Reply on referee comments on Wang et al., Chemical de-staining and the delta correction for blue intensity measurements of stained lake subfossil trees

We thank editor's and referees' comments on this manuscript. The point-by-point responses are listed below.

Reply to referee1:

We greatly appreciate the comments from Dr. Miloš Rydval and his kind line-by-line suggestions to improve the English expressions. Our responses are listed below:

General comments:
The manuscript addresses a highly topical issue related to the dendroclimatic utilization of reflected light (blue intensity) from lake subfossil wood material, specifically a discoloration bias related to staining, which is primarily attributed to Fe oxidation. The study investigates the application of a range of chemical treatment techniques in order to improve light reflectance data by reducing staining bias and also provides additional validation for the applicability of the delta BI correction procedure in the context of using samples affected by staining.

Overall, it is a nice, relevant and focused paper building on previous work on this topic. The study is designed and performed in a methodical manner and the manuscript is generally well organized and logically structured. Although I do not have any major comments on the methods used or analysis performed, the (mostly minor) comments detailed below will hopefully help to further refine this work. I would also recommend checking the manuscript to make minor language improvements (for example in order to clarify the meaning of certain statements) and I include specific suggestions (under minor comments) to indicate parts of the text where I believe most improvements could be made.

Specific comments:

L83: what is meant by 'new lake'? as in 'newly sampled lake' (i.e. data from samples from this lake were not analyzed / published before)? – please clarify

Response: We corrected "new lake" to "newly sampled lake".

L106: Why were the treatment times different for MixC? Please explain briefly.

Response: MixC was the last, but the most efficient and active reagent we used to treat samples. We had known that reactions of other (less active) reagents would terminate in 12–24hrs before using the MixC reagent. We thus used a shorter reaction time (24hrs) for MixC treatments, while we sampled the solutions of MixC more intensively (at 0.5hrs after treatments; Figure S2) in order to capture the change point of reaction (although finally this figure was only placed in the supplementary material). We added a sentence to explain the different time settings "Treatment and sampling times were set shorter for MixC since it was the most active reagent." in Lines 106–107 in the revised manuscript.

L117-118: Please add relevant reference(s) here.

Response: We added reference to (Rydval et al., 2014).

L135-138: Please include information about the measurement resolution of the photo sensor used (i.e. the step size along the density measurement profile) – e.g. 10µm or variable?

Response: The resolution of the density profile is 10µm. This information was added at Line138.

L147-148: In what way were they affected – structurally, their color? If this is true for the unstained samples, could this play a role in affecting the properties of stained samples to some degree as well? Undoubtedly, the treatments lead to improvement, but it would also be worth discussing if there are (or could be) some undesirable effects as well that may perhaps limit the observed improvement?

Response: The major effects are related to colors. As we can see from the unstained+living tree group on the right panel of Figure S7 (in the revised supplementary material), raw LBI, DBI, and MXD measurements were similar for MixA, MixB, and Control treatments, while MixC altered the LBI and DBI but not the MXD measurements. We think the bleaching effect of MixC is weak on the subfossil wood since DBI of stained LSTs did not diverge while DBI of living trees and unstained trees diverged. We added a few sentences at the end of the Result section 3.1 (Line 184–187) to discuss this effect. "MixC, the only reagent with bleaching function (Table 1), could have an additional bleaching effect on the wood, resulting in smaller LBI and DBI values in living and unstained trees compared to untreated control (Fig. S7). However, DBI of the stained LSTs was only slightly modified by the MixC treatment (Fig. S7), indicating that the bleaching effect of MixC is weak for the stained samples."

L150: Was there any apparent difference in the sapwood / heartwood of the subfossil samples and if so, could that then potentially also have some effect on the results?

Response: The sapwood portion of subfossil samples is generally decayed due to their long stay in water/sediments. We thus did not observe such apparent sapwood-heartwood difference on the subfossil samples. In addition, black spruce in general does not show apparent color difference between the sapwood and heartwood. Living-tree samples of L20 may represent a special case where this color issue occurred.

L152-153: Just to clarify - 'regional chronologies' here represents different parameters / treatment methods and not the two sites (since those were pooled together)? A slight re-wording might help to ensure that this is clear.

Response: We changed the sentence to "Regional chronologies for each tree-ring parameter (LBI, DBI, and MXD) and treatment (MixA, MixB, MixC, and Control) were generated by pooling standardized series from both sites using the Tukey's bi-weight robust mean." in Lines 153–154.

L263: Slight reformulation is required here since the higher replication is needed to obtain a robust and representative (DBI) chronology rather than a chronology with an equivalent (or similarly strong) climatic signal to MXD (although it is true that these two things usually go hand in hand).

