Interactive comment **- Anonymous Referee #2 -** Received and published: 11 May 2017
our responses in BLUE

I apologize for "over-reading" the 100-y length condition and comments regarding this topic. Sorry to hear that people are still so uncooperative regarding sharing data that have been published >5-10 years ago. This sure is a problem for advancing the science and has been recognized (or better finally publicly "criticized") recently in Babst et al. 2017 (Improved tree-ring archives will support earth-system science. NEE).

With our effort of setting up a cooperative and open access dendro group in Italy, we will give the opportunity to freely access our data in the online supplementary materials.

Regarding RCS: Yes, for retaining low-frequency it's superior - given that your dataset actually allows a robust regional curve - but prone to a lot of biases. I am not really concerned about the MXD data, because the slope in MXD is usually pretty flat, so you won't run in to big troubles there.

However, I would be still very interested to see to see the Italy-only MXD chronology detrended with a 150-year spline (if I remember correctly) for a direct comparison of the different oscillations against the Trouet reconstruction. I would consider the RCS application as a second and final step to investigate how much more low-frequency there actually is (or might be).

We could run some tests indexing the MXD series with 150y splines as suggested by Referee2, construct the new HSTC and directly compare its oscillations with our reconstructions and with the Trouet's 2014. The results will be showed in the response to Referee2 and if are interesting, they will be put also in the online supplementary materials, not in the paper. In fact, the paper already presents several different approaches (for climate sensitivity analysis, HSTC construction and two approaches to temperature reconstruction - regression and scaling), we think that this additional analysis would overload the reader.

Moreover, in our opinion, the RCS approach is better performing than splines or the classic negative exponentials (previous Refereee2's comments) for preserving low-frequency signals, especially with TRW that present larger widths in the 'young' period of the individual series. By using the same standardization approach both for MXD and TRW data we avoid the possible introduction of different frequency responses that would then impede all the comparisons between MXD and TRW in the analyses of 'Climate sensitivity' and 'Climate sensitivity through time'.

Additionally, I am still extremely cautious of the application of RCS on TRW at sites with n<10-15. If you use a "10% spline" (which in your case comes close to 15-20 years with the "younger" broadleaf samples) to build the RC, the RC is potentially very noisy (or wiggly). And if your 3-9 samples have a narrow age range you essentially take out most of the low frequencies you intended to retain and your

RC at higher ages is probably more flexible at higher ages (due to only very few samples) than the stiff tail of a negative exponential curve.

At each site the individual series passed a rather robust selection: one of the fixed criteria was “ii) the individual series correlation with the respective site chronology had r > 0.3” (p. 6 l. 14; Table 2). This criterion, together with the minimum age of 100 yr length for each series, let us be rather sure that the resulting Regional Curve used for indexing the raw series is also representative of the growth trends at each site (please note that for the construction of a RC, all series are aligned to tree age, therefore the portions with lower sample replication due to the different series lengths are the older ones, where tree-growth is usually more stabilized).

At sites presenting so few (3-9), albeit well correlating, individual series, we only took the resulting indexed individual series for constructing the HSTC chronologies (how many series from what site finally entered in the HSTC used for the climate reconstruction could be further investigated). Of course having more series at all sites would be better, however with our approach no biases were introduced in the subsequent analyses. Actually, within all sites and parameters only two TRW chronologies presenting less than 15 trees were used (namely the ITRDBITAL017 14 trees, and the ITRDBITAL008 12 trees; Table 2), whereas for MXD only one chronology presenting less than 15 trees was used (ITRDBITAL008 12 trees). We underline again that no site chronology constructed with the RCS method was based on less than 12 trees.

Not giving an actual number, Esper et al. 2003 and Briffa & Melvin 2011 propose "the more samples the better", which between the lines is a minimum replication per year at 10 but coming from a population of >30 in total. Specifically Melvin (2004, Historical Growth Rates and Changing Climatic Sensitivity of Boreal Conifers, Section 6.3.3), stated you actually would need 62 samples per year for RCS to get the same per year standard deviation and confidence intervals as a 30-year spline chronology with n=10 (using Torneträsk and Finish-Lapland chronologies). "The cost for the inclusion of lowfrequency variance is a requirement for greater tree replication in order to maintain similar confidence levels."

More samples is better for getting closer to the population mean (TRW o MXD), and for stabilizing the useful statistics used for assessing the chronology quality in relation to sample replication and variability. However having more samples does not mean a better climatic signal in the chronology, as it is also evident in Leonelli et al., 2016 Climatic Change (DOI 10.1007/s10584-016-1658-5), Fig. 4:

Climatic Change (2016) 137:275–291 287

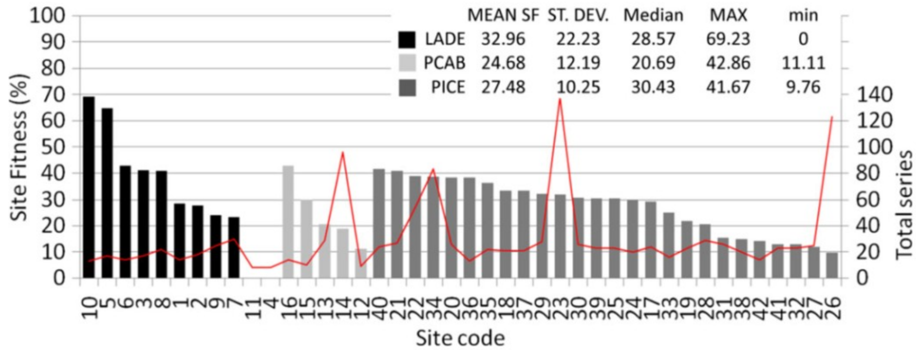

Fig. 4 Site fitness (SF) expressed as the percentage of HSTT series with respect to the total of series available at each site (*red line*). Mean values, standard deviation, median, maximum and minimum values of SF are reported in the included table

The site fitness (SF) index, expressed as the percentage of HSTT (highly sensitive to temperature) series with respect to the total of series available at each site, is sometimes very low even at sites presenting more than 100 individual series.
For example, at ITRDB SWIT219 site (https://www.ncdc.noaa.gov/paleo/study/12790), code 26 in the above figure) with a total of 123 available series, the SF index barely reached the value of 10%.

Although somewhat arbitrary it is common practice to set the EPS threshold to 0.85. The inclusion of EPS values down to 0.7 in your study tells me a lot about the "weak" coherence within your RCS chronologies (even the ones with "higher" replication of 16) during the common 1880-1980 interval. I assume the statistics would be higher (=more robust chronology) if you used a stiffer spline (~150 years) or negexp detrending instead. What are the statistics for the final RCS-HSTC-chronologies, are they >0.85?

We would have liked higher EPS values for our sites, however this statistics is not the only way for assessing the chronology quality. Chronologies with low EPS were more frequent in the TRW than in the MXD where only 3 sites over the 8 used, presented an EPS<0.8 This statistics will be added also for the HSTC in the revised version of the ms.

The Authors, May 12, 2017