Clim. Past Discuss., 9, C2800–C2806, 2013
www.clim-past-discuss.net/9/C2800/2013/

Climate of the Past Discussions

Interactive comment on "Testing long-term summer temperature reconstruction based on maximum density chronologies obtained by reanalysis of tree-ring datasets from northernmost Sweden and Finland" *by* V. V. Matskovsky and S. Helama

T.M. Melvin (Referee)

t.m.melvin@uea.ac.uk

Received and published: 4 December 2013

Testing long-term summer temperature reconstruction based on maximum density chronologies obtained by reanalysis of tree-ring datasets from northernmost Sweden and Finland - V. V. Matskovsky and S. Helama

The authors reprocess two MXD data sets derived from northern Fennoscandia which have been used in recently published climate reconstructions. They find distinct low-

frequency differences between the two chronologies. They examine the effect of various standardisation options and show that differences in the standardisation methods used do not produce substantially different chronologies and so are not the cause of the differences observed. Generally this part of the work is good, addresses an important aspect of climate reconstruction, and advances the science of dendroclimatology.

Having found that the problem is not with tree-ring standardisation they do not manage to attribute another cause. They combine the two data sets and present a new reconstruction which, because of the unresolved problems that exist with these data, does not add to the current understanding of climate variability in Fennoscandia. The paper could be published as a test of standardisation (without the reconstruction). Alternatively the authors need more work to resolve the "problem" if they wish to produce a reconstruction.

Detailed Comments Abstract – contains too many terms like RC1SFC, RC2SF and should be simplified so it can be read by a general audience with the finer details in the discussion and conclusions.

P5662 L3 – whole paragraph – in a paper that is attempting a detailed examination of the problems it is inappropriate to say "there are troubling inconsistencies with temperature amplitudes and timing of events" without explaining which problems have been overcome and which are outstanding.

For Tornetrask there were inconsistencies for which the causes have been identified and corrected and Tornetrask reconstructions now provide roughly consistent long-timescale signals from both MXD and TRW. (Briffa (1992) found chronology values for MXD lower than those for TRW post1700 – the problem was corrected (Briffa 2011) by improvements to the use of RCS. Grudd (2008) found chronology values for MXD lower than those for TRW in the last 2 centuries - the problem was corrected see P5665, lines 24 to 29. Esper et al. demonstrated that the Finnish MXD and TRW chronologies are inconsistent but did not investigate the cause of the discrepancy.

L8 – Briffa 1992 reconstruction was >1000 years.

L14 – “revealed a previously undiscovered millennial scale cooling trend” – as Esper et al.’s findings are not corroborated by any other tree-ring analyses (both MXD and TRW) this statement is misleading.

L18 – Rather than “suggesting an overestimation of the Medieval warmth published by Grudd (2008)” – Melvin et al. clearly showed that the Grudd study contained systematic bias – and should not be considered as suitable to use as a climate reconstruction.

L20 – Having drawn attention to the discrepancies among several reconstructions the authors then state the “.. unsatisfactory, if not worrisome ..” nature of this situation but then cite Briffa et al. 1992 without acknowledging that this reconstruction is not demonstrably more in error (if indeed is is at all) than any of the earlier cited papers (note that the ad hoc correction applied to the MXD in this paper was substantially found to be reasonable see Briffa 2011).

P5663 L1 “Indeed, the flip side of this...” is unsuitable – I suggest “A problem associated with the ...” .

L14 “Yet, these biases do not fully explain the obtained differences between the Torneträsk (Melvin et al., 2013) and larger Fennoscandian MXD data (Esper et al., 2012).”. This statement is not justified. Melvin et al. investigated and corrected for the bias caused by offset mean values of MXD measurements from different contexts while Esper et al. (2012) did not test for this bias. Despite Esper et al.’s claim that “We carried out a number of tests to the MXD network and noted the robustness of the long-term trends, ..” the tests they performed using correlation Table S2 (removes the mean value of sample being compared), using curve-fitting methods Table S2 (removes mean of each series) and creating separate RCS chronologies for each group Figure S4 (sets the mean values of indices of each group to 1.0) did not test the robustness of the long term trend which is obtained from the varying mean value of tree index series over time.

L16 – “all the discussed studies used the same type of standardization method” – they used RCS but selected different options which might explain differences between the resulting chronologies.

P5664 L4 “microdensitometry” – the authors have not examined the possibility of systematic bias in MXD measurements.

L6 – “Tornedalen” - neither Esper et al. nor Melvin et al. use the pre-1860 data because they have not been adjusted for the pre-Stevenson screen period (see Frank 2007 QSR).

P5667 L8 “In the original RCS technique presented by Briffa et al. (1992), the standardization curve is not fitted individually to each data of tree-ring series.” - misleading, perhaps you meant “In the original RCS technique presented by Briffa et al. (1992), the same standardization curve is used to detrend each series of tree-ring measurements.”

