

## ***Interactive comment on “Blue Intensity based experiments for reconstructing North Pacific temperatures along the Gulf of Alaska” by Rob Wilson et al.***

**Rob Wilson et al.**

rjsw@st-andrews.ac.uk

Received and published: 1 June 2017

WE VERY MUCH APPRECIATE THE DETAILED CRITIQUES FROM BOTH REVIEWERS AND WE HOPE WE HAVE ADDRESSED ALL THEIR ISSUES TO BOTH THEIRS AND THE EDITOR'S SATISFACTION. I SHOULD STATE THAT I HAVE ALREADY MADE CHANGES TO THE MANUSCRIPT AND AM HAPPY TO SUBMIT THE EDITED VERSION IF THE EDITOR AGREES WITH THE CHANGES WE HAVE MADE.

Review 2

Specific comments:

C1

Considering the experimental nature of the LWB and particularly DB parameters, it would have been useful to develop even a limited MXD dataset on at least part of the samples (e.g. from one site) in order to enable a direct comparison of the lower frequency trends in the BI data. Although it is argued that the structure of mountain hemlock wood makes it more difficult to prepare and measure these samples for density, it is not impossible and has been done in past studies. This would have been helpful in evaluating and constraining the utility of differently detrended BI chronologies and therefore considerably benefited this study in further strengthening the case for the use of DB as a better, less biased parameter relative to LWB and a suitable alternative to MXD for this species, especially since this is the first study to measure BI on mountain hemlock samples. Was this option at all considered? MXD HAS NOT BEEN MEASURED ON MOUNTAIN HEMLOCK TREES IN THIS REGION AS FAR AS WE ARE AWARE. ALTHOUGH WE AGREE THEORETICALLY WITH THE REVIEWER'S COMMENT THAT THIS COMPARISON COULD BE USEFUL, MXD WAS NEVER FACTORED INTO THE RESEARCH DESIGN OR BUDGET AND THE FOCUS OF THE STUDY WAS ALWAYS ON BI AND RELATED PARAMETERS. WE (AS A COMMUNITY) SHOULD ALSO BE CAREFUL NOT TO ASSUME THAT MXD IS THE "TRUTH" SO HAVE ASSESSED THE DATA AS BEST AS POSSIBLE USING INSTRUMENTAL DATA AND OTHER PROXY ARCHIVES FOR THE REGION.

I am somewhat surprised that a higher number of samples was not used for the individual sites. According to Table 1 replication should ideally be 12-28 series for LWB and 14-36 for DB depending on the site. In several cases, the actual maximum number of series used is below (and in some cases well below) this optimal level. Is this not a problem? The weaker signal strength of BI data and the need for higher replication in order to develop 'robust' chronologies is acknowledged (e.g. L368-369). Also, the relatively low replication may even affect the RW data as stated in L229-230. As stated in the text, a subset of samples was selected for this study from earlier work so why not aim for 25-30 samples per site? That would have at least reduced uncertainty about the representativeness of some of the BI site chronologies, particularly in ear-

C2

lier periods when replication is likely even lower. It would be nice to see a replication plot over time (and EPS plot) for separate sites as well as for the 'all series' pooled version (perhaps as an SI figure) to give a better indication of which periods might potentially be affected by low replication. THIS IS AN EXPLORATIVE PAPER AND WE HAVE BEEN CLEAR, AS THE REVIEWER STATES, THAT THE NUMBER OF RECORDS ARE LOWER THAN WOULD BE IDEAL. DATA DEVELOPMENT IS ONGOING (DATA BEING ADDED CONTINUALLY). HOWEVER, WE WANTED TO WRITE THIS INITIAL EXPLORATIVE PAPER TO PARTLY INFORM OURSELVES AS TO THE CONTINUED STRATEGY OF THIS WORK AS WELL AS THE WIDER COMMUNITY. PREVIOUS RW BASED CALIBRATIONS (WILSON ET AL 2007; WILES ET AL. 2014) SHOW THAT WHEN REPLICATION IS HIGH, WE CAN EXPLAIN AROUND 40% OF THE TEMPERATURE VARIANCE. USING THIS NON-IDEAL DATA-SET WE ONLY EXPLAIN 27% FOR RW, BUT >40% USING LWB OR DB. THIS SUGGESTS THAT THE RESULTS CAN ONLY IMPROVE AS WE INCREASE REPLICATION. WE ARE HAPPY TO MAKE A STRONGER SPECIFIC COMMENT IN THIS REGARD.

