Reply on RC2

The paper „Prospects for dendroanatomy in paleoclimatology – a case study on Picea engelmannii from the Canadian Rockies“ by K. Seftigen et al., explores the paleoclimatic potential of a broad set of dendro anatomical proxies, in particular focusing on the relationships between “new” proxies (mostly related to wood anatomy) and proxies that are relatively more known, such as X-ray maximum latewood density (MXD) and blue intensity (MXBI), and already used in the boreal environments of North America. My general impression about this manuscript is very positive, the paper deals with a timely issue, and such an assessment of the strengths and weaknesses of dendro anatomical proxies is a much-needed help for anyone working with “alternative” tree-ring paleoclimatic proxies. It is also notable that long chronologies of wood anatomical parameters are finally emerging even for the North American continent. For this reason, I can easily foresee that this work has the potential to become a highly cited reference in the field.
That said, I think there are still a few points that require attention.
We are glad that referee #2 found our research worthy of publication and we appreciate the opportunity to revise our manuscripts following the helpful suggestions. Below we provide further detailed description of how we have incorporated the requests and comments. Reviewer comments are followed by our replies in blue.

Major issues
The first major issue is related to the chronology development of wood anatomical parameters: in lines 159-165 it is specified that wood anatomical parameters were calculated as the 75th percentile within 20 µm wide bands parallel to the ring boundary. Nothing to say against this tree-ring partitioning approach, but it is not clear how the data obtained for each band were treated: usually, when the tree ring is partitioned, then multiple chronologies are developed. Since I do not see this in the manuscript, I assume that these data were somehow averaged. If so, it should be specified. In any case, the procedure of chronology development should be clearer, avoiding pointing to third papers.
This may be a misconception. We move the 20micron band radially over the tree-ring and for all cells enveloped by the band we take the 75percentile. When the band has been moved across the full ring, we obtain a profile of values. We take the highest value in the resulting profile for all the maximum parameters, the lowest value of the resulting profile for the minimum parameters and the average of all values within the latewood for all latewood parameters etc.. We have now clarified this methodological aspect in the manuscript (see also related comment by referee #1).

Secondly, I have major concerns related to the procedure for testing temporal stability of dendroclimatic relationships in dendro anatomical proxies (Ln250 - Ln259 & Ln452 - Ln 473):

1) if the goal is to test the temporal stability of dendroclimatic relationships obtained for the monthly data, why shift from using monthly data to daily data, using a different climatic dataset (the Berkeley Earth dataset) that is, in addition, experimental and that You seem not to trust completely (see Ln. 461-464)? I think this is simply wrong and cannot be accepted unless a thorough comparison between the two datasets (CRU TS v4.03 and Berkeley Earth dataset) is performed to check if there exist substantial differences between the two.
We thank the reviewer for suggesting this complementary test. We have now evaluated the similarity between the CRU and Berkley products by aggregating the daily Berkley product into monthly resolution (by averaging the daily values for each month) and correlating it against the tree-ring parameters (the same way as we have performed the correlation analysis between the CRU temperatures and tree-ring parameters in fig 3). These new results, added to the supplement (Fig. S6), show identical patterns to the CRU-based tree growth-climate

correlations (fig. 3), which makes us draw the conclusion that the two climate products are indeed comparable and can be used in a experimental setup as the one that we use in our work.

2) if daily data are going to be used (and I am a huge supporter of daily data, whenever available, especially with wood anatomical data) why use them in 30-days fixed windows (so basically reconducting to a monthly analysis)? This is not the approach proposed by Jevsenak and Levanic (2018), which I would actually recommend following, testing moving correlations over the same 30-year period but using temporal windows starting from i.e. 14 days up to whatever length You'd like to test (30, 60, 180 days). That would help better target the temporal windows of climatic sensitivity described in Figure 3 and Figure 5.
Actually, the R-script developed by Jevsenak and Levanic (2018) both offers a fixed (similar to our approach) and variable (as pointed out by the referee) window width when calculating the correlations between daily met data and tree-ring chronologies.
We would like to keep the analysis at it stands because the 60, 180 days widows are redundant since we already check the smaller window size of 30 days. The similarity of the correlation in one pixel to the next means that the window of temperature signal can be extended one day at the time. If the correlation coefficient is stable 30 days in a row, it means that the signal extends for two whole 30 day windows -> 60 days and so forth.. We refrained from doing the analysis with 14-day windows because the analysis would be sensitive to spurious correlations and difficult to interpret. Moreover, since we usually have significant monthly correlations with two months (Figure 3), we can detail the results insofar that if we find strong correlations for 15 pixels in a row and thereafter a drop, it means that the signal window length is 45 days starting on the Julian day of X.

