Reply to comments by Referee #1 (Comments from the referee in bold font and replies in italics) regarding manuscript: Is blue intensity ready to replace maximum latewood density as strong temperature proxy? A tree-ring case study on Scots pine from Northern Sweden, Björklund et al.

We sincerely thank Referee #1 for the thorough examination of the manuscript, and insightful comments. We hope that we have addressed/answered all the expressed concerns and changed all the errors included. We think that the comments from the referee have greatly improved the revised manuscript.

General Comment

There is no doubt that the relatively new and essentially untested parameter Blue Intensity holds great promise and Björklund et al. present a timely paper on utilising this parameter (along with modifications - i.e. ΔBI) for dendroclimatic reconstruction of past summer temperatures.

My major worry about this paper is their insistence that BI must be converted/transformed to proxies of density. Why? Firstly, are BI and MXD actually measuring the exact same wood properties? I think wood density is related to cell wall thickness, which is in turn controlled, by cellulose, hemi-cellulose and lignin content. As far as I understand the theory, BI is measuring the absorption and related reflectance of the light from lignin at the surface of the sample which is strongest in the blue part of the frequency spectrum.

Therefore, MXD and BI are similar parameters w.r.t. wood properties they measure, but ultimately are measuring slightly different things. Therefore, converting BI to density seems an unnecessary step and in my mind they should NOT be treated in the same way and we should not expect both parameters to show exactly the same characteristics.

This ultimately does not change the main result of the paper w.r.t. the use of ΔBI and ΔMXD but would cut out some of the complexities and rambling text (in some sections) of the paper.

As stated x-ray absorption is foremost controlled by lignin, cellulose and hemi-cellulose. When it comes to BI, again lignin, cellulose and hemicellulose are reflected, and here lignin is perhaps dominating the reflectance more than in the x-ray absorption. We agree that the different methods likely do differ in what they measure and to what degree they "pick up" different components in the wood, however lignin-, cellulose- and hemicellulose-content vary together in almost indistinguishable ways in the wood (personal communication: Dr. Michael Jarvis, University of Glasgow, Expert on biological materials like wood and on their constituent polymers like cellulose and lignin). And even if deviations from this relationship can occur in compression wood (Leonardon et al. 2009) we argue that

the lignin content and wood density should be highly correlated and linearly related, since compression wood likely is distributed randomly in the sample-population over time. And if the x-ray method after calibration is a measure of density then BI should arguably also be calibrated to be a reflection of lignin content. At least, the possibility of a non-linear relationship with what is of interest should be explored.

Since BI and X-Ray transmission are so excellently correlated and describe a linear relationship (Fig 6, supplementary material) and to model X-Ray Density with BI, a non-linear relationship is the most appealing (Fig. 6 lower panel, supplementary material) we suggest that also BI is non-linearly related to lignin-content, wood density or other wood structural components. Consequently, we still argue that the transformation of blue light intensity into a proxy for density is a significant contribution to the research on BI. The manuscript is not aiming to state that BI is density, it is just that BI behaves so suspiciously like density, that it is interesting and useful to try to reconstruct density.

However, we realize that it is complicating the manuscript, and have chosen to comply with the referee and remove the calibration sections. Suggestions for future work including calibration is however still recommended either in the current form or modified.

Also - the authors bounce between 1st difference transforms for response function analysis (RFA) and showing RCS processed chronologies for MXD but not BI. I do not see why they do not do the RFA using the non-1st differenced transformed chronologies. And why not also consider individual series data adaptive detrending options such as linear or Hugershoff functions. The RFA will be susceptible to biases in the mid- lower frequency domain and 1st differencing removes that information.

We have now included a climate correlation analysis for the Arjeplog data in the supplementary material. Here, all the parameter-chronologies are standardized with traditional RCS. Also the residual chronology from the RCS is included and these two on the background of the first differenced RCS chronology. The standardized chronologies are correlated against the monthly climate variables mean temperatures and precipitation sums.