L96-98 and L217-219 – What is the rationale for this statement? To my knowledge this issue has not been investigated in any previous study. Presumably a higher correlation between EWB and LWB would imply that EWB expresses in part the same information as LWB, but does that necessarily mean that this information is related to climate? How do you define 'weakly correlated'? Or in other words, what correlation would be acceptable and what would not? Ideally, this statement could be supported with an example and actual data. If nothing else, I would suggest elaborating further on this statement to more clearly express the justification for this claim. THIS WAS BROUGHT UP BY REVIEW 1 AND WE AGREE THE CURRENT WORDING IS AMBIGUOUS AND NEEDS CLARIFICATION. BASICALLY WE WANT TO PROVIDE A THEORETICAL BASIS THAT DB WILL LIKELY BE USEFUL WHEN EWB AND LWB ARE UNCORRELATED AT HIGH FREQUENCIES.

L102-105 As a general comment, some of the limitations of BI (specifically LWB) have

C3

already been explored in other studies. Clearly DB is a major improvement, although I wonder just how well DB resolves these issues and specifically whether DB could still have some problems at lower or other frequencies. It is interesting that in some cases the calculation of DB weakens the common signal, suggesting that information which should ideally be preserved is to some extent being removed in the process, yet in other instances the strength of the common signal remains relatively unaffected or even shows improvement. I suppose these questions can only be answered by various future studies that will further explore DB and I would not expect this to be fully covered here. But perhaps a statement could be included somewhere to caution and emphasize that considerable uncertainty remains with respect to the performance of DB and more work is required in this area. THIS IS A VALID COMMENT AND WE SHOULD EMPHASISE BETTER THAT IN FACT THE YEAR-TO-YEAR SIGNAL IN THE LWB DATA IS OFTEN STRONGER THAN THE DB DATA. ALTHOUGH NOT RELEVANT FOR THIS SPECIFIC PAPER, A SIMILAR BAND-PASS APPROACH AS UTILISED BY RYDVAL ET AL. (2016 AND 2017) COULD WORK VERY WELL IN THE GOA REGION. WE CAN ADD MORE DISCUSSION IN THIS REGARD BUT THIS WILL BE THE FOCUS FOR A LATER PAPER.

L109-111 - Is there any indication to what degree early instrumental biases could be a limitation (if at all) in achieving the stated aim? THIS IS AN IMPORTANT POINT AND WE ARE HAPPY TO ADD FURTHER DISCUSSION ABOUT THE ISSUES OF USING EARLY INSTRUMENTAL DATA. THE MAJOR ISSUE HERE IS THE LARGE DIFFERENCE BETWEEN LWB (WARM LIA) AND DB (COOL LIA) WHICH CANNOT BE ADDRESSED USING THE LONGER INSTRUMENTAL DATA. WE MUST RELY ON OTHER PROXY RECORDS FROM THE REGION (I.E. GLACIAL RECORD) WHICH SUGGEST A LIA ABOUT 1 DEGREE COOLER THAN THE 20TH CENTURY WHICH FIT WELL WITH THE DB AND RW DATA.

(L133-138) Is there actually any need to detrend DB series? What is the justification for this? Hypothetically, if both LWB and EWB contain the same ontogenetic trends

C4

then by the nature of the DB calculation this trend would be automatically removed. I do not know whether or not that is true. This may be a more complex issue - perhaps only the LWB contains this trend or the LWB and EWB trends related to age differ in some way. But is it not possible that by detrending the DB data some lower frequency climatic information may be unnecessarily removed? Was the development of DB chronologies without performing detrending considered or explored in the analysis? The DB chronology in Figure 5 actually looks like a reasonable chronology variant and so I wonder how non-detrended DB chronologies would perform in terms of calibration and validation statistics relative to the detrended versions. THIS IS A VALID POINT. THEORETICALLY, ONE WOULD ASSUME THAT DB WOULD BE SIMILAR TO MXD IN TREND, BUT I DON'T THINK THE EXPERIMENTATION WITH BI BASED PARAMETERS ARE ADVANCED ENOUGH TO ADDRESS THIS. I DON'T THINK WE CAN ASSUME THAT EWB AND LWB NECESSARILY WILL SHOW THE SAME ONTOGENETIC TREND. APPENDIX FIGURE 4 CLEARLY SHOWS A MORE COMPLEX MEAN AGE RELATED CURVE FOR DB THAN LWB, BUT THE LATTER WILL BE IMPACTED BY THE HW-SW COLOUR CHANGES, SO IT IS DIFFICULT TO JUDGE THIS SPECIFICALLY. TO PARTLY ASSESS THIS, AN AGE DEPENDENT SPLINE IS NOW ALSO USED IN THE ANALYSIS PRESENTED IN FIGURE 6 AND 7. SEE COMMENTS FOR REVIEWER 1.

