

Peer Review File

Manuscript Title: Tree-ring anatomy provide new benchmarks for medieval and current warming

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

I read this manuscript with considerable interest as the authors address a long-standing issue in tree-ring based reconstructions of climate over the past millennium. They leverage recent advances in quantitative wood anatomy to revisit one of the canonical northern hemisphere temperature reconstructions and address the well-described discrepancy between tree-ring based reconstructions of temperatures and outputs from climate model ensembles for the Medieval Warm Period. I think this is an exciting and important piece of work and deserves wide attention. I suspect, though, that it would be better suited to one of Nature's discipline-focused journals (eg, Nature Geosciences) as it resolves a specific technical issue—the relative stationarity of the climate signal in tree rings vs the anatomical features of individual cells within the tree rings—albeit for an issue of great significance.

In addition, I felt that the argument for the A-FEN chronology was a bit fragile. As I understood the manuscript, the main reason to believe the A-FEN over the X-FEN is that it better aligns with the CMIP ensemble temperatures. This sails very close into the wind of confirmation bias. “We like this reconstruction over that reconstruction because it aligns with what we think is the right answer.” Perhaps this is a subtle issue, but absent some independent confirmatory data, the primary argument in favour of the A-FEN reconstruction is that it lines up better with the model results. While I realise that independent data to confirm this is not available, I think you would need to be more careful in how you frame this. It would make more emphasise that they align, rather than the QWA supports the models. For example, you could more clearly articulate that 1) the models integrate current knowledge of the system physics and are consistent with past climate dynamics at the event scale (eg, volcanoes) and at broader trends (eg, CWP dynamics) and 2) the QWA is less prone to a number of the potential biases that tree rings (or MXD) suffer from. As such, their alignment is strong evidence for a more moderate MWP (rather than the QWA support the model results on this), etc. Perhaps if you could show better agreement between A-FEN and models over the instrumental period compared to X-FEN that would make for a stronger argument. The inset box on the bottom right of Figure 2c shows this, but the point is not well-articulated in the manuscript. In the Supplementary Material the only comparison was in relationship to the spectral properties of the A-FEN and X-FEN vis-a-vis the ensemble range (ED Fig 8a,b). Perhaps this is a subtle distinction, but I felt that the manuscript did not address this issue as strongly as it could have.

The manuscript itself is very well presented (although the prose can be somewhat purple on occasion). I provide some minor copy-editing comments below.

Some specific comments:

L72 Delete “vast”

L81 Delete “an”

L88 Why is arctic circle italicised?

L91 “reconstruction”

L93 and L95 “state-of-the-art reconstructions” and “the latest...reconstructions” surely are similar enough to collapse into a single item in this list?

Figure 2. No subfigure labels for a) and c)?

L170 Replace “That is” with “Specifically”

L179 “would thus arguably” isn’t very convincing. It sounds quite arbitrary when presented this way.

L203 Replace “likely means” with “it appears”

L210 What does “quietly” forced mean? This seems rather subjective. Perhaps “minimally forced”?

See also L278

L233 “at the regional”

L239-242 This is an odd statement. Why would someone reconstructing climate knowingly do any of these things? It may well be that these things occur, but “systematically incorporated”? That seems a bit much.

L248 Delete “stunning”

L250 “A-FEN’s”

L261 Replace “By contrast” with “Rather”

L269-270 “involves” and what is “brute decision-making” Perhaps word is missing? (“brute-force decision-making”?)

L278 “constraint”

L290 “...in the Wilson...”

L306-307 “year for the period...”

L312-335 It isn’t clear from this description of the QWA methods how you determine what is latewood. Figure 1 states that this is explained in this section. But it is not.

L346 “Standardisation approach.”

L429 This doesn’t make sense. Binning the 100 warmest extremes each century? There are 100 years in a century, so this would include all years, warm or cold...

Referee #2 (Remarks to the Author):

Nature manuscript 2023-02-02393A.

Tree-ring anatomy provide new benchmarks 1 for medieval and current warming.

In the current manuscript, the authors use a new tree-ring proxy based on cell wall thickness of pine tracheids to develop a millennial reconstruction of May through August temperature in northern Fennoscandia. This new reconstruction agrees with Northern Hemisphere reconstruction and climate model ensembles much better than previous reconstructions based on maximum wood density from some of the trees used in the new reconstruction. As the manuscript correctly reflects, extending these reconstructions based on quantitative wood analyses of tree rings to larger spatial

and temporal scales would bring much needed more precise climate reconstructions that better represent past climate variability.

