**Climate
of the Past**

Clustering climate reconstructions

G. Bürger

Pacific Climate Impacts Consortium, Victoria, Canada

Abstract. A systematic coherence analysis is presented for the set of the most prominent millennial reconstructions of northern hemispheric temperature. The large number of mutual coherences underwent a clustering analysis that revealed five significant, mutually incoherent ("inconsistent") clusters. The use of multiple proxies seems to be causing the clustering, at least in part, but not in an easily definable, physical way. Alternatively, a multidimensional scaling is performed on the same set of coherences. This results in a graphic, two-dimensional rendering of the reconstructions whose geometry (location and distance) is given by the coherences. Both approaches offer complementary ways in dealing with the inconsistencies.

1 Introduction

How inconsistent do two models have to be in order to dismiss at least one of them? – For example, if model M1 purports that at least $60\% \pm 5\%$ of all crows are green and model M2 purports $55\% \pm 5\%$ are red, then, using classic logical and arithmetical reasoning, the models are inconsistent and at least one model must be dismissed as being wrong. (Of course, both can be wrong.) But what happens if the uncertainty is slightly larger (10% instead of 5%)? And are these arguments still valid in times of "reasoning under uncertainty" (Shafer and Pearl, 1990; Parsons, 2001) and an emergence of "paraconsistent" logics (Priest, 2000, 2002; Arieli, 2008)? In the latter, for example, logical reasoning does not become useless as soon as an "inconsistency" occurs in a system, in the form of a proposition (A) together with its opposite ($\neg A$).

Such questions may arise when investigating the modeling – or reconstructing – of past millennial northern hemispheric (NH) temperature. They arose in me, at least, in an attempt to understand the reconstructions of the latest IPCC report (Jansen et al., 2007; Fig. 6.10), of which the following extend back to the year 1000: (Jones et al., 1998; Mann et al., 1999; Briffa, 2000; Esper et al., 2002; Mann and Jones, 2003; Moberg et al., 2005; d'Arrigo et al., 2006). The figure in that report displays an overlap of the 1σ and 2σ uncertainty bands of the reconstructions, weighted accordingly, that approximates the "most likely" temperature for any given year. In statistical terms, the figure entails what is also known as a "probability mixture model" (McLachlan and Peel, 2000), with all sub-models weighted equally. The reconstructions are thus tacitly considered as mutually consistent, and conflicting variations between any two of them are not resolved, but instead add to an overall uncertainty of a unique, albeit unknown NH temperature.

The present study investigates in more detail whether this consistency assumption is actually justified, by analyzing "consistency" very simply in terms of spectral coherence. I will not delve, however, into any logical implications of potential inconsistencies, or whether a paraconsistent framework is indicated in this case or not, but leave the semantic details of the notion to the reader. For the purpose of this study I have also included the reconstructions (Crowley and Lowery, 2000) and (Mann et al., 2008), making a total of ten reconstructions listed in Table 1. That all reconstructions are usually weighted equally, such as in the IPCC Figure, is mainly due to the lack of better evidence. Many verification attempts (Briffa et al., 1988; Mann et al., 1998, 2008; Rutherford et al., 2005; Wahl et al., 2006) suffer from insufficient independent verification data, which severely obscures the corresponding statistics (Bürger, 2007; Christiansen et al., 2009). This lack of data can partly be evaded by using the synthetic data of a climate simulation, where "pseudo" proxies, which are temperature grid points degraded by noise, are

Table 1. The ten reconstructions used in this study, with target season and proxy type.

Symbol	Reference	Season	Proxies
Br00	Briffa, 2000	Summer	trees
dA06	d'Arrigo et al., 2006	Annual	trees
Es02	Esper et al., 2002	Annual	trees
MJ03	Mann and Jones, 2003	Annual	multi-proxy
Mo05	Moberg et al., 2005	Annual	multi-proxy
CL00	Crowley and Lowery, 2000	Annual	multi-proxy
Jo98	Jones et al., 1998	Summer	multi-proxy
Ma08L	Mann et al., 2008	Annual	multi-proxy
Ma08	Mann et al., 2008	Annual	multi-proxy
Ma99	Mann et al., 1999	Annual	multi-proxy

used to track temperature (Von Storch et al., 2004; Mann et al., 2005; Lee et al., 2008; Christiansen et al., 2009). The variance of the noise, or the signal-to-noise ratio (SNR), is determined from local temperature-proxy correlations. According to these studies none of the tested methods revealed a performance conclusive enough to provide reliable temperature estimates for the entire millennium, at least not for the appropriate setting, that is, a small proxy network with a low SNR. It should be noted, moreover, that the reported performance measures are likely too optimistic anyway, as the local temperature-proxy error model – independent white noise – has been shown to be inadequate (Bürger et al., 2006; Bürger, 2007); see also (Von Storch et al., 2006).

With verification being thus poor, debates about competing approaches to climate reconstruction, such as regional curve standardization (Briffa et al., 1992; Esper et al., 2002), or different variants of regression (Von Storch et al., 2004; Mann et al., 2005) remain largely undecided.