L146-147 – This is a fairly short validation period. Why not choose an equally long calibration and validation period which has been a common approach in other similar studies? How sensitive are the results to this choice? THE JUSTIFICATION FOR THESE PERIODS WAS NOT MADE AND THIS SHOULD BE CLARIFIED. AS MUCH OF THE TR DATA IN ALASKA EXPERIENCE CALIBRATION AND VALIDATION ISSUES, SPECIFICALLY IN THE RECENT PERIOD (SO CALLED DIVERGENCE), THIS APPROACH WAS USED FOR THE INITIAL CALIBRATION/VALIDATION TESTS. THERE WILL CERTAINLY BE SOME SENSITIVITY TO THE PERIODS USED, BUT IT SHOULD NOT BE FORGOTTEN THAT THE FULL PERIOD CALIBRATION (1901-2010) WAS VALIDATED AGAINST THE 19TH CENTURY DATA (FIGURE 7), SO MUL-

C5

TIPLE PERIODS HAVE BEEN USED.

L244-245 acknowledges that this may be an issue. For example, would a different period affect the significance of any results in Table 4? WE BELIEVE THAT CONSISTENCY OF THE METHOD FOR TESTING EACH OF THE PARAMETERS AND THEIR COMBINATIONS IS MORE IMPORTANT THAN USING DIFFERENT PERIODS PER SE. THIS IS A VALID POINT HOWEVER AND CERTAINLY THE RESULTS WILL CHANGE SOMEWHAT IF DIFFERENT PERIODS WERE USED. HOWEVER, THIS WAS NOT THE AIM OF THE PAPER AND THE RESULTS OF TABLE 4 AND 5 MUST BE POOLED TOGETHER TO DERIVE AN OBJECTIVE ASSESSMENT OF THE VARYING STRENGTHS AND WEAKNESSES OF THESE DIFFERENT PARAMETERS.

L147-148 and top panel in Figure 8 - Why not perform a nested PC reconstruction? By excluding even just one or two of the shortest sites this reconstruction could be extended to the mid or early 18th century, providing more information for comparison with the other reconstructions. IT IS NOT THE SPECIFIC AIM, AT THIS TIME, TO DERIVE A NEW DEFINITIVE RECONSTRUCTION SPECIFICALLY. THIS PAPER DETAILS MULTIPLE EXPERIMENTS TO EXPLORE THE UTILITY OF THESE DIFFERENT PARAMETERS FOR CLIMATE RECONSTRUCTION AND TO HIGHLIGHT POTENTIAL PARAMETER SPECIFIC BIASES. IT IS NOT CLEAR WHY A NESTING APPROACH WOULD CHANGE THE CONCLUSIONS ALREADY DISCUSSED IN THE PAPER.

164-173 - Was the variance stabilized to account for changing replication /  $\bar{r}$  through time? Even just a look at the changing variance with replication in Figure S4 in the supplementary material suggests that this should be performed. Either I have missed this or it is not stated in the methods section. YES – SORRY – VARIANCE STABILISATION WAS PERFORMED AND THIS SHOULD BE MENTIONED IN THE METHODOLOGY.