Response: We changed the sentence to "Firstly, a higher tree replication is often needed for DBI than MXD data, in order to obtain a robust chronology (Rydval et al., 2014; Wilson et al., 2019)." in Lines 270–271.

L264: At the same time, DBI appears to calibrate more weakly compared to LWB over the instrumental period. Could you suggest a possible explanation for this? Could you also add some moving window EPS statistics somewhere (e.g. in Figure 6) for the final chronologies used for the reconstruction in order to get a better idea which parts of the chronologies might be stronger / weaker.

Response: We added the moving EPS to Figure 6e in the revised manuscript (See Figure in the supplementary reply letter). We also discussed the more variable EPS of DBI in contrast to LBI and MXD at Line 273.

L278-279: I think this sentence could be reformulated somewhat and expressed more clearly.

Response: We changed the sentence to "The chemical de-staining experiments, though not satisfactory regarding the robustness of LBI data, suggest that the post-sampling chemical Fe oxidation most likely result in the staining issue." in Lines 286–287.

Table 2: It would be helpful to clarify a few things. To avoid any possible misinterpretation, it would help to specify in the caption that RGB refers to separate (R, G, and B) color channels rather than for example a full color RGB light values. Also, some more details should be included in relation to no. 3 and 5 to clarify the difference between these two terms – presumably delta RGB refers to all of the color channels whereas DBI only refers to the blue channel? Or is there some difference when it comes to how delta B (in RGB) was calculated compared to DBI? Unless I am mistaken, the settings for calculating DBI (i.e. how LBI and EBI are determined) in CooRecorder are adjustable – the main settings for this calculation should probably also be stated here.

Response: We clarified each parameter in the Table 2 and its caption (see supplementary reply letter). In short, the main difference between No. 1–3 and No. 4–6 is that No. 1–3 provide one R, G, and B value for each measured lath since all tree-ring measurements of each type (R, G, and B) were averaged by lath. Another difference between No.3 and No.5 is that No.3 is based on differencing R, G, and B intensities measured from 100% of pixels in latewood and earlywood, while DBI, i.e. No.5, is based on differencing LBI (measured from 30% of darkest pixels) and earlywood BI (measured from 100% of the pixels).

Figure 1 caption: 'The gray shades in (a) correspond to . . .' / 'The gray shading in (a) corresponds to . . .'. Also, maybe simply use the term 'replication' instead of 'distribution range' here?

Response: The shading in a) shows the range distribution of black spruce, rather than the replication of trees. We corrected "shade" to "shading" and added the source of the tree distribution map.

Figure 2: Certainly in the main text and maybe also here (e.g. in the figure caption) define MP-AES. For text in right-middle box consider: 'Residual iron content per gram of wood . . .'. Also, it may be more suitable to use past tense in the text boxes: e.g. 'lath1 digested using . . .'; 'lath 2 rinsed in de-ionized water for 2 hrs, then air dried and sanded', etc.

Response: We added the full name of MP-AES in the text (at Line 112) and figure caption. The past tense was used in Figure 2.

Figure 3 caption: The description for (a) should make it clear that these are multiple laths from one LST - for example: '(a) shows examples of differently treated laths from one LST and one living tree sample (last row)'. When referring to panels

(b) and (c), it would be clearer to write it using the following format: (b) description, and (c) description – (instead of (b), and (c) description)

Response: We modified the Figure 3 caption:
"Figure 3. Residual iron (Fe) and wood RGB intensities (see definitions in Table 2) of LST laths treated with seven chemical reagents. (a) shows examples of treated laths from one LST sample and one living tree sample (last row). The gray outer part of the example LST is discolored due to decay. (b) shows the mean wood RGB intensities according to treatment. (c) shows the mean concentrations of residual Fe according to treatment. Error bars in (b) and (c) refer to standard deviations of corresponding group. Percentages in (c) refer to the Fe removed by de-staining treatments relative to the Fe concentrations of untreated stained LSTs. (d)–(f) show the linear regressions of earlywood, latewood and delta RGB intensities against the log of residual Fe. Regressions are based only on the LST data (circles). Living-tree data are plotted as triangles but are excluded from the regressions.".

Figure 6: Consider also adding some statistics in the panels such as the full period calibration. In the caption, maybe also specify the type of filter used.

Response: Done. See the revised Figure 6 in the supplementary reply letter.

Minor comments:
L36: Consider changing to e.g. '. . . high cost of X-ray densitometric equipment' . . . unless this is meant to refer to the relatively high costs associated with processing samples at a 'facility' that is equipped to perform X-ray densitometry. In any case, a small edit is needed.

Response: Corrected.

L37: It may be more accurate to state that the production of BI is 'relatively cheap' rather than 'cheap'.

Response: Corrected to "In contrast, BI is more affordable because it uses commercial flatbed scanners and image analysis software to measure the blue light reflectance of tree rings (Rydval et al., 2014)" in Lines 37–39.