I found the manuscript to be very well written, and all results and discussion well developed and interpreted. The authors test their data and results very thoroughly for different potential biases and artifacts (frequency domains, sample sizes, periods in time etc) which really helps to convey the robustness of their results. I believe that with very minor corrections and clarifications the paper could be accepted for publications.

I think the title could be improved by adding “in northern Fennoscandia” or “in Fennoscandia” at the end. It would probably be a less attractive title, but it would be more precise.

L42: annually resolved

L43: because tracheids are cells it seems incorrect to use “tracheid cell measurements” here.

L44: European observed temperatures or instrumental temperatures.

L49: tree ring quantitative wood anatomy

L49: uncertainties in past “temperature variability including” past and cold and warm episodes

L61: this may just be a typo, but there were also women in Norse settlements in Greenland. Please correct.

L86: I don’t see how the number of tracheids measured is useful here. It may be in the methods section. Here, it would be better to add the number of trees, and cores which actually show sample replication.

L92: this sentence is imprecise. Those tests were applied to the QWA data, but mostly to the resulting temperature reconstruction.

L127: which timescales: frequency domain, time periods, annual periods?

L128-130: would these problems be solved by adding or removing samples from these analyses? This sentence is probably not clear enough.

L160: “from 850-1850” should probably be rephrased as “between 850 and 1850” or “for the period 850-1850” or something similar.

L168: was A-FEN used indirectly in any of these products?

L197: “overestimating the response of ...” should only be used when the estimation is compared to a relatively known true value which is probably not the case here.

L232: change “thought” with “reported”

L311. I found this part of the methods not as well written as the rest of the manuscript. There may be too much detail or it maybe it just needs to be better organized.

L337. Clarify between exactly what delta radial cell wall thickness is. Was the difference computed between the mean earlywood CWT and mean latewood CWT? Something else.

Referee #3 (Remarks to the Author):

Review of Bjorklund et al., Tree-ring anatomy provide new benchmarks for medieval and current warming.

I enjoyed reading your paper. This study makes a substantial contribution to the effort to better understand and represent paleotemperatures. The focus on measurement of anatomical parameters strongly improves upon one of the most well-established and replicated tree-ring proxy records for temperatures in northern Europe and clarifies our understanding of global temperature trends. The study is carefully designed, established and proven field, laboratory, and analytical techniques were followed closely, novel approaches in QWA are robust, and the manuscript is nicely organized and clearly presented. The benefit of QWA is particularly convincingly presented, providing a strong argument for the application of these techniques to enhance efforts in dendroclimatology. The calibration and verification statistics are strong, and the qualitative validation with independent data provides further evidence of the robustness of the results.

-Better at capturing extreme values

Below I address the specific standards provided to me by the journal and detail the degree to which each standard has been met.

Key Results – A more mechanistic connection between environment and tree growth, and a more statistically robust assessment of past and current temperature trends that more closely match theoretical and modeled expectations of climate compared to previous efforts based on density measurements.

Validity – I do not see any flaws that would prohibit publication.

Originality and significance – QWA techniques have been in development for several years now (principally via the efforts of the authors of this study), but the application to a millennial-length reconstruction of temperature is novel, and the robustness and skill of the proxy in representing observations and theoretical expectations is unprecedented as far as I am aware. Global interest in this study will certainly be immediate and extend across disciplines.

Data & methodology – I reviewed the figures and tables in the main manuscript, and extended & supplementary material. I am in agreement with the priority placement of the data and they are clearly presented with sufficient detail. The RCS approach was appropriate for the study and appears to have been properly carried out, providing another improvement over previous work. The figures are clear and nicely prepared. Please note however that the X-axis titles for Extended data figs 3, 6, and 7 have been slightly cut-off by the caption.

Appropriate use of statistics & treatment of uncertainties – I see no problems with the data and appropriate treatment/presentation of error. Weaknesses of the study are also acknowledged and properly addressed.

Conclusions: robustness, validity, reliability – the conclusions/summary provided in the abstract and main manuscript are appropriate and within the bounds of the data. If anything, I might even suggest a stronger title. “New benchmarks” is appropriate but the “revision” language in the abstract (L 46) and conclusions (L 159) is stronger and more direct (and appropriate given the data). The calibration/verification stats are very strong, and the qualitative match of the study data with independent reconstructions (L 167-169) provides further evidence of the robustness and validity of the results.