Also, RCS could be used on the BI data, but the sub-fossil and living data would have to be divided into two groups to take into account the different "reflective" properties between these sample sub-sets.

A division can be done, with very good results (Poster PAGES Goa 2013, Björklund et al. 2013), but the question of a subdivision between different ages of dead material, living material would maybe soon arise and ultimately give uncertainty of how many subdivisions are just right, and be associated with loss of lower frequency information and disputability of arbitrary choices. Therefore, the idea in the poster was not

realized in the manuscript, but the poster is still available at: http://www.pages-osm.org/ysm/643-prizewinning-presentations

Finally, the authors do not mention ring-width at all. I find this rather puzzling. Although I would agree that the inter-annual climate signal in TRW is weaker than MXD and BI, I am not sure this is the case at decadal to centennial scales - especially when replication is high (presumably the authors have 250 RW series). Esper et al. (2012) have hypothesised that there could be millennial scale biases in TRW versus MXD, but that hypothesis was only generated from the N-SCAN data and has not yet been tested using any other data-set yet. So - if the problem ultimately with BI is in the mid-longer time-scales, surely this can be partly tested by comparison to TRW data as well. This seems to be a missed opportunity.

We agree with the referee that the relationship between TRW and MXD as well as BI is very interesting and that RW and Density should co-vary on decadal to centennial scale. But a further evaluation of this kind could be an entirely new paper on its own. It should include methods and discussions of differential standardization (TRW vs. MXD), and also interpretation of comparing different types of climate signals with different spectral properties. Furthermore, it is probably going to be addressed in future work concerning Arjeplog and other sites in Fennoscandia.

Ultimately, this paper should be accepted after appropriate revision. However, the authors need to better rationalise why the raw BI data needs to be transformed to a proxy of density. This seems a needless step in my mind and just makes the whole paper more complex than it needs to be.

CP specific questions 1.Does the paper address relevant scientific questions within the scope of CP?

Yes

2.Does the paper present novel concepts, ideas, tools, or data?

Yes

3.Are substantial conclusions reached?

Yes - but conversation of BI to a proxy of MXD seems an irrelevant step.

4.Are the scientific methods and assumptions valid and clearly outlined?

Mostly - but again - I think conversation of BI to a proxy of MXD seems an irrelevant step.

5.Are the results sufficient to support the interpretations and conclusions?

Focussing on ∆BI and ∆MXD - yes.

6.Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?

No - see my detail comments below.

7.Do the authors give proper credit to related work and clearly indicate their own new/original contribution?

Yes

8.Does the title clearly reflect the contents of the paper?

Yes

9.Does the abstract provide a concise and complete summary?

Yes - but I believe changes will need to be made w.r.t. clarification of how BI can be measured from a grey scale image and why BI data needs to be transformed to a density proxy in the first place.

10.Is the overall presentation well structured and clear?

Yes

11.Is the language fluent and precise?

Mostly - have made some minor successions below.

12.Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?

Mostly - although possibly some confusion between g/cm3 and g/dm3 and the 0 - 255 scale for intensity. See comments below.

13.Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?

See detailed comments below.

14.Are the number and quality of references appropriate?

Yes

15.Is the amount and quality of supplementary material appropriate?

na

Detailed Comments

P. 5228, Line 13: Grey scale? Surely you mean Blue scale?

Changed to "blue intensity" where appropriate

P. 5228, Line 25: change "e.g." to "specifically" and place a "(" before Schweingruber.

Changed

P. 5229, Lines 15-20: Another reason for inverting BI is that one could use the same single series data adaptive detrending methods for both MXD and BI.

True

P. 5230, line 8: Can the authors clarify why they used the mean density of the earlywood and not the minimum density value? Was this option tested?

The minimum density measurements are more affected by systematic bias than the earlywood mean measurement in the Windendro software. Since the min and mean measurements are very similar we chose the less biased.