L39: 'coherence' instead of 'coherences'

Response: Corrected.

L43: Perhaps be a bit more specific here by changing 'color issues' for example to something like 'color inconsistencies' or 'inconsistent color properties'?

Response: Corrected to "heterogeneous colors".

L44/45: As this is a property of the wood, 'leading to' is not really appropriate here – instead perhaps ' . . . exhibiting darker heartwood than sapwood'

Response: Yes, that is true. We corrected this sentence to "The best-known issue is the sapwood-heartwood color difference of several tree species such as pine and larch, which does not co-vary with density" in Lines 44–45.

L47: ‘occurs’ instead of ‘happens’

Response: Corrected.

L48: ‘LSTs are . . .’?

Response: Corrected.

L49: just ‘replication’

Response: Corrected.

L53: The second part of this sentence could be improved for example along the lines of ‘. . . to realize the potential of the promising BI technique . . .’

Response: We changed this sentence to “Therefore, it is critical to develop unbiased BI data from a variety of wood materials, in particular from the LSTs, to make the promising BI technique widely applicable in future dendroclimatic reconstructions.” in Lines 52–53.

L58: ‘consists of’ instead of ‘consists in’; the word ‘delta’ can be removed from this part of the sentence

Response: Corrected.

L59: Consider changing ‘DBI is suitable to recover . . .’ for example to ‘DBI suitably represents . . .’ or ‘DBI corrects for . . .’

Response: Corrected.

L64: The last part of the sentence should be re-phrased e.g. ‘. . . without utilizing low frequency information of the more temperature-sensitive BI data’ / ‘. . . without benefitting from / exploiting any potential improvements in the low frequency domain from the more temperature-sensitive BI data’

Response: Corrected.

L69/70: Consider changing the last part of the sentence to something along the lines of ‘. . . to a standard comparable to MXD-based reconstructions’

Response: Corrected.

L80/81: minor change needed here, e.g. ‘. . . LSTs after falling into the water and eventually becoming buried in lake sediments’

Response: Corrected.

L88: ‘fungal’ rather than ‘fungi’? Also, is there a better word that could replace ‘invasion’ here? Perhaps simply ‘fungal discoloration’?

Response: Corrected.

L89: ‘before the year’

Response: Corrected.

L94: Should this be ‘radius’ instead of ‘radii’ if laths are cut along a single radius, otherwise specify the number of radii

Response: Corrected.

L96: This should probably be ‘weighed’ instead of ‘weighted’

Response: Corrected.

L102: I would recommend presenting the figures in the order that they are first mentioned in the text (Fig. S1 is first mentioned here while Fig. S2 is first mentioned in L131).

Response: We have corrected the order of all figures and tables in the revised manuscript.

L108: This could probably be improved slightly – e.g. ‘one lath from each pair of laths’?

Response: Corrected.

L111: Please mention first that MP-AES stands for "microwave plasma-atomic emission spectrometer"

Response: Added at Lines 112.

L112: ‘. . . weight of the corresponding . . .’

Response: Corrected.

L113: The meaning of this sentence was not immediately clear to me. Consider the following minor revision for the sake of clarity: ‘. . . represent the combined total . . . cannot be separately distinguished by MP-AES’

Response: Corrected to "Fe concentrations in this study represent the total amount of ferrous and ferric Fe, because MP-AES does not distinguish the type of Fe ions." in Lines 114–115.

L115: 'grit' rather than 'grits'

Response: Corrected.

L118: 'interference' instead of 'inferences'

Response: Corrected.

L122: I think this could be explained a bit better – in a more specific way. Also, 'consistent with' rather than 'consistent to'

Response: We changed the expression to "Because high RGB values represent light colors (i.e. high brightness), they were subtracted from a value of 256 such that smaller RGB values are associated with lighter colors" in Lines 122–123.

L141: perhaps '(i.e. each lath pair)'?

Response: Corrected.

L143: 'An age-dependent spline . . .'

Response: Corrected.

L148/L150: Also in relation to Fig. S8 / S7, see earlier comment about sequential order of figures mentioned above.

Response: We have corrected the order of all figures and tables in the revised manuscript.

L151: maybe 'poor (tree) health' would be better than 'unhealthy tree growth'

Response: Corrected.

L154: 'averaged into'? Also, maybe consider showing the individual MXD chronologies in the SI.

Response: Added to SI material, Fig. S9. See Fig. S9 in the supplementary reply letter.

L161: This could use a bit of re-wording.

Response: We changed this sentence to "The reconstructions were based on the scaling method (Esper et al., 2005; Rydval et al., 2017) by adjusting means and standard deviations of the chronologies to those of the temperature target over the 1901–2015 time interval" in Lines 161–163.