If such "unilateral" validation approaches fail, bilateral analyses may offer some guidance to assess millennial climate reconstructions. I am aware of only one systematic analysis of such kind. (Juckes et al., 2007) calculate cross correlations of six reconstructions, four of which are also considered here. But their analysis has a number of caveats. For example, the corresponding low-pass filtered versions (21-year running mean) have been described as "highly correlated", but no significance analysis has been supplied that would put "high" into context. Moreover, their estimates are optimistically biased as they include the instrumental period which was used for calibration. Keep in mind, however, that bilateral methods provide necessary but probably insufficient validation criteria for climate reconstructions.

I follow a similar approach here by systematically analyzing the mutual consistency of the ten reconstructions of Table 1. To avoid the "synchronization" effect from the calibration, the analysis will be based exclusively on pre-instrumental variations. Additionally, spectral coherence is used as a consistency measure. Significance estimates of coherence rely on little more than very general stationarity as-

sumptions on the time series, so that this approach presents a better protection against, e. g., spurious significance in correlation measures (Granger and Newbold, 1974). This is further discussed in Sect. 3.

Using a distance measure that is based on coherence, aggregated across relevant frequencies, a clustering analysis is performed on the set of reconstructions. This results in a structured view of the reconstructions, with any two clusters being called "incoherent" or "inconsistent" if the coherence of at least two of their members cannot be established in a significant way. Based on the same distance metric, a multi-dimensional scaling (MDS; Hastie et al., 2001) of the ten reconstructions is performed. In two dimensions the technique, which is briefly described in the next section, produces a very graphic rendering of the reconstructions that may already be useful for, e.g., detecting outliers, and may help to design new reconstructions from the ones given. And it may ultimately lead to a better logical understanding, as indicated above, of what has actually been reconstructed.

2   Clustering reconstructions

The following coherence analysis establishes statistically whether corresponding covariations represent coherent behavior or just pure chance. The analysis is constrained to reconstructed data prior to 1850, to ensure that the estimated coherence is not inflated by calibrating effects from the instrumental period. All reconstructions are rescaled to have zero mean and unit variance.

We use the multitaper spectral estimator (Percival and Walden, 1993). Coherence, κ, as a spectral measure depends on frequency, f. An appropriate summary measure is given by the quantity

$$\overline{\kappa} = 1/0.2 \int_{0 \le f \le 0.2} \kappa(f) df, \tag{1}$$

representing the average coherence in the spectral band $0 \le f \le 0.2$, which means variability above 5 years. This is the time scale where significant temperature-proxy interaction is to be expected (Cook et al., 1998, 2000, 2004; Biondi et al., 2001; d'Arrigo et al., 2001; Briffa et al., 2002; Gray et al., 2003, 2004; Wilson et al., 2007). Table 2 shows the complete set of mutual coherences $\overline{\kappa}$ for the millennial reconstructions. Corresponding significance levels can be estimated analytically as follows. For a spectral smoothing filter of length m, the quantity

$$Y = 2m \frac{|\hat{\kappa}(f)|^2}{1 - |\hat{\kappa}(f)|^2} \tag{2}$$

Table 2. Mutual coherence κ between millennial (1000 to 1850) reconstructions, averaged over frequencies $\geq 5y$. Significant values ($\kappa \geq 0.47$ as the 99% level) are highlighted.

$\overline{\kappa}$	Br00	dA06	Es02	MJ03	Mo05	CL00	Jo98	Ma08L	Ma08	Ma99
Br00	**1**									
dA06	**0.65**	**1**								
Es02	**0.55**	**0.6**	**1**							
MJ03	0.4	0.38	0.44	**1**						
Mo95	0.46	0.41	0.42	**0.54**	**1**					
CL00	0.38	0.35	0.31	0.32	0.33	**1**				
Jo98	0.42	0.4	0.36	0.31	0.27	**0.54**	**1**			
Ma08L	0.3	0.28	0.22	0.35	0.31	0.28	0.25	**1**		
Ma08	0.42	0.36	0.36	**0.48**	0.42	0.32	0.27	**0.48**	**1**	
Ma99	0.37	0.34	0.38	0.38	0.31	0.34	0.44	0.26	0.3	**1**

follows an F-distribution with 2 and $4m$ degrees of freedom (Brockwell and Davis, 1991). Therefore, using a significance level of α the null hypothesis of zero (= random) coherence is rejected if

$$Y > F_\alpha(2, 4m) \qquad (3)$$

This criterion, which is obviously independent of frequency, is valid under very broad conditions, and holds for example if only the series are stationary (Brillinger, 2001). Consequently, it applies uniformly, whether the process under consideration has no memory, finite memory, or even infinite memory (so called long-range dependence). It is verified easily using Monte Carlo experiments, an example of which is given in the Interactive Discussion of the current paper. Note that the allowance for long-range dependence in null hypothesis testing crucially affects other significance estimates, e.g. of correlation coefficients, a fact that will be further discussed in the next section.