(When sanding a woodpiece on the radial surface, resinducts are filled with white-reflecting saw dust. The white resinducts have a greater biasing impact on the minimum density/reflectance than on the mean density/reflectance.)

P. 5230, line 17: Reword to, "....are favourable and result in dense/dark..."

changed to "....are favourable and related to dense/dark..."

P. 5230, line 20: How can blue light be measured from a grey scale image? Surely you need to start with a full colour image? Actually, this is a potentially important observation. I had a quick play in Photoshop and once a figure has been converted to grey-scale, NO COLOUR information can be gleaned from that resulting image. Is it possible the authors are not measuring blue intensity, but rather grey intensity?

The RGB image is loaded into Windendro and then RG spectra are filtered out leaving only blue light. The blue light is represented as a greyscale image in the software.

The sentence "Here the blue light that is reflected is captured in grey scale images." is removed.

P. 5230, line 26: Delete "the" before "similar"

Done

P. 5230, line 28: Can "wood in a volume" be better worded?

Not applicable in revised manuscript

P. 5231, line 1: Why proxies for density? Why can't BI be used as a proxy of lignin content. In fact, this is my one major discussion point about the whole of this paper (see general comment above). I do not see any advantage in viewing BI as a proxy of MXD. Why is the transform needed in the first place? Why not simply use BI data as a measure of latewood lignin content which is controlled by summer temperatures.

See general reply above

P. 5231, line 6: w.r.t. Figure 2. I am pretty sure light intensity scales are from 0-255 - not 256.

Figure updated, without 0-256

P. 5231, line 7: w.r.t. " and consequently if a direct comparison between MXD and BImax is going to be made, BImax must also be transformed into density " WHY? Again - I do not see why these two TR parameters cannot be treated independently and the resultant reconstructions from them compared. Comment also relevant for line 21.

See general reply above

P. 5232, line 11: Not sure "finest" is an appropriate word to use. How about best calibrated? or even longest? I always thought Tornestrask was the gold standard - why was this site chronology not used as well?

Not applicable in revised manuscript

We do not have access to Torneträsk earlywood density among the authors and N-Scan is partly sampled in Torneträsk so it does not add more information to the paper.

P. 5232, line 17: Maybe include a reference for N-SCAND in figure 3 caption. Also does not N-SCAND utilise sub-fossil data from other sites as well?

It does contain much more sites, but the recent material used for climate RFA all comes from the three sites in the map. References now included in the Figure 3 caption

P. 5233, line 7: 250 dated samples but only 140 used. Could the authors please clarify why only a sub-set of the samples were used? Was this simply related to cost - if so, you must say it.

Not applicable in revised manuscript

P. 5233, line 26: reword to, " .....X-Ray analysis, WERE sanded with increasingly finer grit sandpaper, with 600-grit paper FOR the final round."

Not applicable in revised manuscript

P. 5234, line 12: See earlier comments. I do not see the need to calibrate BI to MXD. Calibration to IT8 colour card already made (line 6). The BI data should therefore be simply a measure of light reflectance intensity on a 0-255 scale. I have never used WinDendro, so I might be missing something here.

See earlier reply. We were following the standardprotocol for windendro proposed by Campbell et al with some modifications, like inverting the blue color scale, and slope of the relationship between colourtarget and actual measurement to mimic the mean of x-ray density.

Fig 1 in supplementary material is made to illustrate how the adjustment has been done.

change "like with" to "as with".

Not applicable in revised manuscript

P. 5235, line 15: Consider re-wording "response analysis" to "response function analysis". This is the more standard terminology.

Changed

P. 5235, lines 21-23: The 1st difference transform is all well and good, but response function analysis will help partly evaluate the difference in the mid and low frequency domains if non-transformed chronology versions are used as well. At the very least, I would expect analysis using 1st differenced transforms, the RCS version (non transformed) and possibly even the use of so-called standard chronologies where regression, Hugershoff or [stiff] spline functions have been used for detrending.

The suggested climate correlation has been added in supplementary material, see reply above.