L165: maybe slightly re-word – e.g. '. . . in order to assess the role of Fe in the staining issue'

Response: We changed this to "in order to assess the roles of Fe in the staining issue" in Line 166–167.

L167: Perhaps just briefly specify the advantage of applying this specific filter in the context of high-pass / low-pass filtering.

Response: We used this method because it is the default option of the dplR package, not because of a specific advantage. The filtered series were very similar between using the Butterworth filter and Chebyshev filter.

L168: 'Performance of the reconstructions was assessed ...'

Response: Corrected.

L170: '. . . while the 1961-2015 period . . .'

Response: Corrected.

L178-179: Just as a general point, I wonder what the reason for this might be? Could it be related to the 'color properties' of the staining caused by Fe?

Response: Yes, this is correct. This is related to the properties of Fe oxides that stained the wood. In general, Fe oxides have stronger red light reflectance. From this aspect, our conclusions are also consistent with this general phenomenon. We feel slightly confused by the real-world colors of the stain on wood, which look more blue than red, however such "blue" stain reflect more red color according to our analysis (Fig. 3b).

L182: 'intensities'

Response: Corrected.

L186-187: '. . . of treated earlywood and latewood'?

Response: Corrected.

L189: Please be more specific here and elaborate on what the difference is.

Response: We changed it to "Note that wood delta BI and DBI were not calculated exactly in the same way (see Table2; delta BI is the averaged difference between BI of entire latewood and earlywood from all tree rings in a sample, while DBI is a tree-ring parameter which presents the difference between LBI and BI of entire earlywood for each tree ring).” in Lines 194–196.

L191: Consider something along the lines of ‘. . . four treatments examined in more detail . . .’

Response: Sorry, but we could not understand this comment

L192: ‘prior to 1900 CE’ or ‘prior to the year 1900 CE’

Response: Corrected.

L193: ‘coherence’ rather than ‘coherences’

Response: Corrected.

L196: ‘few differences’ instead of ‘little differences’

Response: We changed the sentence to “In addition, few differences were found between the control DBI series and chemically treated DBI data, although the colors of wood samples were visually distinct (Fig. 3a).” in Lines 203–204.

L203: ‘Briefly’ is probably not needed here

Response: We prefer to keep it. The term is suggested by one of the co-author working in geochemistry, because the chemistry of iron is complex.

L208: I would recommend re-wording ‘combine to wood’ – ‘combine with wood’ or maybe ‘bind to wood’?

Response: We corrected it to “bind”.

L225: just ’etc.’ instead of ’and etc.’

Response: Corrected.

L228: maybe ‘in our samples’ rather than ‘from our samples’?

Response: Corrected.

L229-230: Minor edit needed here: e.g. ‘is not sensitive to sulfur and phosphorous’ / ‘is not designed to detect sulfur and phosphorous’ or ’is not sufficiently sensitive to detect . . .’?

Response: We made some changes to clarify this. “However, the MP-AES instrument is not sufficiently sensitive to verify our hypothesis regarding those Fe complexes (detection limits are ~6500 ppb and 125 ppb for dissolved sulfur and phosphorus, respectively, compared to ~4.6 ppb for dissolved Fe)” in Lines 236–238.

L234: ‘. . . very little residual Fe . . .’

Response: Corrected.

L236: probably replace ‘such’ with ‘the’

Response: Corrected.

L240: maybe something like ‘. . . when staining is present in subfossil wood’ would be better

Response: Corrected.

L242: ‘a’ should be removed

Response: Corrected.

L247: something is missing here – e.g. ‘. . . for example as with the most efficient MixC protocol’ or something similar

Response: Corrected.

L250-251: ‘This evidence suggests that DBI is not only an excellent solution to resolving ...’

Response: Corrected.

L251: ‘but also efficiently resolves ...’

Response: Corrected.

L256-257: The last part of this sentence could be re-worded and expanded a bit to clarify what is meant by this.

Response: We made some corrections. “Although DBI is theoretically sufficient to solve the sapwood-heartwood color issue (Björklund et al., 2014), in our case it could only partially correct this problem (Fig. S8). Old living trees were collected from lakeshore forests at the L20 site and they often displayed declining ring widths compared to healthy trees sampled later at the same site (not shown). DBI of L20 is likely influenced by these narrow tree rings (Björklund et al., 2019) because DBI of black spruce is not only correlated to MXD but also to the ring-width data (Wang et al., submitted). We thus speculate the divergence of DBI reflects mostly a specific issue related to the declining growth of unhealthy trees.” in Lines 262–267.

L258: replace ‘declined’ with ‘declining’

Response: Corrected.

L259: ‘decline’ instead of ‘declines’? Also, the reference Björklund et al., 2019 (Reviews of Geophysics) may also be relevant to this point.