Based on Eq. (3), of all $\binom{10}{2} = 45$ pairs, only a small fraction (7) turns out to be significantly nonzero, indicating nonrandom behavior. Among these, the pairs dA06, Br00 and dA06, Es02 (abbreviations from Table 1) stick out with values of 0.65 and 0.6; and, adhering to what can be called the transitive law of coherence, Br00, Es02 follow with $\overline{\kappa} = 0.55$. It should be added that in these cases, the phase spectrum was always close to zero, indicating vanishing time shift as one would expect for any two significantly coherent reconstructions.

More systematically, a hierarchical clustering analysis (Hastie et al., 2001) is applied to Table 2, using as a distance metric the term

$$D = 1 - \overline{\kappa}. \qquad (4)$$

Starting from each single reconstruction as a cluster, new clusters may be formed recursively from any given set of clusters by merging the two nearest (most coherent) clusters, the distance of any two clusters being taken as the maximum
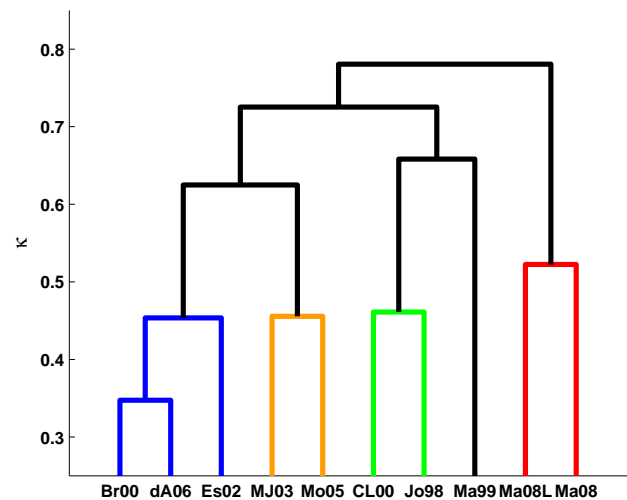
Fig. 1. Dendrogram of reconstructions, with distance metric D based on coherence κ (see text). Each node immediately below the 99% significance level of $D=0.525$ corresponds to a significant cluster, signified by the coloring.

of *all* member distances ("complete linkage"). If that distance is smaller than the distance corresponding to the level of κ significance in Eq. (4) (from the 99% level), the two clusters are called *coherent* and merged to form a new cluster. A looser criterion of forming clusters is "single linkage" where maximum distance is replaced by minimum distance. But in that case, two clusters are merged if only *any* two members are significantly coherent, and so clusters are populated with mutually incoherent members (reconstructions) which should be avoided after all. Therefore, single linkage clustering is generally dropped from this analysis. The clustering is shown in Fig. 1, the resulting group of internally coherent but mutually incoherent clusters signified by different colors. The height of a node is given by the distance of its
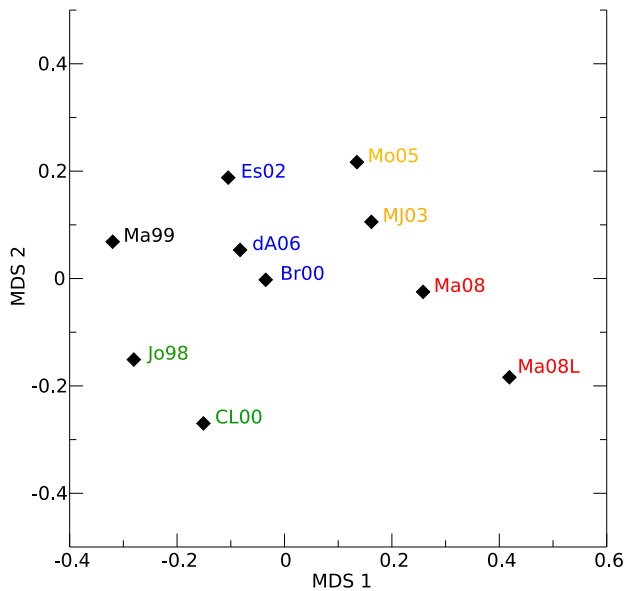
Fig. 2. MDS image of the ten climate reconstructions, based on mutual coherence. The five colors represent the five inconsistent clusters (complete linkage), {Br00, Es02, dA06} (blue), {MJ03, Mo05} (orange), {CL00, Jo98} (green), {Ma08L, Ma08} (red), and {Ma99} (black).

two constituents. Five clusters are so obtained: {Br00, Es02, dA06}, {MJ03, Mo05}, {CL00, Jo98}, {Ma08L, Ma08}, and {Ma99}.

A more graphic representation of the reconstruction clustering is obtained from performing a MDS analysis. In MDS, a matrix of dissimilarities, **D**, between a number of objects is mimicked as a distance matrix of an abstract set of points in some higher dimensional Euclidean space. This is achieved by embedding the objects into that space in a way that their distance matrix is as close as possible to the matrix **D**. Beyond this resemblance the mapped points, respectively their coordinates MDS1, MDS2,..., have no particular physical meaning. In this study, the ten reconstructions, with **D** being the matrix of Table 2, are embedded into a two-dimensional Euclidean space, the result of which is shown in Fig. 2. Br00 occupies the center of the plot, with relatively moderate (albeit mostly inconsistent) distances to the other reconstructions; dA06 is similar. In this display, Ma08L appears as the most "excentric" reconstruction, followed by CL00, Ma99, and Mo05. Ma08L and Ma99 show the greatest distance, that is, of all pairs they are maximally inconsistent. Note that all five clusters are well represented in the plot (which is not too surprising as this is exactly the purpose of MDS).

