. (2015)

Pyrenees	Pinus uncinata	RW/MXD	Mean, max, min temperature, precipitation, SPEI	Visual, comparison, three-period correlation comparison, moving correlations, multi-period spatial correlations	Changes in limiting factors	Low	50%n splines, cut-off 150 and 300 yrs, negative exponential/linear regression, RCS	Y	Galvàn et al. (2015)
Argentina	Nothofagus pumilio	RW	Monthly mean temperature, precipitation	Visual comparison, two-period correlations, moving correlations, VS-Lite model	Changes in limiting factors, threshold effects	High	10-yr cubic smoothing spline	N	Lavergne et al. (2015)
Alaska, USA	Picea glauca	RW/MXD	Mean summer temperature, precipitation, PDSI	Two-period correlation comparison	Phenological changes linked to climate change	High	140-yr 50% cut-off spline, negative exponential/linear regression, RCS, raw data	N	Andreu-Hayles et al. (2011a)
Iberia	Pinus nigra, Pinus sylvestris, Pinus uncinata	RW/d13C	Mean, minimum, maximum temperature, precipitation, VPD	Visual comparison	Methodological approach? Changes in water use efficiency?	Not specified (high implied)	None	NA	Andreu-Hayles et al. (2011b)
Latvia	Quercus robur	RW, EW vessel area	Mean temperature	Moving correlation and response functions	Changes in limiting factors, tree age	Not specified (low implied)	Negative exponential	N	Matisons et al. (2012)
California, USA	Pinus longaeva	RW	Sep-Aug mean temperature, precipitation	Visual comparison	Increasing temperature	Not specified (low)	None	NA	Salzer et al. (2009)
Iberia	Pinus nigra	RW	Mean temperature, precipitation	Visual comparison, moving correlations	Changes in limiting factors (including season). Greater (lesser) spring (summer) sensitivity in latter years	Not specified (low implied)	Negative exponential/linear regression	N	Martìn-Benito et al. (2010)
France	Picea abies	EWD/LWD	Jul-Sep mean temperature	Visual comparison, moving correlations	NA	Low	Developed linear mixed-effects models	N	Franceshini et al. (2012)
Spain	Abies alba, Pinus sylvestris, Pinus uncinata	RW	Mean and minimum temperature, precipitation, soil moisture	VS-Lite	Climate change (but study concerned with projections, not retrospective)	Not specified	25-year splines	N	Sànchez-Salguero Camarero et al. (2017)
Czech Republic	Picea abies	RW	Temperature, precipitation	VS-Lite	Changes in limiting factor	Not specified	Double detrending: negative exponential, 50% frequency-response cut-off 67 yrs		Tumajer et al. (2017)
Austria	Pinus nigra	RW	Precipitation	Visual comparison, moving correlations, pointer year analysis	Increasing CO2 concentration? Increased N deposition?	Not specified (high implied)	Negative exponential	N	Leal et al. (2008)
Canada	Picea mariana	d18O, d13C	Precipitation, maximum temperature	Moving correlations	Increased moisture stress (d13C)	Not specified	Not specified	NA	Naulier et al. (2015)