Response: Corrected and reference added.

L261: ‘need to be’ rather than ‘need be’?

Response: Corrected.

L268-269: ‘. . . further attention / investigation is needed . . .’

Response: Corrected.

L281: should probably be ‘. . . used as part of Fe extraction protocols . . .’

Response: Corrected.

L282: ‘which also face / are also susceptible to / also suffer from’?

Response: Corrected.

Reply to referee2:

We are grateful for the useful comments from Dr. Jesper Björklund. Our responses are listed below:

The manuscript by Wang et al presents a very interesting sample material for temperature reconstructions and examine how to best utilize this in conjunction with the popular and affordable BI technique. The paper is foremost dedicated to a very novel and clever de-staining experiment which I thoroughly enjoyed and have the potential to be highly cited in future BI studies. The second component was a careful comparison of LBI, DBI and MXD from parallel X-ray measurements to evaluate the performance of the chemical de-staining and LBI and DBI parameters with MXD as reference. Although the authors conclude that the simple DBI was more successful in replicating the low-frequency variance of the MXD, they have made some very important discoveries in terms of de-staining of relict wood material. The DBI parameter appears to be quite successful, but has documented problems as the authors also mention in the final sentences. Therefore, all tools available for de-staining prior to DBI transformation must be considered of great value. I congratulate the authors to a fine, and from what I can tell labor intensive, experiment and I consider the manuscript suitable for publication following minor revisions and clarifications. I also look forward to learn more about the planned follow-up manuscript.

Detailed comments:
L32 We would not say BI is recently developed anymore, it has been around almost 20 years now.

Response: Yes, we totally agree. We removed "recently developed".

L32-33 The BI technique is an alternative to the X-ray technique in producing proxy parameters such as MXD.

Response: We re-phrased the first sentence as: "The blue intensity (BI) technique is an alternative to the more expensive X-ray densitometric methodology in producing tree-ring proxy parameters such as maximum latewood density (MXD) for dendroclimatology." in Lines 32–33.

L37 -38 Consider changing to something like: "In contrast, BI is more affordable because of the utilization of commercial flatbed scanners to generate images of reflected blue light analyzed in potentially affordable image analysis software. . ."

Response: We re-phrased this sentence as "In contrast, BI is more affordable because it uses commercial flatbed scanners and image analysis software to measure the blue light reflectance of tree rings." in Lines 37–39.

L38-42 Strange sentence, some of the studies encouraging more studies were made later than the encouraged studies. Work a bit more on this sentence and consider also these references: Björklund et al., 2014, 2015; Dolgova, 2016; Fuentes et al., 2018; Kaczka et al., 2017; McCarroll et al., 2013; Rydval, Gunnarson, et al., 2017.

Response: Several of the suggested references are cited elsewhere in the manuscript. Here we only wish to cite studies specific to latewood BI rather than delta BI, because the topic of delta BI is detailed in the next paragraph. So we only added McCarrol et al., 2013. We improved the sentence to make it more fluent. "Excellent coherence was reported between the latewood BI (LBI) and MXD data measured from living-tree materials of a number of coniferous tree species across the northern hemisphere (Campbell et al., 2007; Kaczka et al., 2018; Österreicher et al., 2015; Rydval et al., 2014; Wilson et al., 2014), suggesting the potentials to use BI method in dendrocliamtic reconstructions (McCarrol et al., 2013; Rydval et al., 2017; Wilson et al., 2019)." in Lines 39–42. The later three references (McCarrol et al., 2013; Rydval et al., 2017; Wilson et al., 2019) are real reconstruction works.

L45 Should perhaps add something like: "..not accompanied by a similar difference in density.."

Response: I corrected this sentence to: “The best-known issue is the sapwood-heartwood color difference of several tree species such as pine and larch, which does not co-vary with density.” in Lines 44–45.

L83 newly exploited lake?

Response: This has been changed to “newly sampled lake”.

L84 millennium-long?

Response: We think the meanings of “millennium-long” and “millennial” are very similar. We keep “millennial” here.

L96 What was the purpose of the weighing? Were the laths also weighed after the chemical analysis? Could not find any more use of these measurements in the manuscript

Response: The weights of wood laths were used to calculate the Fe concentrations in the wood. This is because Fe concentrations in wood depend strongly on the weights of wood laths. The iron data shown in the original manuscript had already been adjusted by the wood weights. It was described at Line 113 of the revised manuscript, “data were adjusted to “milligram of Fe per gram of wood” according to the dilution and weight of the corresponding wood lath.”.
We only weighed a few post-treatment wood laths. This is because we set two control groups in addition to the chemical treatments, i.e. untreated stained samples and living tree samples. For these samples, we would not know their post-treatment weights. A consistent measure of weight from laths should be used, i.e. pre-treatment weight. Second, the amount of wood burrs contacted on the surface of laths might also affect the post-treatment weights. These burrs were often found detached from the laths after treatments (samples are placed in a tube on a shaker for at least 24 hrs). This mass loss varies tree by tree, and we cannot quantify the weight losses. For example, being similarly treated with MixC, one lath lost 0.2% of weight relative to a pre-treatment weight (0.2866g) while another lath could loss 1.03% of its weight (pre-treatment weight was 0.1935g).