Figure 3 displays the reconstructed time series grouped by cluster. Cluster {Br00, Es02, dA06} shows warm conditions at about the years 1000, 1400 and 1550, and cooler conditions from 1200 to 1350 and at 1450 and 1600. Cluster {MJ03, Mo05} is, like all remaining clusters, dominated by

a fairly strong negative trend. On top of that there is an extended cooling in the 17th century, followed by much warmer conditions in the 18th century. Not much variability is in cluster {CL00, Jo98}, only the apparent negative trend which seems stronger for CL00. The series are weakly coherent. Finally, the clusters {Ma08L, Ma08} and {Ma99} are both characterized by little variability, interrupted by sporadic outbreaks of strong cooling (1350, 1450, 1700) that might be related to volcanic events.

To exemplify the inter- and intra-cluster coherence I have plotted in Fig. 4 typical coherence spectra from the clusters {Br00, Es02, dA06}, {MJ03, Mo05}, {Ma08L, Ma08}, and {Ma99}, together with the 90%, 95%, and 99% confidence band of no coherence (which is known to be independent of frequency). Br00 and dA06 are significantly (99%) coherent on all timescales, whereas MJ03 and Mo05 are coherent at the lower frequencies ($f \leq 0.2$) only. An extreme case of cross-cluster inconsistency are the two most distant reconstructions Ma99 and Ma08L, which are nowhere coherent except at very small frequencies, signifying their common negative trend.

A potential cause of the cluster incoherence may lie in the different target areas of the reconstructions. For example, Br00 reconstructs the NH extratropical land temperature only, while Mo05 is targeted at the entire NH. Inspecting Table 3 shows that in fact the five clusters are nicely lined up with their respective target configurations, with the exception of {Ma08, Ma08L} which are distinguished by using sea surface information. But this characterization is not unique, as, e.g., {MJ03, Mo05} and {Ma99} are incoherent but share the same targets. Moreover, the different targets are not very different in the first place, as Table 4 shows: a millennial climate simulation (Gonzalez-Rouco et al., 2003) shows that the various target areas are strongly coherent for the relevant time scales ≥ 5 y, with $\overline{\kappa} \sim 0.95$ or $D \sim 0.05$.

If the different target areas cannot sufficiently account for the different clusters, having a sufficiently even type and processing of proxies seems to lead to coherent reconstructions. This applies to the cluster {Br00, Es02, dA06}, all whose reconstructions are based on tree rings and a similar technique (age band decomposition and regional curve standardization) to retain low-frequency information for the proxy standardization.

It should be noted that choosing a narrower spectral band, such as for example $0 \leq f \leq 0.1$, that is, decadal and longer, does not alter the dendrogram of Fig. 1 substantially. Similarly, the 95% significance level ($D = 0.61$) yields identical clusters, while under the 90% level ($D = 0.65$) clusters {Br00, Es02, dA06} and {MJ03, Mo05} merge.
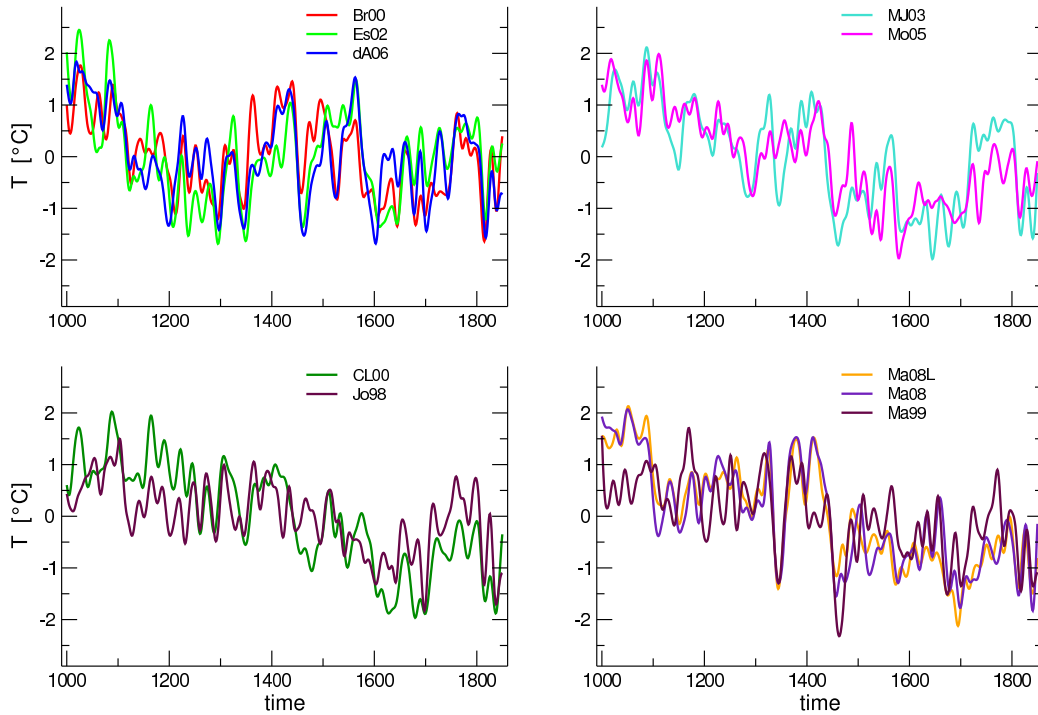
Fig. 3. The five clusters of reconstructions (smoothed). Note that Ma99 forms a single cluster.