NE USA	Picea rubens	RW	Mean temperature	Kalman Filter	Elevation-dependent change in limiting factors	Not specified (low implied)	Not specified	N	Cook and Johnson (1989)
Australia	Lagarostrobos franklinii	RW	Nov–Apr mean temperature	Visual comparison, Kalman Filter	Issues related to climate data	Low	Signal-free negative exponential/negative linear	N	Allen et al. (2014)
Eastern USA	Picea rubens	RW	Mean temperature, precipitation	Comparison of modelled with actual growth (based on step-wise linear regression)	Climate stress, atmospheric pollution	Not specified (low implied)	Not directly specified	N	Downing and McLaughlin (1987)
Alaska, USA	Picea glauca	MXD	Jul–Aug mean temperature	Visual comparison, 'temporal' scatter plot, pseudo-proxy experiment	NA	None	RCS, signal-free, data-adaptive	Y	Anchukaitis et al. (2013)
Austria	Pinus cembra	RW	Jul mean temperature	Moving response function	Change in limiting factor, tree physiology, climate?	Both low and high implied	Double detrending: negative exponential, 67%n cubic spline	N	Oberhuber et al. (2008)
Russia	Not specified (likely Larix Larix gmelinii, Larix sibirica)	RW/ MXD	Summer temperature	PRECON Simulation model	Increasing winter precipitation, delaying on set cambial activity	Not specified (low implied)	Not specified	N	Vaganov et al. (1999)
Alaska, USA	Picea galuca	RW/ MXD/ d13C	May–Aug mean temperature	Visual comparison	Change in limiting factors	Not specified (low implied)	None	NA	Barber et al. (2000)
NH high latitude	Picea galuca, Larix dahurica, Larix sibirica, Pinus sylvestris	RW	Mean temperature	Visual comparison	Increased temperatures and sensitivity to microsite conditions	Not specified (low implied)	Negative exponential/ linear regression	N	Wilmking et al. (2005)
Siberia, Russia	Larix gmelinii	RW	May–Sep mean temperature	Visual comparison	Not specified	Low	Negative exponential/ linear regression (zero/negative slope)	N	Jacoby et al. (2000)
NH	Various	RW/ MXD	Mean temperature	Visual comparison	Various hypotheses suggested: e.g., CO2, increased pollutants, increased UV-B, changes in soil chemistry	Low	RCS, more flexible technique (not explicitly identified, see reference for description)	Y	Briffa et al. (1998)
Alaska, USA	Not directly specified (Picea glauca implied)	RW	Mean temperature	Three-period correlation comparison	Change in limiting factor	Not specified	Negative exponential	N	Lloyd and Fastie (2002)
Canada	Picea glauca	RW	Mean temperature	Two-period correlation comparison	Threshold effects	Not specified (low implied but high somewhat apparent also)	Not specified	N	D'Arrigo et al. (2004)
Alaska, USA	Picea glauca	RW	Apr–Jul mean temperature	Visual comparison of 'positive' and 'negative' responders to climate series	Different climate responses of sub-groups	Not specified (low implied)	Negative exponential	N	Driscoll et al. (2005)

Italian Alps	Larix decidua	RW	Mean temperature, precipitation	Two-period correlation comparison, moving correlations	Not specified	Negative exponential, ratio (subsequently 50% frequency response spline of 20 yrs)	N	Carrer and Urbinati (2006)
Central Alps	Pinus cembra	RW	May–Sep mean temperature	Comparison of average ring width for age classes through time	CO2 increase	Not specified (low implied)	Form of RCS	N	Nicolussi et al. (1995)
Alaska, USA	Picea glauca	RW/ d13C	Temperature, drought	Moving correlations, variance analysis	Unsure, probably multiple. Not drought stress	Not specified, (low implied)	Signal-free RCS, no pith adjustment	Alluded to	Brownlee et al. (2015)

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.dendro.2018.02.002.

References

Àlvarez, C., Veblen, T.T., Christie, D.A., Gonzàlez-Reyese, À., 2015. Relationships between climate variability and radial growth of Nothofagus pumilio near altitudinal treeline in the Andes of northern Patagonia, Chile. For. Ecol. Manage. 342, 112–121.

Alexandersson, H., 1986. A homogeneity test applied to precipitation data. Int. J. Climatol. 6, 661—675.

Allen, K.J., Ogden, J., Buckley, B.M., Cook, E.R., Baker, P.J., 2011. The potential to reconstruct broadscale climate indices associated with Australian droughts from Athrotaxis species, Tasmania. Clim. Dyn. 37, 1799–1821.

Allen, K.J., Drew, D.M., Downes, G.M., Evans, R., Cook, E.R., Battaglia, M., Baker, P.J., 2013. A strong regional temperature signal in low-elevation Huon pine. J. Quat. Sci. 28, 433–438.

Allen, K.J., Cook, E.R., Buckley, B.M., Larsen, S.H., Drew, D.M., Downes, G.M., Francey, R.J., Peterson, M.J., Baker, P.J., 2014. Continuing upward trend in Mt Read Huon pine ring widths –temperature or divergence? Quat. Sci. Rev. 102, 39—53.

Allen, K.J., Cook, E., Evans, R., Francey, R., Buckley, B., Palmer, J., Peterson, M., Baker, P., 2018. Lack of cool, not warm extremes, distinguishes late 20th Century climate in 979-year Tasmanian summer temperature reconstruction. Environmental Research Letters In Revision.