L104 .., to identify the most effective. . .? Remove “(see results below)”. The results are always be presented after the methods description.

Response: Ok, we removed “(see results below)”.

L118 sensu Rydval et al., 2014?

Response: We added this reference here.

L118-119 Great initiative
L121-122 Very strange statement. Real world observations? Do you mean: lower RGB values corresponds to lighter densities?

Response: We changed the expression to “Because high RGB values represent light colors (i.e. high brightness), they were subtracted from a value of 256 such that smaller RGB values are associated with lighter colors.” in Lines 122–124.

Section 2.2, 2.3 and 2.4 Consider re-structuring here. Perhaps one section for chemical de-staining description. One section for BI and X-ray data development and one section for chronology development for climate analysis, and sample average RGB data?

Response: The manuscript comprises a de-staining experiment and a dendroclimatic assessment. Accordingly, we kept section 2.2 and merged sections 2.3, 2.4 and 2.5 to obtain one section for the de-stanning experiment (section 2.2) and one section for the dendroclimatic assessment (section 2.3)

L139 Did you use the full RGB spectrum or only the blue spectrum? If the latter, it is consistent with the use of BI based parameters. Same comment in L165.

Response: We extracted the red, green and blue intensity values from earlywood and latewood of each ring, then averaged values by colors and laths. Of course, for the LBI and DBI we considered only the blue spectrum. A revised version of Table 2 (in the supplementary reply letter) clarifies the measured parameters. We could not find the comment in L165.

L145 N.B. residuals are most often used for density related parameters. This is not a major problem here since you compare results from BI and X-ray, but may be important in pure climate reconstructions.

Response: Ok, we will think of calculating residuals in future reconstructions.

L168 "coherence" can also be a type of statistical analysis, perhaps change to the more general term of "agreement", or simply not explain correlation since more or less the entire readership is familiar with this.

Response: We changed it to "agreement".

Figure S4 Spelling of replication

Response: Corrected.

Figure s6 spelling of earlywood. It seems odd that the area of the 30% of the darkest pixels in the latewood are differently sized even though the latewood area is roughly the same (compare ring 4 and ring 5). Please check the definition you used and clarify why this is the case.

Response: Misspelling Corrected. We systematically used the 30% of the darkest latewood BI as LBI. However, the Figure S6b was generated using the densitometer function of CooRecorder (See Figure1 in the supplementary reply letter). We replaced this figure with the output of the actual LBI measurements (Figure2 in the supplementary reply letter).

L169-171 Would be great to have running Rbar or EPS, to evaluate the difference between the different parameters. Perhaps this can explain why the DBI perform so badly in the post 1960 period compared to LBI and MXD. Both in terms of trend and correlation.

Response: We added EPS values in the revised Figure 6 (in the supplementary reply letter). Some relevant discussions were made on the EPS data (at Line 273 in the revised manuscript). It is true that EPS of DBI was more unstable than LBI.

L182 spelling intensities

Response: Corrected.

Figs. S7-S8 Would be interesting to also present the Earlywood measurements. Would be even more interesting if you also presented Delta density and Earlywood density. It is puzzling why LBI and DBI has such similar trends in S7. Is there a HW/SW transition in these trees, if so why so weak in the earlywood? Are the rings in the post 1960 period very narrow? If so, I think that your measurement resolution is causing some problems here. Consider that the measurement resolution is affecting your latewood measurements more than your earlywood measurements. That is, your latewood BI is deflated because of adjacent contamination of earlywood BI. Ergo the delta BI will be artificially lowered and similar in trend to LBI. Not completely relevant to your nice study, but could not resist :)

Response: We did not extract the earlywood BI and earlywood density data for this manuscript, which are in very raw forms, and consequently delta density data are not available. We responded to the rest of this comment below, under the L253–260 comment.

L196 check grammar

Response: We changed the sentence to "In addition, few differences were found between the untreated control DBI series and chemically treated DBI data, although colors of wood samples were visually distinct (Fig. 3a)" in Lines 204–205.

L208 combine to wood? Not clear, rephrase..

Response: We replaced "combine" with "bind".

L241-242 This is not surprising. If you would calculate delta density and correlate with delta BI you would probably find equally high correlation as between LBI and MXD. This is not needed in revision, I am merely pointing this out.