Table 3. Target area for reconstructions.

	Br00	dA06	Es02	MJ03	Mo05	CL00	Jo98	Ma08L	Ma08	Ma99
NH				■	■			■	■	■
NH extratropics	■	■	■			■	■			
land + sea				■	■	■	■		■	■
land	■	■	■					■		

3 Coherence vs. correlation

The Interactive Discussion raised some concern about the preference for spectral coherence, κ, over correlation, ρ, as a consistency measure. My argument in favoring κ is its simplicity with respect to significance estimates. These estimates depend on the degrees of freedom, dof, used to calculate the statistic in question. Because spectral estimates are in most cases independent for different frequencies, this number dof is directly and simply related to the spectral smoothing used for the estimates (see Eqs. 2 and 3). For ρ, however, the determination of dof is often a matter of debate, such as, for example, in the controversy around the so called "hockey stick" (e.g. McIntyre and McKitrick, 2005).

Analogously to Table 2, I have calculated the matrix of correlations between the ten reconstructions, restricting the analysis to variations ≥ 5y, shown in Table 5. The overall magnitude of ρ is similar to the values reported in (Juckes et al., 2007), but single values differ quite considerably (e.g.

$\rho = 0.71$ there vs. $\rho = 0.36$ here, for the Es02, Ma99 correlation). This difference is caused solely by the restriction to pre-instrumental temperatures and/or the somewhat weaker smoothing used in this study.

To decide the significance question the correlations need to be assessed against a realistic, albeit "pessimistic" null hypothesis of them being based on pure chance, which brings up the problem of the $dofs$. Treating all 851 points of the time series as independent without any memory, so that each represents one degree of freedom, leads to a white noise null distribution which is quite unrealistic and too optimistic. Autoregression introduces more memory into a process, but still remains in the realm of "short-range" where memory fades after some time. "Long range" dependent processes (with a fractional differencing parameter $d > 0$) have instead a memory that never dies off completely, but require very long time series to obtain sufficiently robust estimates of d. Of several existing procedures I applied the local Whittle estimator
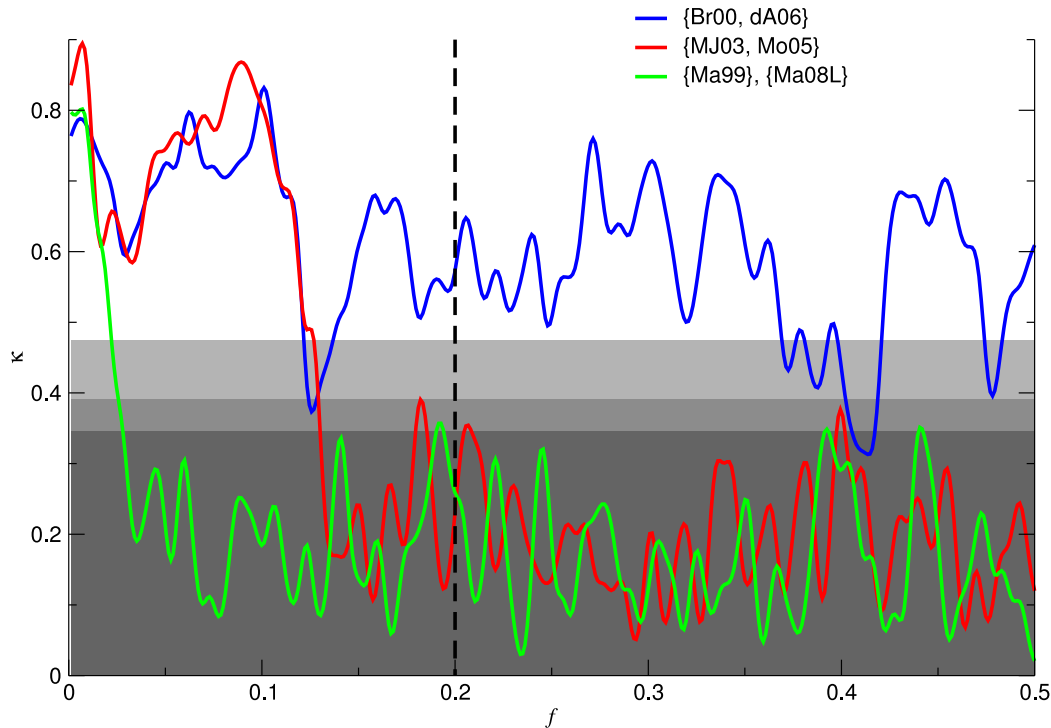
Fig. 4. Intra- and inter-cluster coherence spectrum (smoothed). The gray areas mark, from dark to light gray, the 90%, 95%, and 99% significance level. The vertical dashed line indicates the frequency threshold below which reconstructions are compared for clustering.