P5668 L22 – Updated reference - Melvin and Briffa 2013 addresses signal-free RCS (2008 is SF curve fitting) see http://www.cru.uea.ac.uk/cru/papers/melvin2013dendrochronologia/.

P5669 L21 – why only the “first 100 years” of each tree. Using all years reduces the risk of inconsistencies due to suppressed early or late growth (e.g. the crossing of RCS curves of Fig 3a).

L15-18 (and P5672 L12) Replication differences – The TORN data are based on “mean tree” series where the FENN data have multiple cores for each tree (which artificially inflates EPS values and reduces error margins).

L25 – combining TORN and FENN MXD to produce the FULL dataset requires careful checking for site or sample type inhomogenuity (specifically offsets of mean values of the data series).

P5672 L24 – “eliminate the biases arising from temporal distribution of well and poorly growing trees” – wrong. It will “reduce the ‘modern sample bias’ created by sampling

living trees – which is much larger for TRW than MXD". One idea behind using two RCS curves is that the independent sub-samples can be shown to produce the same common signal – demonstrated clearly for the separate TORN and FENN signals shown in Fig 3b.

Table S2, S3 and S4 – two digits is sufficient (3 is OK) for correlation purposes – any more than this should be deleted. Also your discussion of the values should ignore differences that are likely to insignificant.

P 5674 L4 Replace "an increasing toward" with "a trend increasing towards".

P5675 L4-28 There seems to be little justification for joining measurement series and producing a combined chronology when there are such large unexplained differences outside of the calibration period e.g. the early medieval warmth (8th and 9th centuries) is not consistent.

P5677 L6 – "it may be generally comfortable to rely on massive sample replication in dendrochronology (Büntgen et al., 2012)" – not a useful statement as increasing sample replication does not remove the systematic end effects, caused by tree aging processes, which detrending is designed to remove.

P5678 L25 – "Tree-ring standardization is generally understood as an obstacle for deriving the low frequency climate information from tree-ring ..." – wrong. "Tree-ring standardization is necessary for the isolation of low frequency climate information from tree-ring ...".

P5679 L23 "As a consequence, tree-rings have even become notoriously poor indicators of low-frequency climate variability for wider readership (e.g. Broecker, 2001)." – only for those not familiar with background and techniques. "Tree ring chronologies are one of the few proxies for which sensible estimates of their skill at indicating 'long-timescale' variance can be calculated.

L28 - "This is how the correction procedure significantly reduces the "sample error"

due to uneven age distribution over the AD 1950–1990 period (see Fig. 7)." – the arguments here are not convincing – "is the error associated with an age distribution problem", "was the reduction in error significant", "why not RC2SFC" and "does Fig 7 show errors". It seems likely that a much larger data set is needed to resolve these questions and this statement needs to be qualified with words such as might or likely.

P5680 L7-17 Multiple RCS tends to reduce (two-curve RCS roughly halves the amplitude of the variance derived solely from the mean values of series of tree indices) longer timescale variance and improves the shorter timescale variance. This would reduce the effect of any site or sample type variation of the mean value of MXD measurements provided the differences are equally distributed over time. Correlation over the most recent century is not likely to be a useful test of the effect of multiple v single RCS. Two-curve RCS has less error and so reduced need for correction. TORN has ample trees for two-curve RCS – an examination of uncertainty error #1, of supplementary 4 is likely to show this.

L19-28 – can be tested by plotting the mean values of young/old trees or young/old rings on the same graph and looking for systematic differences (not necessary for this paper).

P5685 L3 "Thus our new reconstruction can be used as the source of information about year-to-year, as well as centennial and longer variations of summer temperature in Northern Fennoscandia for the Common Era." – need a warning about the unexplained discrepancy.

L5 "Nevertheless, the use of other proxies that can reproduce low-frequency past temperature variations is highly preferable in every paleoclimatic study." – this statement needs a qualification about the error estimates – if there is a proxy that can reproduce low-frequency accurately they should mention it if there is not then this statement should be removed/qualified appropriately.

Supplementary 1 Sorry cannot read Russian so this is my only description of the cor-

rection procedure. Your definition needs to clearly distinguish between SF-series, SF-curves by saying what has been removed e.g. SF-measurement series contain aging curve and noise but climate signal removed or SF-indices where aging curve and climate both removed consist of noise.

It seems that the correction is producing an improvement by reducing the low-frequency noise (an informed version of robust mean). Because two-curve RCS removes half the variance of the long-timescale signal the correction will have less effect for multiple RCS.

Tom Melvin 4/12/2013

Interactive comment on Clim. Past Discuss., 9, 5659, 2013.