L171 – Figure S4 suggests that linear detrending may not be the most appropriate

C6

choice to detrend for example the DB data. Based on the initial increase in the juvenile period of growth in the DB results, perhaps a somewhat more flexible detrending alternative could be explored to account for this initial increase? Additionally, Figure S4 raises some intriguing questions about non climatic trends in the BI data. I suggest that a fourth panel showing results for EWB should also be shown here. It appears that the initial juvenile trend is present in the EWB only. Does this not suggest that the EWB should not be used to 'correct' LWB in this initial 30-50 year period?! Because it appears that this initial trend is not present in the LWB data, but is introduced into DB by the EWB data making it necessary to then remove this trend again from the DB series. Due to the experimental nature of this TR variable, several methodological considerations such as the one discussed here remain unaddressed and have not been explored in this study. Considering the nature and aims of this manuscript, I would not expect a detailed evaluation of DB, however, methodological issues such as this and the need to explore them further should at least be acknowledged and more clearly highlighted in the text. ATTACHED AN EXTENDED FIGURE S4 WHICH NOW INCLUDES THE EWB AGE ALIGNED CURVE. THE REVIEWER IS CORRECT THAT THERE REMAINS MUCH AMBIGUITY W.R.T. TRENDS AS A FUNCTION OF AGE. WE AGREE THAT LINEAR FUNCTIONS MAY NOT REPRESENT THE BEST "FIT" FOR DETRENDING AND REVIEWER 1 MENTIONED THIS ISSUE AS WELL. WE USED THE CURRENT CHOICES AS THEY ARE THE COMMON APPROACH IN THE LITERATURE. HOWEVER, WE HAVE NOW ADDED AN AGE DEPENDENT SPLINE OPTION INTO THE MIX. SEE OTHER UPDATED FIGURES W.R.T. REVIEWER 1.

Is it reasonable to develop a RW + DB chronology / reconstruction considering the difference in the seasonal response of RW and BI data and the acknowledgement that RW may not be primarily controlled by summer temperatures in this region (L349-351 and L379-380)? PLEASE SEE THE RESULTS FROM THE 1ST DIFFERENCED CORRELATION RESPONSE FUNCTION ANALYSIS RESULTS. YES – WE BELIEVE THAT THESE PARAMETERS CAN BE COMBINED ALTHOUGH PLEASE NOTE THAT THE RECONSTRUCTION PRESENTED IN FIGURE 7 ARE SINGLE PARAMETER

C7

ONLY.

L187-189 – Maybe already refer to Figure 2 at the end of the first sentence. THIS SECTION FOCUSSES ONLY ON THE COMMON SIGNAL AND NOT THE CLIMATE SIGNAL EXPRESSED BY IT. FIGURE 2 IS RELEVANT FOR THE NEXT SECTION.

Perhaps specifically state somewhere in section 3.3 that JJAS was selected for further analysis for BI (presumably because this is the optimal season). YES – THIS CAN BE DONE.

L196-198 – What might be the reason for such a broad seasonal response in the RW data in this region? Has this been discussed in any of the previous studies (e.g. Wilson et al. 2007 or Wiles et al. 2014)? YES – THIS WAS DISCUSSED IN PREVIOUS PAPERS AND IS NOT RELEVANT TO THE CURRENT PAPER. PLEASE NB THE 1ST DIFFERENCED CRFA RESULTS WHICH SHOW THAT THIS EXTENDED SEASON IS STILL SEEN FOR THE FEB-AUGUST SEASON.

L216 – Just an observation, it is a little surprising that RW agrees more strongly with DB and LWB than DB does with LWB. THE PRIMARY AUTHOR HAS NOTED THIS FOR MULTIPLE SPECIES/LOCATIONS THAT DB OFTEN SHOWS SIMILARITIES TO RW, BUT WITH LOWER 1ST ORDER AUTOCORRELATION. REMEMBER THAT ALL OF THESE PARAMETERS SHOULD BE EXPRESSIONS OF SUMMER TEMPERATURES SO THEY SHOULD SHOW SOME SIMILARITY. THIS ISSUE NEEDS SPECIFIC ATTENTION AT SOME LATER STAGE BUT IS BEYOND THE PRIMARY AIM OF THIS CURRENT PAPER.

L219-220 - Maybe worth clarifying here that EWB would not be expected to contain an actual climate signal in the first place. I DO NOT THINK WE KNOW THAT SPECIFICALLY AND WORK BY REVIEWER 1 FOCUSING ON THE NH SCHWEINGRUBER WORK CLEARLY SHOWS A RELATIONSHIP BETWEEN MINIMUM EARLYWOOD DENSITY AND SPRING TEMPERATURES. IT IS HOWEVER, NOT AS STRONG AS MXD FOR LATE SUMMER TEMPERATURES.

C8

L241-242 – Is it possible that the failed validation could be related to the quality of instrumental data for this early period? THAT IS INDEED LIKELY AND WE CAN EXPAND ON THE DISCUSSION IN THIS REGARD – BUT SEE PREVIOUS COMMENTS.