P. 5236, line 4: Please define better what you mean by residuals. I understand it is the difference between two chronology series, but with the use of residuals in regression, and so called residual chronologies, the terminology might be a little confusing.

Not applicable in revised manuscript

P. 5236, line 14-16: Why did blue light levels have to be adjusted? This again all comes down to the issue of transforming BI to MXD. I really don't see why this needs to be done and when I see word like "adjustment" in this context, I hear alarm bells. Please clarify.

The blue levels did not have to be adjusted, but they were adjusted so that they could occupy the same part of the scale as the x-ray density and to be compared on the same axis, for pedagogical reasons. The adjustment is not controversial in this context because all measurement means are adjusted, and at the same time the variance is adjusted. It is the same data but on another part of the scale. See reply above.

Figure 1 in supplementary material has now been extended to illustrate how the blue values were adjusted to the x-ray density scale.

P. 5236, line 16 onwards: All chronologies were compared and analysed raw. Is this why 1st differenced transforms were needed for the RFA? If so, why not at least undertake individual series detrending approaches? The authors then say that RCS is used on the MXD data but not on the BI data. Again - this is all rather confusing. Are the chronologies in Figure 10 raw means and not RCS detrended versions. This all needs further clarification.

We use first differences here because it is simple and because it is good practice. Detrending is now included to evaluate BI climate proxy performance.

P. 5236, line 26: There is one possible problem with Figure 5 and that is the images are presented using the full light spectrum. The methods is ONLY interested in the blue light reflectance, so why not filter the images in figure 5 to show only the blue part of the spectrum. My gut feeling is that the colour differences seen with the full spectrum colour figure will change considerably when filtered to blue only.

Figure 5 now figure 1, is now changed so that only blue light spectrum is represented. The blue light is represented in a grey scale image

P. 5237, line 13: I am not sure what the authors mean by a "rational climate signal". Please re-word.

We mean a signal that can be explained in a mechanistic way, "Rational" changed to "plausible"

P. 5237, line 18: It looks to me as if it is ONLY MXD which shows an EPS < 0.85 around 1600. BI is fine.

"in the MXD data" is added after "around 1600 A.D."

P. 5237, line 23: Not fully clear. Can the authors clarify how the data are generated for figure 7. The EW and LW data are summed? Is that correct? Is that why the BI y-axis is not in the 0-255 range? Sorry, but this is all a little confusing. Please clarify. Why the * after BLI?

The figure 7 now figure 5 in supplementary material is changed and clarified. The scale was modified in compliance with themethodology in figure 1 supplementary material.

*The * after BI is removed and replaced with ACT (adjusted colour target), at first it was intended to explain that the scale was adjusted but unfortunately this was later forgotten.*

P. 5238, line 9: All the calibration r2 values are using 1st differenced transforms - right? Can this be clarified in the caption. Ultimately, I think all the BI, MXD and their Δ versions should be detrended in more standard data adaptive ways to ascertain the mid-frequency response.

Added words: "With first difference transforms, the JJA"

Will be clarified also in the caption

Mid frequency response is presented in Supplementary material (Figs. 2 and 3)

P. 5238, line 25: "point" should be "points"

Changed

P. 5238, line 27: insert "the" before "same".

Changed

P. 5239, line 1-6: Would Figure 10 be relevant if the BI data were not calibrated to density values???? Ultimately, I am struggling with this figure as I do not see why the BI data need to be calibrated and assuming that the statistical properties will be exactly similar to MXD is wrong.

Figure is removed, not applicable in revised manuscript

P. 5239, line 11-17: Is there a chance that there was some timber extraction from this region between the 16th and 17th centuries that could also be partly to blame for the low replication for this period?

This is a very good point, I, (j.björklund), recently took a course in Fennoscandian Boreal forest history, and used my newfound knowledge on this material and I think that it is probably a very likely explanation. If it happened it likely occurred in the 19th century. I did however, not find any stumps from high-grading activities, but this

does not mean that they were absent.