("LW": Shimotsu and Phillips, 2005) and the method of discrete variations in presence of outliers and/or an additive noise ("DV": Achard and Coeurjolly, 2010). Both methods revealed a very strong dependence on the choice of parameters, especially DV. This latter method, moreover, using the default setting offered for the ten reconstructions the entire range from white ($d = -0.4$ for Jo98) to brown noise ($d = 1.3$ for Ma08) as a null model, which I therefore dropped as being unrealistic. For the LW method I generated pairwise null distributions from the default parameters and obtained corresponding significance levels for ρ. Based on this, the number of insignificant correlations was much smaller than the corresponding number for coherence (9 vs. 38). It increased to 23, however, (and 39 for coherence) after removing the negative trend that is common to all reconstructions.

Comparing the intricacy and uncertainty of this estimate with the analytic estimate of Eq. (3), and because of the strong impact of the millennial trend on correlation, I put more confidence into the coherence estimates.

Table 4. Coherence between target areas as simulated by ECHO-G Erik.

$\overline{\kappa}$	NH	NH extratropics
land + sea	0.93	0.89
land	0.90	0.91

4 Discussion

By avoiding the (calibrating) instrumental period, and by using a fairly robust spectral measure for low-frequency performance, the above coherence analysis has uncovered several inconsistencies among the group of millennial reconstructions that figured prominently in the latest IPCC report and elsewhere. An immediate lesson from this is that simple visual inspection of smoothed time series, grouped and overlaid into a single graph, can be very misleading. For example, the two reconstructions Ma99 and Ma08L, which have previously been described to be in "striking agreement" (cf. Mann et al., 2008), turned out to be the most incoherent of all in our analysis.

Table 5. Mutual correlation ρ between millennial (1000 to 1850) reconstructions, using variations ≥ 5 y. Bold numbers indicate significance (99%) for the detrended series.

	Br00	dA06	Es02	MJ03	Mo05	CL00	Jo98	Ma08L	Ma08	Ma99
Br00	**1**									
dA06	**0.74**	**1**								
Es02	**0.64**	0.74	**1**							
MJ03	**0.60**	**0.62**	0.52	**1**						
Mo95	**0.61**	**0.51**	0.36	**0.80**	**1**					
CL00	**0.54**	0.43	0.36	0.78	**0.75**	**1**				
Jo98	**0.53**	0.44	0.31	0.62	0.57	**0.69**	**1**			
Ma08L	0.53	**0.49**	0.28	**0.74**	**0.74**	0.79	0.57	**1**		
Ma08	**0.56**	0.55	**0.32**	**0.71**	**0.69**	0.72	0.51	**0.92**	**1**	
Ma99	**0.42**	0.54	0.40	**0.68**	**0.56**	0.68	0.58	0.60	0.60	**1**

The most obvious, pragmatic, response to the inconsistencies is to inspect the methods and try to improve and harmonize them. But as I have pointed out, without a functioning, uncontroversial verification procedure this will not lead very far.

Having therefore to live, for now, with pairwise inconsistent reconstruction clusters there is more than one way to interpret the coherence results meaningfully. Two complementary views regarding the "true" NH temperature are possible, depending on the focus lying on the clustering or on the MDS:

(a) five inconsistent clusters each representing a possible truth

(b) ten independent approximations of an otherwise unknown truth

ad (a) With no obvious means at hand to dismiss any of the five inconsistent reconstructions, one would have to deal with derivations involving inconsistent statements. As mentioned in the beginning, this requires a non-standard approach to the logical discourse, perhaps along the lines of, e.g., (Arieli, 2008).

ad (b) This viewpoint, which may be somewhat more realistic than a), is closer to the conventional approach where all reconstructions are seen as approximations to a single, true temperature curve. However, the error metric is fundamentally different here. The conventional metric would operate on the reconstructed temperatures themselves and construct a real temperature average. The view suggested here (mainly through Fig. 2) is that the best estimate of truth is near the "center" of the reconstructions in the MDS rendition. But this rendition is non-physical, or not directly physical, as the MDS dimensions are not related to the original temperature series in a simple way. Least-squares approaches do not work here, so that estimating the center by simple temperature averaging is impossible. That center represents a compromise of the reconstructions, in the sense that it would be, on average, maximally coherent with all of them. It is likely to be "close" to Br00 and Ma08 and may be found by prudently merging techniques and proxies from both approaches. Otherwise, one would probably have to resort to trial and error.

Favoring correlation over coherence raises serious questions about the significance levels, and any corresponding null hypotheses will be a matter of debate. Moreover, without the common negative millennial trend considerable inconsistencies remain.