A third major concern is related to the overall presentation of the paper. Despite being well written and accompanied with high-quality pictures, the manuscript is, in my personal opinion, exceptionally long, and at some point, allow me to use this term, becomes exhausting to read. I refer, for example, to the analysis related to the biases in the MXBI technique that are introduced in lines 190-199 and then discussed in lines 411-450. Honestly, I do not see the point of this long analysis and discussion. I am not criticizing its soundness nor questioning its interest. But to me, it looks like material for a stand-along paper, a deep technicality related to the blue intensity with weak (or at least not immediate) implications for dendro anatomy within the context of this manuscript. I believe that the strength of this work is that it is exploring important technical details related to those dendro-anatomical proxies that are less known in this region. It should not focus on going into deep technical details related to proxies or techniques that are already relatively more known and applied.
Another example is in lines 368-385. This long paragraph discussing the climatic response of EW features could be easily avoided, since it is evident throughout the entire manuscript, and according to Table 1, that EW signal is not predominant at this location. Since the paper is already very long, this long (and sometimes repetitive) discussion on EW density and lumen area could be avoided. The last example could be section 3.1, where several times concepts discussed in the following sections are already mentioned. In general, I suggest careful evaluation throughout the manuscript if certain paragraphs and/or concepts could be removed or rephrased to improve the readability of this paper.
We have followed the advice and shortened the manuscript where we find appropriate. For example, we have removed the paragraph 368-385 according to the suggestion, and also removed the last section discussing long-term trends (following the suggestion made by reviewer 1). We have also split the results/discussion section into sub-sections including new subheadings "Temporal signal stability" and "Possible implications of measurement resolution on climate signal". We hope that this will make the text more "airy" and easy to navigate.

We have decided to keep the discussion around MXBI and its potential resolution bias as these results are significant when interpreting the strength and seasonality of the temperature imprint withing the various parameters. Moreover, the analysis around the MXBI resolution issue was appreciated by referee #2, which indicates that these results may be of interests to the community and worthy a publication.

Minor issues:

The paper introduces the longest dataset of dendro-anatomical parameters for North America but there is not a figure showing such 1585-2014 chronology in the paper for any parameter. I suggest adding it, even if the analysis is focused on the 1700-1994 period. Moreover: why is the chronology truncated in 1994? And the data in Table 1 refers to the 1585-2014 or 1700-1994 timeframe? Please clarify. *We have followed these suggestion and 1) added a new figure to the supplement showing the full 1585-2014 chronology for the aMXD and Max. radial CWT (fig. S4), clarified in the ms the reason for the truncation to year 1994 (to match the X-ray dataset), and also 3) provided info on which period the table 1 statistics are computed (the common 1700-1994 period, info added to the table caption).*

Lines 320-331 Could be integrated into the next section. *Integrated and also partly removed to shorten the text.*

Lines 360-361 “Even though the parameters describe two temporally distinct temperature signals, both are encapsulated within the short June-July-August period”. What are the two temporally distinct temperature signals? Please clarify. *Sentence removed.*

Lines 363-364 “…average monthly temperatures rise above 0 °C only in four months of the year”. In Fig 1c it is actually five months. *Typo, now changed to five months.*

Lines 364-366 “This window is substantially shorter than the single but wide target season observed in the latewood anatomical traits of P. sylvestris growing in temperature-limited environments in northern Scandinavia”. I don’t understand this sentence (i.e. what is the “target season observed in latewood anatomical traits”?), please be more clear. *Rephrased to “[…] comparing our results with the previous study of (Björklund et al. 2020) on latewood anatomical traits of P. sylvestris in northern Scandinavia, where the temperature response window extends from April to September.” .*

Lines 368-385 Regardless of my previous comment (see above), I would like to add a few words about this. The reversing relationship observed between June-July temperature and EW density and lumen area might be due to an effect of precipitation, or in general of moisture availability, at that time. The initial stages of wood formation (hence, the anatomical features of the cells formed firstly along the ring) are highly dependent on water availability, which determines cell turgor in the enlargement phase. This is generally reflected in the EW lumen features, such as lumen area or lumen radial diameter. As a general remark, I would not say that EW features (i.e. lumen area) are not suitable climate proxies (or at least I would properly contextualize it): this might (is) true for this location and for temperature limited environments, but in arid environments earlywood (not latewood) features (in particular lumen size) are a crucial climatic proxy (not cell wall thickness). *The paragraph the referee refer to is now removed. Nevertheless, we agree with the reviewer that moisture availability in many cases in an important driver for cell enlargement in the EW, especially in drought-prone environments. However, correlation analysis with precipitation for our site and species do not show any significant link to moisture availability, which makes us believe that this environmental constraint is of secondary importance.*

Figure 4 The r coefficients listed on the right side, how are the aggregate months' correlation computed? Is it an average of the two r? Why some values are listed for both

single and aggregated months, and in other cases not, even though the correlations seem equally strong? *We believe that the reviewer refer to fig 3 here? The temperature data have first been averaged over the specific season, and then correlated against the tree-ring chronology to obtain the correlation coefficients. The seasons/monthly windows are selected based on the strongest correlation between the tree-ring data and temperatures. We have now harmonized the figure to make it less confusing: for ring width and all EW parameters we now show June correlation coefficients, and for all LW parameters July-August and August temperature correlations.*

Lines 458-459 Please clarify the meaning of "peripheral ends" and of "elusive". *Re-worded, now reads "The temperature associations at the margins of the target season are, however, more unstable.".*

Lines 484-489 This is a key methodological aspect that lead to one of the main findings of this paper, hence I was surprised to find it here and not in the Methods section. Please move it to the Methods section. Done according to suggestion. *The paragraph has been moved to sect. 2.3 in the methods.*