Ambrizzi, T., Souza, E.B., Pulwarty, R.S., 2004. The Hadley and Walker regional circulations and associated impacts on South American seasonal rainfall. Ch. 7. In: Diaz, H.F., Bradley, R.S. (Eds.), The Hadley Circulation: Present, Past and Future. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Anchukaitis, K.J., D'Arrigo, R.D., Andreu-Hayles, L., Frank, D., Verstege, A., Curtis, A., Buckley, B.M., Jacoby, G.C., Cook, E.R., 2013. Tree-ring reconstructed summer temperatures from northwestern North America during the last nine centuries. J. Clim. 26, 3001—3012.

Andreu-Hayles, L., D'Arrigo, R.D., Anchukaitis, K.J., Beck, P.S.A., Frank, D., Goetz, S., 2011a. Varying boreal forest response to Arctic environmental change at the Firth River, Alaska. Environ. Res. Lett. 6. http://dx.doi.org/10.1088/1748-9326/6/4/049502.

Andreu-Hayles, L., Planells, O., Gutièrrez, E., Muntan, E., Helle, G., Anchukaitis, K., Schleser, G.H., 2011b. Long tree-ring chronologies reveal 20th century increases in water-use efficiency but no enhancement of tree growth at five Iberian pine forests. Global Change Biol. 17, 2095–2112.

Büntgen, U., Frank, D.C., Schmidhalter, M., Neuwirth, B., Seifert, M., Esper, J., 2006a. Growth/climate response shift in a long subalpine spruce chronology. Trees 20, 99–110.

Büntgen, U., Frank, D.C., Schmidhalter, M., Neuwirth, B., Seifert, M., Esper, J., 2006b. Growth/climate response shift in a long subalpine spruce chronology. Trees 20, 99–110.

Büntgen, U., Frank, D., Wilson, R., Carrer, M., Urbinati, C., Esper, J., 2008. Testing for tree-ring divergence in the European Alps. Global Change Biol. 14, 2443—2453.

Büntgen, U., Frank, D., Neuenschwander, T., Esper, J., 2012. Fading temperature sensitivity of Alpine tree growth at its Mediterranean margin and associated effects on large-scale climate reconstructions. Clim. Change 114, 651–666. http://dx.doi.org/10.1007/s10584-012-0540-4.

Barber, V.A., Juday, G.P., Finney, B.P., 2000. Reduced growth of Alaskan white spruce in the twentieth century from temperature-induced drought stress. Nature 405, 668–673.

Bianchi, E., Villalba, R., Viale, M., Courreux, F., Marticorena, R., 2016. New precipitation and temperature grids for northern Patagonia: advances in relation to global climate grids. J. Meteorol. Res. 30, 38–52.

Briffa, K.R., Melvin, T.M., 2011. A closer look at Regional Curve Standardisation of tree-ring records: a justification of the need, a warning of some pitfalls, and suggested improvements in its application. Ch 5. In: Hughes, M.K., Diaz, H.F., Swetnam, T.W. (Eds.), Dendroclimatology: Progress and Prospects. Springer, Verlag.

Briffa, K.R., Schweingruber, F.H., Jones, P.D., Osborn, T.J., Harris, I.C., Shiyatov, S.G., Vaganov, E.A., Grudd, H., 1998. Trees tell of past climates: but are they speaking less clearly today? Philos. Trans. R. Soc. Lond. B 353, 65–73.

Briffa, K.R., Osborn, T.J., Schweingruber, F.H., Jones, P.D., Shiyatov, S.G., Vaganov, E.A., 2002. Tree-ring width and density data around the Northern Hemisphere: part I local and regional climate signals. Holocene 12, 737–757.

Briffa, K.R., Osborn, T.J., Schweingruber, F.H., 2004. Large-scale temperature inferences from tree rings: a review. Global Planet. Change 40, 11–26.

Briffa, K.R., Melvin, T.M., Osborn, T.J., Hantemirov, R.M., Kirdyanov, V.S., Shiyatov, S.G., Esper, J., 2013. Reassessing the evidence for tree-growth and inferred temperature change during the Common Era in Yamalia, northwest Siberia. Quat. Sci. Rev. 72, 83–107.