One may as well choose to neglect the inconsistencies altogether. But then the following, and likely more, semantic subtleties regarding the reconstructions have to be resolved:

– Can they *skillfully* represent NH temperatures?

– Can they lie within a *common uncertainty* bound?

– If they suggest an identical *conclusion* – such as the non-existence of a Medieval Warm Period, what does it mean for that conclusion?

Using inconsistent reconstructions to approximate the temperature curve has one particular visual consequence. Whether overlaying them in one figure or forming an average, the result tends to be a cancellation of larger amplitudes, because inconsistency here means to be indistinguishable from random covariations. Together with the mentioned synchronization through the instrumental calibration period, such "synthesis" figures automatically resemble a hockeystick.

It was shown that the target area plays only a minor role. Furthermore, if type and processing of proxies are sufficiently even, coherent reconstructions are produced. If that is true in general, the main source of reconstruction inconsistency is the use of mixed types of proxies ("multiproxies"), and their role for temperature reconstruction should be revised. One should systematically check whether "uni"-proxy reconstructions tend to be more coherent than multi-proxy reconstructions, and if so, which types of proxies actually create the inconsistencies.

Edited by: J. Guiot

References

Achard, S. and Coeurjolly, J. F.: Discrete variations of the fractional Brownian motion in the presence of outliers and an additive noise, Statistics Surveys, 4, 117–147, doi:10.1214/09-SS059, 2010.

Arieli, O.: Distance-based paraconsistent logics, Int. J. Approx. Reason, 48(3), 766–783, 2008.

Biondi, F., Gershunov, A., and Cayan, D. R.: North Pacific decadal climate variability since 1661, J. Climate, 14, 5–10, 2001.

Briffa, K.: Annual climate variability in the Holocene: interpreting the message of ancient trees, Quaternary Sci. Rev., 19(1–5), 87–105, 2000.

Briffa, K. R., Osborn, T. J., Schweingruber, F. H., Jones, P. D., Shiyatov, S. G., and Vaganov, E. A.: Tree-ring width and density data around the Northern Hemisphere: Part 1, local and regional climate signals, The Holocene, 12(6), 737–757, 2002.

Briffa, K. R., Jones, P. D., Bartholin, T. S., Eckstein, D., Schweingruber, F. H., Karlen, W., Zetterberg, P., and Eronen, M.: Fennoscandian summers from AD 500: temperature changes on short and long timescales, Clim. Dynam., 7(3), 111–119, 1992.

Briffa, K. R., Jones, P. D., Pilcher, J. R., and Hughes, M. K.: Reconstructing summer temperatures in northern Fennoscandinavia back to AD 1700 using tree-ring data from Scots pine, Arctic Alpine Res., 20(4), 385–394, 1988.

Brillinger, D. R.: Time series: data analysis and theory, Society for Industrial and Applied Mathematics, 2001.

Brockwell, P. J. and Davis, R. A.: Time series: Theory and Methods, Springer, New York, USA, 577 pp., 1991

Bürger, G.: On the verification of climate reconstructions, Clim. Past, 3, 397–409, doi:10.5194/cp-3-397-2007, 2007.

Bürger, G., Fast, I., and Cubasch, U.: Climate reconstruction by regression-32 variations on a theme, Tellus A, 58(1), 227–235, 2006.

Christiansen, B., Schmith, T., and Thejll, P.: A Surrogate Ensemble Study of Climate Reconstruction Methods: Stochasticity and Robustness, J. Climate, 22(4), 951, doi:10.1175/2008JCLI2301.1, 2009.

Cook, E. R., D'Arrigo, R. D., and Briffa, K. R.: A reconstruction of the North Atlantic Oscillation using tree-ring chronologies from North America and Europe, The Holocene, 8(1), doi:10.1191/095968398677793725, 1998.

Cook, E. R., Esper, J., and D'Arrigo, R. D.: Extra-tropical Northern Hemisphere land temperature variability over the past 1000 years, Quaternary Sci. Rev., 23(20–22), 2063–2074, 2004.

Cook, E. R., Buckley, B. M., D'Arrigo, R. D., and Peterson, M. J.: Warm-season temperatures since 1600 BC reconstructed from Tasmanian tree rings and their relationship to large-scale sea surface temperature anomalies, Clim. Dynam., 16(2), 79–91, 2000.

Crowley, T. J. and Lowery, T. S.: How Warm Was the Medieval Warm Period?, AMBIO, 29(1), 51–54, 2000.

d'Arrigo, R., Villalba, R., and Wiles, G.: Tree-ring estimates of Pacific decadal climate variability, Clim. Dynam., 18(3), 219–224, 2001.

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

Esper, J., Cook, E. R., and Schweingruber, F. H.: Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability, Science, 295(5563), 2250,

doi:10.1126/science.1066208, 2002.

Gonzalez-Rouco, F., Von Storch, H., and Zorita, E.: Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years, Geophys. Res. Lett, 30(21), 2116, doi:10.1029/2003GL018264, 2003.