Response: OK. We deleted this sentence.

L253-260 I think you are right that the narrow ring widths are causing the problem here, but I would not say it is a healthy versus unhealthy tree problem. It is a problem of measurement resolution (see comment above for fig s7). Healthy tree can also have narrow rings.

Response: Yes. The L20 samples showed narrow rings and distinct sapwood-heartwood colors. This latter phenomenon is unusual for black spruce because this species usually does not show distinct boundary between earlywood and latewood. We speculate that this phenomenon was resulted from the unhealthy state of some trees in this site because in the early stages of sampling, we collected trees as old as possible, regardless of their crown shape and growth rate. We agree that the cause of the divergence (in particular for the DBI which in theory is not affected by color) is most likely the narrow rings. We thus modified Lines 155–160 in the original manuscript to:

"Old living trees were collected from lakeshore forests at the L20 site and they often displayed declining ring widths compared to healthy trees sampled later at the same site (not shown). DBI of L20 is likely influenced by these narrow tree rings (Björklund et al., 2019) because DBI of black spruce is not only correlated to MXD but also to the ring-width data (Wang et al., submitted). We thus speculate the divergence of DBI reflects mostly a specific issue related to the declining growth of unhealthy trees." in Lines 263–267.

L262 yes interesting observation. Would be better underpinned if you also presented the rbar for all the parameters.

Response: We added moving EPS values in Figure 6 (see revised Fig. 6 in supplementary reply letter).

Relevant changes were listed in the responses above and marked in red in the revised manuscript below.

Chemical de-staining and the delta correction for blue intensity measurements of stained lake subfossil trees

Feng Wang[1, 2*], Dominique Arseneault[1, 2], Étienne Boucher[3], Shulong Yu[4], Steeven Ouellet[1], Gwenaëlle Chaillou[5], Ann Delwaide[6], Lily Wang[7]

[1]Département de Biologie, Chimie et Géographie, Université du Québec à Rimouski, Rimouski, G5L 3A1, Canada
[2]Centre d'études nordiques, Université du Québec à Rimouski, Rimouski, G5L 3A1, Canada
[3]Département de Géographie, GEOTOP and Centre d'études nordiques, Université du Québec à Montréal, Montréal, H3A 0B9, Canada
[4]Key Laboratory of Tree-ring Ecology of Uigur Autonomous Region and Key Laboratory of Tree-ring Physical and Chemical Research, Institute of Desert Meteorology, China Meteorological Administration, Urumqi, 830002, China
[5]Canada Research Chair in geochemistry of coastal hydrogeosystems, Québec-Océan, UQAR/ISMER, Rimouski, G5L 3A1, Canada
[6]Département de Géographie, Université Laval, Quebec City, G1V 0A6, Canada
[7]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Science, Beijing, 100101, China

Correspondence to: Feng Wang (feng.wang@uqar.ca)

Abstract. The stain of wood samples from lake subfossil trees (LSTs) is challenging the wide application of the blue intensity (BI) technique for millennial dendroclimatic reconstructions. In this study, we used seven chemical de-staining reagents to treat samples of subfossil black spruce (*Picea mariana* (Mill.) B.S.P.)) trees from two lakes in the eastern Canadian boreal forest. We subsequently compared latewood BI (LBI) and delta BI (DBI) time series along with conventional maximum latewood density (MXD) measured from the stained and de-stained samples. Results showed that the stain of our samples is most likely caused by post-sampling oxidation of dissolved ferrous iron in lake sediments that penetrated into wood. Three reagents (ascorbic acid, sodium ascorbate, and sodium dithionite all mixed with ethylenediaminetetraacetic acid) could remove >90% of Fe. However, even for the best chemical protocol, a discrepancy of about +2°C compared to MXD data remained in the LBI-based temperature reconstruction due to incomplete de-staining. On the contrary, the simple mathematical delta correction, DBI was unaffected by Fe stain and showed very similar results compared to MXD data (r>0.82) from annual to centennial timescales over the past ~360 years. This study underlines the difficulty of completely de-staining lake subfossil samples, while confirming the robustness of the DBI approach. DBI data measured from stained LSTs can be used to perform robust millennial temperature reconstructions.