References: appropriate credit to previous work – the literature cited seems thorough, appropriate, and current.

Clarity and context: With very few exceptions, detailed below, the manuscript is very clearly presented.

Suggested improvements: I cannot think of any additional measurements or analysis that would improve upon the study as it has been carried out. Below I detail suggestions and edits on the presentation of the manuscript that I hope will improve upon the final presentation. Otherwise I am very satisfied with the paper and I look forward to seeing it published.

L78: I'd recommend a different word than “salvaging” here (negative connotations), Perhaps “retrieving” or “recovering”?

L87: Why not be more specific here than “several” and “half dozen”?

103: Use lower case for “blue”

104: Comma use here makes it difficult for the reader to determine exactly which parameters were measured.

106: should be “e” not “g”

127: The language is a little fuzzy here. I think you are trying to say that A-FEN performs across multiple timescales compared to X-FEN which only performs well with May-Aug or June-Aug. Perhaps “whereas” instead of “where” would clarify this.

L133 Fig 2: I do not see an “a)” or “c)” on the panels on my copy of the figure.

L128 and 146: Try to avoid the two uses of “moreover” so close together.

L148: The hyphen-use is confusing here. I'd recommend “widespread” as a compound word and probably “pine-tree-canopy damage”.

L154: Probably “artifact” depending on the preference of the journal.

L163: “volcanic, and unperturbed solar forcing” is unclear

L165: I was very confused about “elusive volcanism in 1453” until L190. It makes it sound like you were unable to track/identify any volcanic activity at that point rather than an unidentified eruption. Consider revising.

L202-204: Some awkward phrasing here – “likely means” doesn’t fit the structure of the sentence. Please revise.

Author Rebuttals to Initial Comments:

Dear Editor, first of all we sincerely appreciate this opportunity to be able to revise our manuscript and also wish to thank you and the reviewers for constructive critiques and insightful comments. We believe that the manuscript has benefited greatly from the revision process, and we hope that we have addressed the feedback from the reviewers in a satisfactory manner. In the following, we provide point-by-point responses (in bold face) to the reviewer feedback (in normal font). We highlight the changes also in an appended Track Changes document (TCD) and refer specifically to lines and pages in that document when changes were made.

Referee #1 (Remarks to the Author):

I read this manuscript with considerable interest as the authors address a long-standing issue in tree-ring based reconstructions of climate over the past millennium. They leverage recent advances in quantitative wood anatomy to revisit one of the canonical northern hemisphere temperature reconstructions and address the well-described discrepancy between tree-ring based reconstructions of temperatures and outputs from climate model ensembles for the Medieval Warm Period. I think this is an exciting and important piece of work and deserves wide attention. I suspect, though, that it would be better suited to one of Nature's discipline-focused journals (eg, Nature Geosciences) as it resolves a specific technical issue—the relative stationarity of the climate signal in tree rings vs the anatomical features of individual cells within the tree rings—albeit for an issue of great significance. **Response: We sincerely appreciate the work and constructive critique of R#1 and have made changes accordingly, to the best of our ability.**

In addition, I felt that the argument for the A-FEN chronology was a bit fragile. As I understood the manuscript, the main reason to believe the A-FEN over the X-FEN is that it better aligns with the CMIP ensemble temperatures. This sails very close into the wind of confirmation bias. “We like this reconstruction over that reconstruction because it aligns with what we think is the right answer.” Perhaps this is a subtle issue, but absent some independent confirmatory data, the primary argument in favour of the A-FEN reconstruction is that it lines up better with the model results. While I realise that independent data to confirm this is not available, I think you would need to be more careful in how you frame this. It would make more emphasise that they align, rather than the QWA supports the models. *For example, you could more clearly articulate that 1) the models integrate current knowledge of the system physics and are consistent with past climate dynamics at the event scale (eg, volcanoes) and at broader trends (eg, CWP dynamics) and 2) the QWA is less prone to a number of the potential biases that tree rings (or MXD) suffer from. As such, their alignment is strong evidence for a more moderate MWP (rather than the QWA support the model results on this), etc.* Perhaps if you could show better agreement between A-FEN and models over the instrumental period compared to X-FEN that would make for a stronger argument. The inset box on the bottom right of Figure 2c shows this, but the point is not well-articulated in the manuscript. In the Supplementary Material the only comparison was in relationship to the spectral properties of the A-FEN and X-FEN vis-a-vis the ensemble range (ED Fig 8a,b). Perhaps this is a subtle distinction, but I felt that the manuscript did not address this issue as strongly as it could have.