Brownlee, A., Sullivan, P.F., Csank, A.Z., Sveinbjörnsson, Ellison, S.B.Z., 2015. Drought-induced stomatal closure probably cannot explain divergence in white spruce growth in the Brooks Range, Alaska. Ecology 97, 145–159. http://dx.doi.org/10.1890/15-0338.1.

Buishand, T.A., 1982. Some methods for testing the homogeneity of rainfall records. J. Hydrol. 58, 11–27.

Carrer, M., Urbinati, C., 2006. Long-term change in the sensitivity of tree-ring growth to

climate forcing in Larix decidua. New Phytol. 170, 861–871.
Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D., Worley, S.J., 2011. The twentieth century reanalysis project. Q. J. R. Meteorolog. Soc. 137, 1–28.
Cook, E.R., Johnson, A.H., 1989. Climate change and forest decline: a review of the red spruce case. Water Air Soil Pollut. 48, 127–140.
Cook, E.R., Peters, K., 1997. Calculating unbiased tree-ring indices for the study of climatic and environmental change. Holocene 7, 361—370.
Cook, E.R., Johnson, A.H., Blasing, T.J., 1987. Forest decline: modeling the effect of climate in tree rings. Tree Physiol. 3, 27–40.
Cook, E.R., Bird, T., Peterson, M., Barbetti, M., Buckley, B., D'Arrigo, R., Francey, R., Tans, P., 1991. Climatic change in Tasmania inferred from a 1089-year chronology of Huon Pine. Science 253 (5025), 696–698.
Cook, E.R., Buckley, B.M., D'Arrigo, R.D., Peterson, M.J., 2000. Warm-season temperatures since 1600BC reconstructed from Tasmanian tree rings and their relationship to large-scale sea surface temperature anomalies. Clim. Dyn. 16, 79–91.
Cook, E.R., Palmer, J.G., Ahmed, M., Woodhouse, C.A., Fenwick, P., Zafa, M.U., Wahab, M., Khan, N., 2013. Five centuries of Upper Indus River flows from tree rings. J. Hydrol. 486, 365–375.
Cook, E.R., 1985. A Time Series Approach to Tree-ring Standardisation PhD Thesis, Tree-Ring Laboratory, Lamont-Doherty Geological Earth Observatory. Palisades, New York.
D'Arrigo, R.D., Wilson, R., Liepert, B., Cherubini, P., 2008. On the 'divergence problem' in northern forests: a review of the tree-ring evidence and possible causes. Global Planet. Change 60, 289–305.
D'Arrigo, R.D., Jacoby, G.C., Buckley, B.M., Sakulich, J., Frank, D., Wilson, R., Curtis, A., Anchukaitis, K., 2009. Tree growth and inferred temperature variability at the North American Arctic tree line. Global Planet. Change 65, 71–82.
Downing, D., McLaughlin, S.B., 1987. Intervention detection-a systematic technique for examining shifts in radial growth rates of forest trees. In: Jacoby, G.C., Hornbeck, J.W. (Eds.), Proceedings of the International Symposium on Ecological Aspects of Tree-Ring Analysis. United States Department of Energy, pp. 543–554.
Drew, D.M., Allen, K.J., Downes, G.M., Evans, R., Battaglia, M., Baker, P.J., 2013. Wood properties in a long–lived conifer reveal strong climate signals where ring width series do not. Tree Physiol. http://dx.doi.org/10.1093/treephys/tps111.
Driscoll, W.W., Wiles, G.C., D'Arrigo, R.D., Wilmking, M., 2005. Divergent tree growth response to recent climatic warming, Lake Clark National Park and Preserve, Alaska. Geophys. Res. Lett. 32.
Esper, J., Frank, D., 2009. Divergence pitfalls in tree-ring research. Clim. Change 94, 261—266.
Esper, J., Frank, D., Büntgen, U., Verstege, A., Hantemirov, R.M., Kirdyanov, A.V., 2010. Trends and uncertainties in Siberian indicators of 20th century warming. Global Change Biol. 16, 386–398.
Evans, M.N., Kaplan, A., 2004. The Pacific sector Hadley and Walker circulation in historical marine wind analyses. Ch. 8. In: Diaz, H.F., Bradley, R.S. (Eds.), The Hadley Circulation: Present, Past and Future. Kluwer Academic Publishers, Dordrecht The Netherlands.
Fisher, R.A., 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 10, 507–521.
Franceshini, T., Bontemps, J.-D., Leban, J.-M., 2012. Transient historical decrease in earlywood and latewood density and unstable sensitivity to summer temperature for Norway spruce in northeastern France. Can. J. For. Res. 42, 219–226.