L278-280 - This may simply indicate that the use of LWB and EWB to calculate DB is an imperfect procedure. It would be necessary to look at data from other locations (and also other species) to identify whether the DB trends in Figure S4 actually reflect inherent properties related to age (and should therefore be detrended) or are related to other factors. This is an important issue and it is unfortunate that this paper does not or cannot investigate this type of issue in more detail. But perhaps at least a bit more discussion could be included in relation to this? WE LIKE TO THINK THAT THIS PAPER HAS EXPLORED THIS ISSUE. THERE IS A CLEAR BIAS IN THE LWB DATA WHICH APPEARS PARTLY (COMPLETELY?) MINIMISED IN THE USE OF THE DB VARIANT. YES – CLEARLY THIS NEEDS MORE TESTING, BUT AT THE SAME TIME, THE RESULTS MAY BECOME LESS AMBIGUOUS AS MORE DATA ARE ADDED.

L294-295 - The results in Figure S4 already indicate that a linear detrending curve can lead to a serious underestimation in the early parts of the series so the poor performance of LINsf is not surprising. YES – BUT THE LINRES IS FINE, SO I THINK WE ARE SEEING AN ISSUE OF THE SF APPROACH RATHER THAN MIS-FITTING OF A DETRENDING FUNCTION PER SE.

L314-317 - But as discussed in the text (and in relation to the results in lines 286-291) it is apparent that LWB is inherently biased. So why even consider it as a feasible option? BECAUSE THIS IS A PAPER EXPLORING THE STRENGTHS AND LIMITATION OF THESE PARAMETERS. SURELY WE NEED TO SHOW THAT THESE DATA ARE BIASED IN THE LOW FREQUENCY DOMAIN. THE HIGH FREQUENCY SIGNAL IS HOWEVER ARGUABLY BETTER THAN DB AND DON'T FORGET THAT LWB DOES IN FACT PASS MOST VALIDATION TESTS SO THIS IS A TRICKY ISSUE TO ASSESS AND OTHER PROXY TEMPERATURE ARCHIVES ARE FUNDAMENTAL IN

C9

#### FURTHER ASSESSMENT OF THE PRE-1850 TRENDS.

L330 – Maybe include '(Figure 6 and 7)' in the bracket since this reconstruction appears in both figures and is discussed in relation to Figure 7 in the previous paragraph. YES – CAN EASILY ADD THIS IN.

L329-331 – It may be worth stating here something along the lines that the 'best performing' PC and extended reconstructions are shown here and compared with Wiles and the glacial advance record - i.e. state the reason why these reconstructions are shown in Figure 8. YES – CAN EASILY ADD THIS IN.

L358 – Or for species which do not have this colour difference to begin with. YES – CAN EASILY ADD THIS IN.

Figure 3 – It is interesting that LWB calibrates and validates more strongly than DB in terms of the strength of the relationship. Why might this be the case? Could this difference be related to replication? NO – REPLICATION IS THE SAME – THIS APPEARS TO BE RELATED TO THE FACT THAT DB PORTRAYS SUMMERS TEMPERATURES SLIGHTLY MORE WEAKLY THAN LWB. EWB HAS A WEAK CLIMATE SIGNAL AND MAY IMPACT DB WHEN IT IS CALCULATED. THIS CAN BE CLARIFIED FURTHER IN THE PAPER AND DOES LEAD TO THE POSSIBILITY THAT THE BAND-PASS APPROACH TO CALIBRATION USED BY RYDVAL ET AL. (2016/17) COULD BE A REALISTIC APPROACH TO DENDROCLIMATIC RECONSTRUCTION IN THIS REGION. HOWEVER, THIS WAS NOT THE FOCUS OF THIS PAPER WHICH WANTED TO HIGHLIGHT THE STRENGTH AND WEAKNESSES OF THE DIFFERENT PARAMETERS. WE CAN ADD MORE DISCUSSION ON THIS ISSUE.

Figure 6 - It is somewhat difficult to identify the trend of the LINres curve for LWB and especially DB - is it possible to improve the visibility of these curves? Also, in the figure caption (L596-597) maybe consider changing 'low plots' to 'lower set of plots'. AS MANY OF THE TIME-SERIES ARE VERY SIMILAR, IT IS REALLY NOT POSSIBLE TO ADDRESS THE FIRST POINT IN ANY MEANINGFUL WELL. HOWEVER, WE

C10

ARE HAPPY TO EDIT THE CAPTION ACCORDINGLY.

Figure 7 – Why not also show r2 rather than r for the 1850-1900 period? WE CAN DO THIS IF REQUIRED BUT IT DOES NOT CHANGE ANYTHING W.R.T. INTERPRETATION.