Response: We thank R#1 for this insightful comment. Although, we argue that we show a range of traits in favor of A-FEN both empirically observed and theoretically articulated regardless of any model results, we realize that they were not highlighted enough. We also appreciate R#1 comments about the distinctive roles of proxies and models. We have implemented these clarifications and principles to the best of our ability. Hopefully the manuscript will more directly convey the reasons for favoring A-FEN over X-FEN:

- 1) statistically superior fidelity compared to instrumental data,
- 2) statistically stronger traits such as inter-tree correlation accompanied by a balanced replication outside of the instrumental calibration period, as well as
- 3) a theoretical framework highlighting potential biases in tree-ring measurements, favoring A-FEN over X-FEN.

The alignment with the models comes on top of these advantages, and the alignment may further be related to the modest biological noise structure of the new wood anatomical tree-ring proxy. A-FEN likely contain less biological noise than X-FEN, highlighted especially by the pine canopy damage (but also quantitatively in Extended Data Figures 4 and 5). During the pine canopy event, the trees were physically damaged, and the trees formed rings that reflected this damage (i.e., not climatically driven radial growth). A-FEN, however, seemed unaffected by this, and continued to track climate, in contrast to X-FEN. The non-climatic biological noise can thus obscure real relationships between climate variations derived from proxies if compared with models, because the models do not have this noise structure (Gonzales-Rouco et al 2015 Clim. Past). We have added text to increase clarity around these points at the end of the manuscript.

We agree with the suggested text of R#1 that the point is that A-FEN and Models represent different perspectives on the same target, and that their independent strengths, accompanied by their alignment, are compelling evidence for an only moderately warm MCA. We have implemented a text clarifying this in the last section as well. These texts can be found on P6 L28 – P7 L11 in TCD.

The manuscript itself is very well presented (although the prose can be somewhat purple on occasion). I provide some minor copy-editing comments below.

Some specific comments:

L72 Delete “vast” **Response:** removed from L4 P3 in TCD

L81 Delete “an” **Response:** removed from L12 P3 in TCD

L88 Why is arctic circle italicised? **Response:** removed L20 P3 TCD

L91 “reconstruction” **Response:** We would like to keep the plural on L22 P3 TCD

L93 and L95 “state-of-the-art reconstructions” and “the latest...reconstructions” surely are similar enough to collapse into a single item in this list? **Response:** We meant to distinguish between corresponding regional state-of-the-art reconstructions and the latest

hemispheric and global reconstructions, but see the point by the reviewer and therefore changed the text to “relevant reconstructions” for brevity. L23-26 P3 TCD

Figure 2. No subfigure labels for a) and c)? **Response: Now included**

L170 Replace “That is” with “Specifically” **Response: We condensed this sentence and the expression is not used anymore. On L10-12 P5 TCD we changed to:**

“The multi-decadal fluctuations of A-FEN are similar to those of X-FEN, but X-FEN most notably portrays MCA as exceptionally warm, and CWP only with above average temperatures (Fig. 2c).”*

L179 “would thus arguably” isn’t very convincing. It sounds quite arbitrary when presented this way. **Response: We change arguably for “likely”. L18 P5 TCD**

L203 Replace “likely means” with “it appears” **Response: This sentence was changed:**

From:

Since forced models are unable to reproduce an exceptionally warm medieval period and since the more precise QWA approach also cannot reproduce this warming, likely means that the regional expression of the MCA must have been cooler than previously estimated.*

To:

Since the forced models, as well as A-FEN, are both unable to reproduce an exceptionally warm medieval period, their alignment thus forms compelling new evidence for a more moderate MCA than previously estimated.*

And then moved to L6-9 P7 TCD

L210 What does “quietly” forced mean? This seems rather subjective. Perhaps “minimally forced”? See also L278 **Response: It is a bit difficult to find the right wording here. By quietly forced we mean minimal volcanic forcing and stable (but not necessarily low) solar forcing. We chose the quietly forced term because it is defined in Bradley, R. S., Wanner, H. & Diaz, H. F. The medieval quiet period. The Holocene 26, 990-993 (2016), and that “minimally forced” as suggested, only seems correct for the volcanic forcing. We suggest keeping the current wording, but are open for changes. L8 P6 in TCD**