Paragraph removed and not applicable in revised manuscript

P. 5239, line 19: Replace "negative" with "irrelevant"

Not applicable in revised manuscript

P. 5241, line 20-23: It is not clear if the CRA is undertaken using data transformed to 1st differences. If so, the authors cannot really talk about climate response at time-scales longer than year-to-year.

The analysis is based on the RFA made on first differenced data.

Captions will be clarified regarding first differences.

The statement about longer timescales is based on time series that share most of their information. If time series that are compared share most information, they are likely to be similar in all frequencies. We were making an analogy between the likeness between JJA and AMJJA temperatures and that these would be very similar in all frequencies and probably more so in lower frequencies.

MXD and DMXD also share most information since the latewood measurement is very dominating over the earlywood measurement and thus the similarities between the time series would also be very similar and also here even more so in lower frequencies. There is no implication here of the climate response on longer time scales. Only arguments that, MXD and ΔMXD should have similar overall trends.

P. 5242, line 1: Replace "boosted" with "improved"

Not applicable in revised manuscript

P. 5242, line 3: section sub-heading title (use of "complement") is not really consistent with the paper title which uses "replacement". Consider changing to "replacement".

Not applicable in revised manuscript

Note also that the paper title is changed to better reflect the revised manuscript

Please note that much of the text of section 4.2 is a rather rambling affair and overall needs tightening. Focusing just on BI data, without transforming them to density, might simplify much of the paper in this regard.

Section 4.2 is mostly rewritten, and hopefully improved.

P. 5244, line 15: reword, " .....to drastically improve THE spatial distribution and replication in highly climate sensitive tree-ring chronologies and lead to higher confidence in LARGE-SCALE climate reconstructions "

Not applicable in revised manuscript

Comments on Tables and Figures

Overall, captions could have much more detail.

Table 1: Please clarify if these results are from using 1st differenced versions of the data. If so, it would be interesting to see results using detrended data - STD and/or RCS.
Added sentence: All chronology- and temperature-data are first difference transformed. See Supplementary material.

Figure 2: change "over" to "of". Intensity scale should be 0 -255.

Changed

Figure 3: I think N-SCAND covers a greater region??

Sentence added: "In N-Scan there are more sites in Finland that also include subfossil material, not shown."

Figure 4: Throughout the paper the authors change between g/cm3 and g/dm3. In A, I think it should be g/dm3. Why a ** after BLI? Why are the intensity scales ranging from 0 - 1200 and not 0 - 255. Having never worked with WinDendro, this seems a little unclear to me. Surely the calibration with the IT8 card allows the intensity data to be on a 0 -255 scale?

*Figure 4 caption slightly altered. All * are removed. And g/dm3 is used consistently. See Figure 1, supplementary material for explanation on scale of blue intensities.*

Figure 5: full visual light/colour is perhaps not relevant??? Why not filter to show blue part of the spectrum.

Figure 5 changed to figure 1, showing only blue spectrum colour, however it is represented as a grey scale image.

Figure 6: State that these are raw non-detrended chronologies.

Now stated

Figure 7: Not fully clear why Y-axis scale is not from 0 - 255 - maybe because the ED and LD values have been summed??

A more thorough caption has been written. See also the new fig 1 supplementary material.

Figure 8: These are RFA results from using 1st differenced transforms. If so, state this in caption, but also consider including results from detrended data (STD and/or RCS)

More explanation in the caption has been added. See also Supplementary material for climate correlations with detrended data.

Figure 9: These are RCS chronologies - correct? Please state this in the caption.

Yes they were, now they are raw. In Figs. 2 and 3 in Supplementary material standardization is applied.

Figure 10: Please clarify if these are raw non-detrended chronologies.

Now stated

Figure 11: This again all comes down again to the calibration of BI to MXD - if this is not needed, then figure 11 is not necessary I think.

Figure removed.

Sincerely Björklund et al.