Table 1 - Is there any real meaning in including N-EPS information for EWB data? Presumably these data do not (or should not) contain any common climatic signal and there would be no point in developing a chronology from these data that would be of much use. WHY WOULD EWB NOT PORTRAY A COMMON SIGNAL? WE DISAGREE WITH THIS COMMENT.

Table 2 – Minor detail, but why not arrange the site order from west to east? Table 3 – Why is EWB positively (though weakly) correlated with DB? COMPLETELY AGREE. THE ORDER CAN BE EASILY CHANGES TO A MORE LOGICAL EAST-WEST ORDER AS ALREADY PRESENTED IN TABLE 3.

Technical corrections: L32 – affecting instead of effecting AGREED – THIS CAN BE CORRECTED WHEN “ALLOWED” TO EDIT THE PAPER FOR FINAL PUBLICATION

L71 – ‘as they are a measure of ...’ instead of ‘as they measure’ may be a more accurate statement. AGREED – THIS CAN BE TWEAKED WHEN “ALLOWED” TO EDIT THE PAPER FOR FINAL PUBLICATION

L109 – reconstructions instead of reconstruction AGREED – THIS CAN BE CORRECTED WHEN “ALLOWED” TO EDIT THE PAPER FOR FINAL PUBLICATION

L126 – Please specify the exact calibration target type as there are different versions. IT8.7/1 is a transparency target whereas IT8.7/2 is a reflective target. I assume that the latter was used. YES - IT8.7/2 – TEXT CAN EASILY BE EDITED.

L211-212 – Consider rewording ‘potentially optimal’ to something along the lines of ‘more optimal compared to a PC approach’. WE PREFER THE CURRENT WORDING AS A PC APPROACH WOULD NOT BE POSSIBLE WITH A SUB-FOSSIL EXTEN-

C11

SION FROM ONE LOCATION.

L218 – Consider specifically stating that the correlation between EWB and LWB is not significant. IN ACTUAL FACT THE CORRELATION IS SIGNIFICANT – BUT WEAK. CURRENT WORDING IS PREFERRED.

Also, ‘of’ missing in ‘the utilization DB to’. REVIEWER 1 ALREADY FLAGGED THIS SENTENCE AS BEING CLUNKY. WE CAN REVISE.

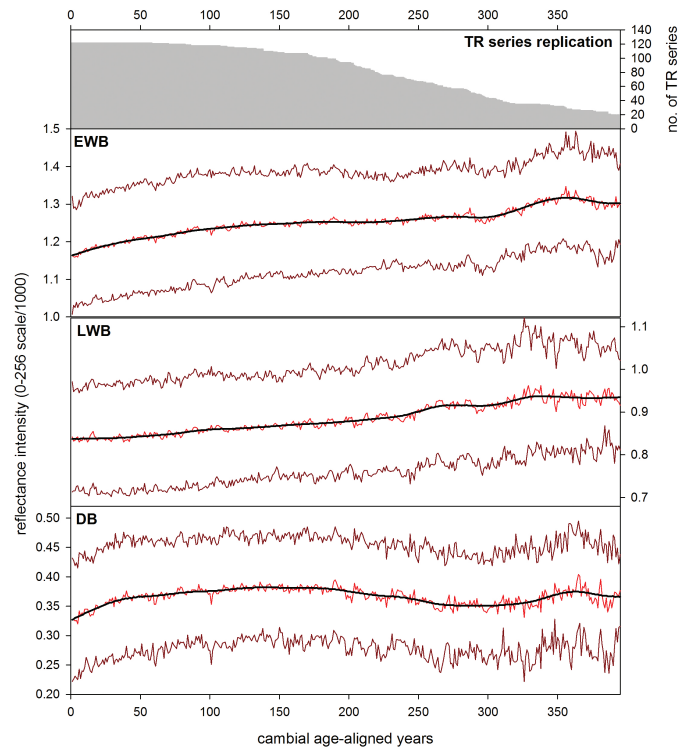
L288 – change ‘particular’ to ‘particularly’ AGREED – THIS CAN BE CORRECTED WHEN “ALLOWED” TO EDIT THE PAPER FOR FINAL PUBLICATION

L637 – Is there a better word than ‘dominant’ which could be used here? CAN BE CHANGED TO “CALIBRATION EXPERIMENTS FOR THE FOUR STRONGEST SEASONS...”

---

Interactive comment on Clim. Past Discuss., <https://doi.org/10.5194/cp-2017-26>, 2017.

C12



**Fig. 1.**