Frank, D., Büntgen, U., Böhm, R., Maugeri, M., Esper, J., 2007. Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. Quat. Sci. Rev. 26, 3298–3310.
Gallant, A.J.E., Phipps, S., Karoly, D.J., Mullan, A.B., Lorrey, A.M., 2013. Nonstationary Australasian teleconnections and implications for paleoclimate reconstructions. J. Clim. 26, 8827–8849.
Galvàn, J.D., Büntgen, U., Ginzler, C., Grudd, H., Gutièrrez, E., Labuhn, I., Camarero, J.J., 2015. Drought-induced weakening of growth-temperature associations in high-elevation Iberian pines. Global Planet. Change 124, 95–106.
Gershunov, A., Schneider, N., Barnett, T., 2001. Low-frequency modulation of the ENSO-Indian monsoon rainfall relationship: signal or noise? J. Clim. 14, 2486–2492.
Grudd, H., 2008. Torneträsk tree-ring width and density AD 500-2004: a test of climatic sensitivity and a new 1500-year reconstruction of north Fennoscandian summers. Clim. Dyn. 31, 843–857.
Harvey, A.C., 1989. Forecasting, Structural Time Series and the Kalman Filter. Cambridge University Press, Cambridge (554pp.).
Jacoby, G.C., D'Arrigo, R.D., 1995. Tree ring width and density evidence of climatic and potential forest change in Alaska. Global Biogeochem. Cycles 9, 227–234.
Jacoby, G.C., Lovelius, N.V., Shumilov, O.I., Raspopov, O.M., Karbainov, J.K., Frank, D.C., 2000. Long-term temperature trends and tree growth in the Taymir region of northern Siberia. Quat. Res. 53, 312–318.
Jones, D.A., Trewin, B., 2000. The spatial structure of monthly temperature anomalies over Australia. Aust. Meteorol. Mag. 49, 261–276.
Jones, D., Wang, W., Fawcet, R., 2009. High-quality spatial climate data sets for Australia. Aust. Meteorol. Oceanogr. J. 58, 233–248.
Körner, C., 1998. A re-assessment of high elevation treeline positions and their explanation. Oecologia 115, 445–459.
Lara, A., Villalba, R., 1993. A 3620-year temperature record from Fitzroya cupressoides tree rings in southern South America. Science 260, 1104–1106.
Lavergne, A., Daux, V., Villalba, R., Barichivich, J., 2015. Temporal changes in climatic limitation of tree-growth at upper treeline forets: contrasted responses along the west-to-east humidity gradient in Northern Patagonia. Dendrochronologia 36, 49–59.
Lavergne, A., Daux, V., Pierre, M., Stievenard, M., Srur, A.M., Villalba, R., 2018. Past summer temperatures inferred from dendrochronological records of Fitzroya cupressoides on the eastern slope of the northern Patagonian Andes. J. Geophys. Res.–Biogeosci. JGRG20942. http://dx.doi.org/10.1002/2017JG003989.
Leal, S., Eamus, D., Grabner, M., Wimmer, R., Cherubini, P., 2008. Tree-rings of Pinus nigra from the Vienna basin region (Australia) show evidence of change in climatic sensitivity in the late 20th Century. Can. J. For. Res. 38, 744–759.
Linderholm, H.W., Gunnarson, B.E., Liu, Y., 2010. Comparing Scots pine tree-ring proxies and standardisation methods among sites in Jämtland, west-central Scandinavia. Dendrochronologia 28, 239–249.
Martìn-Benito, D., del Rìo, M., Cañellas, I., 2010. Black pine (Pinus nigra Arn.) growth divergence along a latitudinal gradient in Western Mediterranean mountains. Ann. For. Sci. 67. http://dx.doi.org/10.1051/forest/2009121.
Matisons, R., Elferts, D., Br¬¬umelis, G., 2012. Changes in climatic signals of English oak tree-ring width and cross-sectional area of earlywood vessels in Latvia during the period 1900–2009. For. Ecol. Manage. 279, 34–44.
Melvin, T.M., Briffa, K.R., 2008. A signal-free approach to dendroclimatic standardisation. Dendrochronologia 26, 71–86.
Melvin, T.M., Briffa, K.R., Nicolussi, K., Grabner, M., 2007. Time-varying response smoothing. Dendrochronologia 25, 65–69.
Melvin, T.M., Grudd, H., Briffa, K.R., 2013. Potential bias in 'updating' tree-ring chronologies using regional curve standardisation: re-processing 1500 years of Torneträsk density and ring-width data. Holocene 23, 364–373.
Mitchell, T.D., Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int. J. Climatol. 25, 693–712.