L233 “at the regional” **Response: Done L22 P6 TCD**

L239-242 This is an odd statement. Why would someone reconstructing climate knowingly do any of these things? It may well be that these things occur, but “systematically incorporated”? That seems a bit much. **Response: We see the point and did not mean that this was made intentionally by any researchers, but the resulting reconstructions may have a systematic bias due to the properties of the used tree-ring data. This statement is removed for brevity and clarity.**

L248 Delete “stunning” **Response: Sentence removed**

L250 “A-FEN’s” **Response: Done, L24 P6 TCD**

L261 Replace “By contrast” with “Rather” **Response: The paragraph starting with “By contrast” is now removed.**

L269-270 “involves” and what is “brute decision-making” Perhaps word is missing? (“brute-force decision-making”?) **Response: sentence is no longer included.**

L278 “constraint” **Response: Done L16 P7 TCD**

L290 “...in the Wilson...” **Response: Done L18 P9 TCD**

L306-307 “year for the period...” **Response: Done L3 P10 TCD**

L312-335 It isn’t clear from this description of the QWA methods how you determine what is latewood. Figure 1 states that this is explained in this section. But it is not. **Response: We apologize for this mistake. We have amended the methods section with the following clarification: The latewood in each ring was defined using a Mork’s index value greater than 1. That is, those tangential cell bands where twice the double cell wall thickness was larger than the lumen diameter across the radial direction on average, were defined as latewood. On L1-4 P11 TCD**

L346 “Standardisation approach.” **Response: Done L16 P11 TCD**

L429 This doesn’t make sense. Binning the 100 warmest extremes each century? There are 100 years in a century, so this would include all years, warm or cold... **Response: We hopefully clarified this by changing the sentence to: “The incidence of warm extremes was quantified by first extracting the 100 warmest years from the common overlap 850-2005 CE of each simulation and reconstruction. The extracted extremes were then summed per each century and compared with confidence intervals based on number of extremes happening by chance.” L22-25 P16 TCD**

Referee #2 (Remarks to the Author):

Nature manuscript 2023-02-02393A.

Tree-ring anatomy provide new benchmarks for medieval and current warming.

In the current manuscript, the authors use a new tree-ring proxy based on cell wall thickness of pine tracheids to develop a millennial reconstruction of May through August temperature in northern Fennoscandia. This new reconstruction agrees with Northern Hemisphere reconstruction and climate model ensembles much better than previous reconstructions based on maximum wood density from some of the trees used in the new reconstruction. As the manuscript correctly reflects, extending these reconstructions based on quantitative wood analyses of tree rings to larger spatial and temporal scales would bring much needed more precise climate reconstructions that better represent past climate variability.

I found the manuscript to be very well written, and all results and discussion well developed and interpreted. The authors test their data and results very thoroughly for different potential biases and artifacts (frequency domains, sample sizes, periods in time etc) which really helps to convey the robustness of their results. I believe that with very minor corrections and clarifications the paper could be accepted for publications. **Response: We sincerely appreciate the work and constructive critique of R#2 and have made changes accordingly**

I think the title could be improved by adding “in northern Fennoscandia” or “in Fennoscandia” at the end. It would probably be a less attractive title, but it would be more precise.

Response: The title is now changed to “Fennoscandian tree-ring anatomy shows modern climate warmer than Medieval” as suggested by the editor.

L42: annually resolved **Response: “annually resolved” inserted before “1170-year” L11 P2 TCD**

L43: because tracheids are cells it seems incorrect to use “tracheid cell measurements” here. **Response: We believe that simply using “tracheid” would be less accessible to the broad readership of the journal, and would prefer to at least use the wording tracheid anatomical measurements”. L12 P2 TCD. “Cell” is now removed.**

L44: European observed temperatures or instrumental temperatures. **Response: “instrumental” inserted L13 P2 TCD**

L49: tree ring quantitative wood anatomy **Response: the relevant section of this sentence is removed.**

L49: uncertainties in past “temperature variability including” past and cold and warm episodes **Response: the relevant section of this sentence is removed.**