Granger, C. W. J. and Newbold, P.: Spurious regressions in econometrics, J. Econometrics, 2(2), 111–120, doi:10.1016/0304-4076(74)90034-7, 1974.

Gray, S. T., Betancourt, J. L., Fastie, C. L., and Jackson, S. T.: Patterns and sources of multidecadal oscillations in drought-sensitive tree-ring records from the central and southern Rocky Mountains, Geophys. Res. Lett., 30(6), 1316, doi:10.1029/2002GL016154, 2003.

Gray, S. T., Graumlich, L. J., Betancourt, J. L., and Pederson, G. T.: A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation since 1567 AD, Geophys. Res. Lett., 31(12), L12205, doi:10.1029/2004GL019932, 2004.

Hastie, T., Tibshirani, R., and Friedman, J. H.: The elements of statistical learning: data mining, inference, and prediction: with 200 full-color illustrations, Springer, New York, USA, 533 pp., 2001.

Jansen, E., Overpeck, J., Briffa, K. R., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W. R., Rahmstorf, S., Ramesh, R., Raynaud, D., Rind, D., Solomina, O., Villalba, R. and Zhang, D.: Palaeoclimate, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. and Miller, H. L., Cambridge University Press, 433–497, 2007.

Jones, P. D., Briffa, K. R., Barnett, T. P., and Tett, S. F. B.: High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures, The Holocene, 8(4), 455–471, 1998.

Juckes, M. N., Allen, M. R., Briffa, K. R., Esper, J., Hegerl, G. C., Moberg, A., Osborn, T. J., and Weber, S. L.: Millennial temperature reconstruction intercomparison and evaluation, Clim. Past, 3, 591–609, doi:10.5194/cp-3-591-2007, 2007.

Lee, T., Zwiers, F., and Tsao, M.: Evaluation of proxy-based millennial reconstruction methods, Clim. Dynam., 31(2–3), 263–281, 2008.

Mann, M. E., Rutherford, S., Wahl, E., and Ammann, C.: Testing the fidelity of methods used in proxy-based reconstructions of past climate, J. Climate, 18(20), 4097–4107, 2005.

Mann, M. E., Bradley, R. S., and Hughes, M. K.: Global-scale temperature patterns and climate forcing over the past six centuries, Nature, 392, 779–787, 1998.

Mann, M. E., Bradley, R. S., and Hughes, M. K.: Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, Geophys. Res. Lett., 26(6), 759–762, 1999.

Mann, M. E. and Jones, P. D.: Global surface temperatures over the past two millennia, Geophys. Res. Lett., 30(15), 1820, doi:10.1029/2003GL017814, 2003.

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over

the past two millennia, P. Natl. A. Sci., 105(36), 13252–13257, 2008.

McIntyre, S. and McKitrick, R.: Hockey sticks, principal components, and spurious significance, Geophys. Res. Lett, 32(3), L03710, doi:10.1029/2004GL021750, 2005.

McLachlan, G. and Peel, D.: Finite Mixture Models, John Wiley & Sons, 419 pp., 2000.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlén, W.: Highly variable Northern Hemisphere temperatures reconstructed from low-and high-resolution proxy data, Nature, 433(10), 613–617, 2005.

Parsons, S.: Qualitative methods for reasoning under uncertainty, The MIT Press, 2001.

Percival, D. B. and Walden, A. T.: Spectral analysis for physical applications: multitaper and conventional univariate techniques, Cambridge Univ Pr., 583 pp., 1993.

Priest, G.: Truth and contradiction, The Philosophical Quarterly, 50(200), 305–319, 2000.

Priest, G.: Paraconsistent logic, Handbook of philosophical logic, 6, 259–358, 2002.

Rutherford, S., Mann, M. E., Osborn, T. J., Bradley, R. S., Briffa, K. R., Hughes, M. K., and Jones, P. D.: Proxy-Based Northern Hemisphere Surface Temperature Reconstructions: Sensitivity to Method, Predictor Network, Target Season, and Target Domain, J. Climate, 18(13), 2308–2329, 2005.

Shafer, G. and Pearl, J.: Readings in uncertain reasoning, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1990.

Shimotsu, K. and Phillips, P.C.B.: Exact local Whittle estimation of fractional integration, Ann. Stat., 33, 1890–1933, 2005.

Von Storch, H., Zorita, E., Jones, J. M., Gonzalez-Rouco, F., and Tett, S. F.: Response to Comment on "Reconstructing Past Climate from Noisy Data", Science, 312, 529c, 2006.

Von Storch, H., Zorita, E., Jones, J. M., Dimitriev, Y., Gonzalez-Rouco, F., and Tett, S. F.: Reconstructing past climate from noisy data, Science, 306, 679–682, 2004.

Wahl, E. R., Ritson, D. M., and Ammann, C. M.: Comment on "Reconstructing Past Climate from Noisy Data", Science, 312, 529b, 2006.

Wilson, R., Wiles, G., D'Arrigo, R., and Zweck, C.: Cycles and shifts: 1,300 years of multi-decadal temperature variability in the Gulf of Alaska, Clim. Dynam., 28(4), 425–440, 2007.