Naulier, Y., Savard, M.M., Bègin, C., Marion, J., Nicault, A., Bègin, Y., 2015. Temporal instability of isotopes–climate statistical relationships – a study of black spruce trees in northeastern Canada. Dendrochronologia 34, 33–42.
Neukom, R., Luterbacher, J., Villalba, R., Küttel, M., Frank, D., Jones, P.D., Grosjean, M., Wanner, H., Aravena, J.C., Black, D.E., Christie, D.A., D'Arrigo, R.D., Lara, A., Morales, M., Soliz-Gamboa, C., Srur, A., Urrutia, R., von Gunten, L., 2011. Multiproxy summer and winter surface air temperature field reconstructions for southern South America covering the past centuries. Clim. Dyn. 37, 35–51.
Neukom, R., Nash, D.J., Endfield, G.H., Grab, S.W., Grove, C.A., Kelso, C., Voegl, C.H., Zinke, J., 2013. Multi-proxy summer and winter precipitation reconstruction for southern Africa over the last 200 years. Clim. Dyn. 42, 2713–2726.
Neuwirth, E., 2014. Package 'RColorBrewer'. URL: https://cran.r-project.org/web/packages/RColorBrewer/index.html.
Nicolussi, K., Bortenschlager, S., Körner, C., 1995. Increase in tree-ring width in subalpine Pinus cembra from the central Alps that may be CO2-related. Trees 9, 181–189.
O'Donnell, A.J., Allen, K.J., Evans, R.M., Cook, E.R., Trouet, V., Baker, P.J., 2016. Wood density provides new opportunities for reconstructing past temperature variability from southern Australian trees. Global Planet. Change 141, 1–11.
Oberhuber, W., Kofler, W., Pfeifer, K., Seeber, A., Gruber, A., Wieser, G., 2008. Long-term changes in tree-ring-climate relationships at Mt Patscherkofel (Tyrol, Austria) since the mid-1980s. Trees 22, 31–40.
Orlowsky, B., 2015. Package 'iki.dataclim'. www.climate-babel.org.
PAGES2k Consortium, 2017. A global multiproxy database for temperature reconstructions of the Common Era. Sci. Data 4, 170088. http://dx.doi.org/10.1038/sdata.2017.88.
Pettitt, A.N., 1979. A non-parametric approach to the change point problem. J. Appl. Stat. 28, 126–135.
Prisarc, M.F., Carey, S.K., Kokelji, S.V., Youngblut, D., 2007. Anomalous 20th century tree growth, mackenzie delta, Northwest territories, Canada. Geophys. Res. Lett. 34. http://dx.doi.org/10.1029/2006GL02139.
Sànchez-Salguero Camarero, J.J., Guitièrrez Rouco, F.G., Gazol, A., Sangüesa-Barreda, G., Andreu-Hayles, L., Linares, J.C., Seftigen, K., 2017. Assessing forest vulnerability to climate warming using a process-based model of tree growth: bad prospects for rear-edges. Global Change Biol. 23, 2705–2719.
Salinger, M.J., Palmer, J.G., Jones, P.D., Briffa, K.R., 1994. Reconstruction of New Zealand climate indices back to AD 1731 using dendroclimatic techniques: some preliminary results. Int. J. Climatol. 14, 1135–1149.
Salzer, M.W., Hughes, M.K., Bunn, A.G., Kipfmueller, K.F., 2009. Recent unprecedented tree-ring growth in bristlecone pine at the highest elevations and possible causes. PNAS 106, 20348–20353.
Schneider, L., Esper, J., Timonen, M., Büntgen, U., 2014. Detection and evaluation of an early divergence problems in the northern Fennoscandian tree-ring data. Oikos 123, 559–566.
Tardif, J., Camarero, J.J., Ribas, M., Gutièrrez, E., 2003. Spatiotemporal variability in tree growth in the central Pyrenees: climatic and site influences. Ecol. Monogr. 73, 241–257.
Tolwinski-Ward, S.E., Evans, M.N., Hughes, M.K., Anchukaitis, K.J., 2011. An efficient forward model of climate controls on interannual variation in tree-ring width. Clim. Dyn. 36, 2419–2439.
Tumajer, J., Altman, J., Stepànek, T., Reml, V., Dolezal, J., Cienciala, E., 2017. Increasing moisture limitation of Norway spruce in Central Europe revealed by forward modeling of tree growth in tree-ring network. Agric. Forest Meteorol. 247, 56–64.
Vaganov, E., Hughes, M.K., Kirdyanov, A.V., Schweingruber, F.H., Silkin, P.P., 1999. Influence of snowfall and melt timing on tree growth in subarctic Eurasia. Nature 400, 149–151.
Vaganov, E., Hughes, M.K., Shahskin, A.V., 2006. Growth Dynamics of Conifer Tree Rings Images of Past and Future Environments. Springer, Berlin Heidelberg, New York (351 pp).
Van Deusen, P., 1990. Evaluating time-dependent tree ring and climate relationships. J.