L61: this may just be a typo, but there were also women in Norse settlements in Greenland. Please correct. **Response: “Norse men to settle” is changed to “Norse settlements” L26 P2 TCD**

L86: I don't see how the number of tracheids measured is useful here. It may be in the methods

section. Here, it would be better to add the number of trees, and cores which actually show sample replication. **Response: “c. 50 million tracheid cell measurements from” is now removed, the number of trees is kept in the sentence. L18 P3 TCD**

L92: this sentence is imprecise. Those tests were applied to the QWA data, but mostly to the resulting temperature reconstruction. **Response: QWA data changed to “A-FEN” L23-24 P3 TCD**

L127: which timescales: frequency domain, time periods, annual periods? **Response: various timescales is replaced by “inter-annual to decadal timescales” L14 P4 TCD**

L128-130: would these problems be solved by adding or removing samples from these analyses? This sentence is probably not clear enough. **Response: The problem of X-FEN’s fading quality could be addressed if one would remove 60-100 samples covering the past 150 years, but then the calibration statistics would likely reflect this change for the worse (worse climate calibration – but no reason to expect fading quality back in time), and the difference in climate fidelity between X-FEN and A-FEN would thus be larger than found in this study. Hopefully, the sentence is now more clearly phrased on L14-17 P4 TCD:**

“A-FEN is potentially less susceptible to fading reconstruction quality than X-FEN. This is because X-FEN Rbar and replication calculated over the calibration period (1850-2019) substantially exceeds corresponding statistics for the reconstruction period (850-1849) (Fig. 2b).”

L160: “from 850-1850” should probably be rephrased as “between 850 and 1850” or “for the period 850-1850” or something similar. **Response: changed to: between 850 and 1850. L2 P5 TCD**

L168: was A-FEN used indirectly in any of these products? **Response: You could probably say that because the samples used to make QWA for this study were also used to make X-ray MXD and ring width in previous studies, and those latter data were used in the large-scale products. The sentence is now removed.**

L197: “overestimating the response of ...” should only be used when the estimation is compared to a relatively known true value which is probably not the case here. **Response: we agree, the sentence is now shortened, and the expression excluded.**

L232: change “thought” with “reported” **Response: Done L21 P6 TCD**

L311. I found this part of the methods not as well written as the rest of the manuscript. There may be too much detail or it maybe it just needs to be better organized. **Response: We have now tried to streamline the structure and clarity of this section. L7 P10 – L5 P11 TCD**

L337. Clarify between exactly what delta radial cell wall thickness is. Was the difference computed between the mean earlywood CWT and mean latewood CWT? Something else.

Response: Now clarified on L10-11 P11 TCD

Referee #3 (Remarks to the Author):

Review of Bjorklund et al., Tree-ring anatomy provide new benchmarks for medieval and current warming.

I enjoyed reading your paper. This study makes a substantial contribution to the effort to better understand and represent paleotemperatures. The focus on measurement of anatomical parameters strongly improves upon one of the most well-established and replicated tree-ring proxy records for temperatures in northern Europe and clarifies our understanding of global temperature trends. The study is carefully designed, established and proven field, laboratory, and analytical techniques were followed closely, novel approaches in QWA are robust, and the manuscript is nicely organized and clearly presented. The benefit of QWA is particularly convincingly presented, providing a strong argument for the application of these techniques to enhance efforts in dendroclimatology. The calibration and verification statistics are strong, and the qualitative validation with independent data provides further evidence of the robustness of the results. **Response: We sincerely appreciate the work and constructive critique of R#3 and have made changes accordingly**

-Better at capturing extreme values

Below I address the specific standards provided to me by the journal and detail the degree to which each standard has been met.

Key Results – A more mechanistic connection between environment and tree growth, and a more statistically robust assessment of past and current temperature trends that more closely match theoretical and modeled expectations of climate compared to previous efforts based on density measurements.

Validity – I do not see any flaws that would prohibit publication.

Originality and significance – QWA techniques have been in development for several years now (principally via the efforts of the authors of this study), but the application to a millennial-length reconstruction of temperature is novel, and the robustness and skill of the proxy in representing observations and theoretical expectations is unprecedented as far as I am aware. Global interest in this study will certainly be immediate and extend across disciplines.

