

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

J. Björklund (Referee)

jesper.bjorklund@gvc.gu.se

Received and published: 29 January 2014

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 reanalyze the two longest existing MXD records from Fennoscandia. They assume that they should have similar signals in all frequencies due to that they have

C3280

similar forcing agent (summer temperature). In addition to finding that the two records covary most of the time they also find that they diverge significantly from time to time. They address the problem by applying different standardization-techniques but fail to conclude that that is the root of the problem. However, it is very interesting to note how different techniques perform together, but this section could be expanded to include other techniques that are mentioned in the text. Further this analysis would greatly benefit from addition of more chronologies, for instance TRW from the same dataset and MXD and TRW from other datasets, they do not have to be 1500 years. This would add credibility to the otherwise miniscule differences in correlations, with regard to standardization configuration skill. Further it is a little bit confusing to compare how high the correlation is between two different standardization techniques, I do not find that information useful. The attempt of making a new reconstruction of evidently diverging chronologies is advised against since one of the chronologies could hypothetically be more in error than the other. Additional chronologies might give answers to this. The paper should after proper revisions be published, but would benefit a great deal if some of the discussion-points about potential sources were analytically addressed instead of only discussed, see detailed comments.

Detailed comments P5661 L21 Consider adding Grudd (2008) and McCarroll et al (2013)

P5661 L27 Change “in the most recent” to “In a recent” or specify that you mean MXD from torneträskmaterial

P5662 L4 change “indications” to “records”

P5662 L6-L10 Grudd 2008 has on several occasions been shown to be incorrect (e.g. Melvin et al. 2013, Björklund et al. 2013) and should not be used in this comparison. Again P16-P18.

P5662 L2-L24 This paragraph would benefit from studying McCarroll et al. (2013). The discussion, as it is now, only relates to two different datasets treated in different ways.

C3281

P5662 L27-L29 Perhaps also sampling biases? Some trees may be preserved better than others. The preservation conditions are likely different for a dry dead-tree chronology than from a wet subfossil chronology. Also, the living tree material was sampled with different strategies for FENN and TORN.

P5663 L14-L16 Agree with Referee T. M

P5663 L22-L26 Since the Matskovsky 2011 paper is in Russian and most of the scientific contribution of this paper is to apply this method, perhaps it deserves a thorough description in the main article.

P5663 L26-L29 McCarrol et al. 2013 uses the same RCS method to standardize the MXD chronologies included in that paper. Forfjorddalen Torneträsk and Laanila.

P5664 L3-L5 How did you deal with effects from microdensitometry? You are addressing effects of standardization methods in this paper?

P5664 L6 I do not see a problem using Tornedalen temperatures as long as the analysis is made with high-pass filtered data. But low-frequency correlations, where pre-stevenson data is included, should be avoided.

P5664 L7-8 I agree with T.M that making a reconstruction is premature when the problems identified between the chronologies are not resolved. Instead I suggest to also include TRW and perhaps also more chronologies, does not have to be >1000 years, to have a larger sample when comparing skill of methods.

P5664 Materials and methods. Previously the authors described that problems do exist between the TORN and FENN. Then they are assuming that signals must be the same or very similar because of the homogenous nature of the dominating forcing agent (summertemperature) upon them. They go on to try to solve these problems with various standardization-techniques, and fail. Could the assumption perhaps be in error considering the heterogeneity of the datasets, it is clear that the elevation, the size, preservation medium, number of subsites, geographical distribution of subsites of the

C3282

sample-sites vary considerably. The amount of low-frequency variation retrieved in an RCS chronology would arguably be larger in a more homogenous sample. Would there be any way to investigate this possibility by separating the subsites from each other?

P5667 L23 Why ratios when convention says residuals for MXD data?

P5668 Correction procedure. This is a central part of the paper and should probably be detailed more extensively in the manuscript, maybe even with figures. This because large part of the readership is not Russian-speaking and because it seems as this methodology performs the best in general and thus is the most valuable finding of the paper. Further there are several other methodologies that are left out from comparison that the author mentions, I particularly think of Nicault et al. (2010).

P5670 Design of experiments. The main analytic tool used to evaluate standardization performance is pearson correlation. This is done with untreated chronologies as well as for smoothed chronologies. I would like to see p-values or $p = 0.05$ significance levels for all correlations, especially for the smoothed data, because they are associated with substantial loss of degrees of freedom. In the discussion where, one procedure is deemed to perform better than another, it would also be nice to analyse if the performance is significantly higher/lower than the other. Without this, one procedure performance over another could be purely by chance, because one procedure consequently must yield higher correlations than another if chronologies are slightly changed.

P5670 Reconstruction. See comment above.

P5673 L8-L11 See comment above about correlations. For example what is the p-value for the 300 yr smoothing where $r=0.44$, when adjusted for loss of degrees of freedom, see below. Dawdy, D.R., and Matalas, N.C., 1964, Statistical and probability analysis of hydrologic data, part III: Analysis of variance, covariance and time series, in Ven Te Chow, ed., Handbook of applied hydrology, a compendium of water-resources technology: New York, McGraw-Hill Book Company, p. 8.68-8.90

C3283

P5674 L13 I do not see the need for the quasi-periodicity analyse. It is not associated with the comparison of standardization techniques. But it could perhaps be used early in the paper to establish the assumption that the two dataset are very similar or that they differ in some frequencies.

P5676 Comparison of datasets. The authors discuss potential sources of the differences between TORN and FENN. This should be more the motivation for the paper, and to construct experimental design to try to investigate these, because it is obviously not only standardization that can improve the coherence between the two.

P5694 F3 Chronologies produced with different standardization-techniques. To me it looks like the different standardization techniques have larger effect on the FENN material and that TORN seem inert to choice. What can be learned from that? Also the change caused by standardization techniques is marginal to the differences between the datasets. Clearly this must be further addressed before a reconstruction can be made, see comment above. Alternatively, more chronologies must be added to arrive in higher certainty around the mean. It can hypothetically be the case that one of them is more correct than the other, and then a new composite reconstruction would be worse than then previous publications of them separately.

P5696 F5. Why are spectrums for differently standardized chronologies shown?

P5697 F6 Figure must be redesigned, now it is impossible to see legends and differences between chronologies. Why not use residuals between summer temperatures and chronology-configuration? A trend analysis of the residuals can be performed and see if some technique produce less or more trend?

P5698 F7 Tornedalen record seems odd in the beginning and in the context of the discussion with Stevenson-screens, should perhaps be omitted from analysis?

Jesper Björklund 29/01/2014

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

C3284