Environ. Qual. 19, 481–488.
Villalba, R., Lara, A., Boninegna, J.A., Masiokas, M., Delgado, S., Aravena, J.C., Roig, F.A., Schmelter, A., Wolodarsky, A., Ripalta, A., 2003. Large-scale temperature changes across the southern Andes: 20th-Century variations in the context of the past 400 years. Clim. Change 177, 177–232.
Villalba, R., 1990. Climatic fluctuations in northern Patagonia during the last 1000 years as inferred from tree-ring records. Quat. Res. 34, 346–360.
Visser, H., Molenaar, J., 1988. Kalman filter analysis in dendroclimatology. Biometrics 44, 929–940.
Visser, H., 1986. Analysis of tree ring data using the Kalman filter technique. IAWA Bull. 7 (4), 289–297.
Von Neumann, J., 1941. Distribution of the ratio of the mean square successive difference to the variance. Ann. Math. Stat. 13, 367–395.
Wilmking, M., D'Arrigo, R.D., Jacoby, G.C., Juday, G.P., 2005. Increased temperature sensitivity and divergence growth trends in circumpolar boreal forests. Geophys. Res. Lett. 32, L15715. http://dx.doi.org/10.1029/2005GL023331.
Wilson, R., Luckman, B., 2002. Tree-ring reconstruction of maximum and minimum temperatures and the diurnal temperature range in British Columbia, Canada. Dendrochronologia 20, 1–12.
Wilson, R., D'Arrigo, R.D., Buckley, B., Büntgen, U., Esper, J., Frank, D., Luckman, B., Payette, S., Vose, R., Youngblut, D., 2007. A matter of divergence: tracking recent warming at hemispheric scale using tree ring data. J. Geophys. Res. 112 (D17103 1029/2006JD008318).
Wilson, R., Miles, D., Loader, N.J., Melvin, T., Cunningham, L., Coopr, R., Briffa, K., 2013. A millennial long March-July precipitation reconstruction for southern-central England. Clim. Dyn. 40, 997–1017.
Zang, C., Biondi, F., 2013. Dendroclimatic calibration in R: the bootRES package for response and correlation analysis. Dendrochronologia 31, 68–74.
Zang, C., Biondi, F., 2015. Treeclim: an R package for the numerical calibration of proxy-climate relationships. Ecography 38, 001–006.
Zhang, Y., Wilmking, M., Gou, X., 2009. Changing relationships between tree growth and climate in northwest China. Plant Ecol. 201, 39–50.