Data & methodology – I reviewed the figures and tables in the main manuscript, and extended & supplementary material. I am in agreement with the priority placement of the data and they are clearly presented with sufficient detail. The RCS approach was appropriate for the study and appears to have been properly carried out, providing another improvement over previous work. The figures are clear and nicely prepared. Please note however that the X-axis titles for Extended data figs 3, 6, and 7 have been slightly cut-off by the caption. **Response: We will prepare separate figures for the resubmission so this will hopefully not be a problem anymore.**

Appropriate use of statistics & treatment of uncertainties – I see no problems with the data and appropriate treatment/presentation of error. Weaknesses of the study are also acknowledged and properly addressed.

Conclusions: robustness, validity, reliability – the conclusions/summary provided in the abstract and main manuscript are appropriate and within the bounds of the data. If anything, I might even suggest a stronger title. “New benchmarks” is appropriate but the “revision” language in the abstract (L 46) and conclusions (L 159) is stronger and more direct (and appropriate given the data). The calibration/verification stats are very strong, and the qualitative match of the study data with independent reconstructions (L 167-169) provides further evidence of the robustness and validity of the results.

References: appropriate credit to previous work – the literature cited seems thorough, appropriate, and current.

Clarity and context: With very few exceptions, detailed below, the manuscript is very clearly presented.

Suggested improvements: I cannot think of any additional measurements or analysis that would improve upon the study as it has been carried out. Below I detail suggestions and edits on the presentation of the manuscript that I hope will improve upon the final presentation. Otherwise I am very satisfied with the paper and I look forward to seeing it published. **Response: We are delighted about this positive evaluation and provide point-by-point responses in the following.**

L78: I'd recommend a different word than “salvaging” here (negative connotations), Perhaps “retrieving” or “recovering”? **Response: changed to “retrieving” L10 P3 TCD**

L87: Why not be more specific here than “several” and “half dozen”? **Response: We like to keep this a bit vague as we do not have 100% overview of all datasets, their length, distribution, etc. But we are certain that Fennoscandia currently hosts a disproportionate share of these long datasets (Anchukaitis et al. 2017). Therefore, we rephrased this to: “...hosts a disproportionate share of the millennial...” L19 P3 TCD**

103: Use lower case for “blue” **Response: Done L27 P7 TCD**

104: Comma use here makes it difficult for the reader to determine exactly which parameters were measured. **Response: We have now tried to improve clarity here. L30-31 P7 TCD**

106: should be “e” not “g” **Response: Done L3 P8 TCD**

127: The language is a little fuzzy here. I think you are trying to say that A-FEN performs across multiple timescales compared to X-FEN which only performs well with May-Aug or June-Aug.

Perhaps “whereas” instead of “where” would clarify this. **Response: Actually, we meant that both A-FEN and X-FEN perform similarly well regardless of which target season (MJJA or JJA) is used. And that A-FEN is consistently stronger for both target seasons at all scales that we tested (high-pass filter, untreated, and low-pass filtered data). We hope the wording is clarified by changing the sentence to: “Although the climatic imprint in A-FEN is similar to the MXD-based corresponding reconstruction 9 (X-FEN), A-FEN is significantly stronger (Fig. 2a), and consistently performs better across inter-annual to decadal timescales.” L12-14 P4 TCD**

L133 Fig 2: I do not see an “a)” or “c)” on the panels on my copy of the figure. ” **Response: Apologies, now inserted.**

L128 and 146: Try to avoid the two uses of “moreover” so close together. ” **Response: we removed one “moreover” on L14 P4 TCD**

L148: The hyphen-use is confusing here. I’d recommend “widespread” as a compound word and probably “pine-tree-canopy damage”. ” **Response: changed to: “widespread pine tree canopy damage” L18-19 P4 TCD**

L154: Probably “artifact” depending on the preference of the journal. ” **Response: Done L27 P4 TCD**

L163: “volcanic, and unperturbed solar forcing” is unclear ” **Response: changed to “stable solar forcing” L5 P5 TCD**

L165: I was very confused about “elusive volcanism in 1453” until L190. It makes it sound like you were unable to track/identify any volcanic activity at that point rather than an unidentified eruption. Consider revising. **Response: changed to “...still unidentified volcanism around...” L6-7 P5 TCD**

L202-204: Some awkward phrasing here – “likely means” doesn’t fit the structure of the sentence. Please revise. **Response: Sentence removed.**

Sincerely, Jesper Bjorklund on behalf of all co-authors