1. Introduction

The blue intensity (BI) technique is an alternative to the more expensive X-ray densitometric methodology in producing proxy parameters such as maximum latewood density (MXD) for dendroclimatology (Björklund et al., 2019; McCarroll et al., 2002). MXD is the most suitable tree-ring parameter for summer temperature reconstructions in northern and high-altitude regions (Esper et al., 2014; Frank and Esper, 2005). However, compared to the less climate-sensitive ring-width data, millennial MXD series have been much less frequently developed worldwide, mainly due to the high cost of densitometric equipment (Anchukaitis et al., 2017; St. George and Esper, 2019; Wilson et al., 2016). In contrast, BI is more affordable because it uses commercial flatbed scanners and image analysis software to measure the blue light reflectance of tree rings (Rydval et al., 2014). Excellent coherence was reported between the latewood BI (LBI) and MXD data measured from living-tree materials of a number of coniferous tree species across the northern hemisphere (Campbell et al., 2007; Kaczka et al., 2018; Österreicher et al., 2015; Rydval et al., 2014; Wilson et al., 2014), suggesting the potentials to use BI method in dendrocliamtic reconstructions (McCarrol et al., 2013; Rydval et al., 2017; Wilson et al., 2019).

However, the BI technique is also facing challenges due to heterogeneous colors of various wood materials. The best-known issue is the sapwood-heartwood color difference of several tree species such as pine and larch, which does not co-vary with density (Björklund et al., 2014; Rydval et al., 2014; Wilson et al., 2019). Some wood types, such as dead trees and historical materials, may also be discolored by decay or weathering (Wilson et al., 2014). Another less documented color issue occurs with lake subfossil trees (LSTs), which often have darker wood than living trees (Wilson et al., 2019). LSTs are a very interesting source of material to extend tree-ring chronologies from centuries to several millennia and can greatly improve the replication of reconstructions, especially in regions where only short-lived tree species occur (Arseneault et al., 2013; Grudd et al., 2002). All the above color issues may potentially alter the accuracy of BI data (e.g. LBI) and introduce biases in BI-based climate reconstructions, particularly for the low-frequency domain (Björklund et al., 2014; Wilson et al., 2019). Therefore, it is critical to develop unbiased BI data from a variety of wood materials, in particular from LSTs, to make the promising BI technique widely applicable in future dendroclimatic reconstructions.

Some solutions have been proposed to overcome the sapwood-heartwood color issue. Sheppard and Wiedenhoeft (2007) used hydrogen peroxide to bleach wood colors and attenuate the sapwood-heartwood difference, although an earlier study claimed that chemical bleaching likely degraded the climate signal (Sheppard, 1999). A later and seemingly more promising approach consists of a simple mathematical correction computed as the BI difference between the earlywood and latewood, (i.e. the delta BI; hereafter DBI) of each tree ring. DBI corrects for the low-frequency (decadal–centennial) variations distorted by the sapwood-heartwood color issues (Björklund et al., 2014, 2015). Only a few studies attempted to use BI data measured from LSTs to develop long-term temperature reconstructions, but the staining issue was not addressed and these BI series were not directly compared to MXD data. For example, Rydval et al. (2017) combined the high-frequency temperature signals extracted from LBI with the low frequency from ring-width data of LSTs in order to reconstruct temperatures, without using the low-frequency information of the more temperature-sensitive BI data.

In this study, we explore the potential of generating unbiased BI series from stained black spruce LSTs from two eastern Canadian lakes (Fig. 1a). More specifically, we compared chemical de-staining with the DBI approach as well as with conventional MXD data. The following hypotheses were formulated: (1) the stain of LSTs is mainly caused by oxidation of Fe (Hyacinthe et al., 2006; Kostka and Luther, 1994; Pelé et al., 2015; Zhang and Xi, 2003), which can be removed using some anti-oxidant reagents; (2) chemical de-staining and DBI can greatly improve BI-based temperature reconstructions that are comparable to MXD-based reconstructions. We first treated thin wood laths of stained LSTs using seven potential chemical reagents and quantified the proportion of Fe extracted by chemicals. Subsequently, we performed dendroclimatic assessments by comparing LBI and DBI chronologies of the de-stained and stained LST samples with MXD data over the past ~360 years.

2. Materials and methods

2.1 Study sites and staining issue

The two studied lakes (L20 and L105) are located approximately 450 km apart in the eastern Canadian boreal forest of the Quebec-Labrador Peninsula (Fig. 1a). The climate of this region is characterized by short, mild summers and long, cold winters (Environment Canada, 2020). Regional forests are strongly dominated by black spruce, mixed with balsam fir (*Abies balsamea* L.) and eastern larch (*Larix laricina* (Du Roi) K. koch) (Payette, 1993). Lakes are extremely abundant and cover up to about 25% of the landscape. Numerous black spruce trees in the lakeshore forests thus become LSTs after falling in water and eventually becoming buried in lake sediments (Arseneault et al., 2013; Gennaretti et al., 2014a, 2014b). Black spruce LSTs from L20 (54.56º N, 71.24º W) were previously included in a millennial temperature reconstruction (Gennaretti et al., 2014c). L105 (50.81º N, 67.80º W) is a newly sampled lake where more than a thousand black spruce LSTs were extracted, and many of them were successfully cross-dated to develop a millennial ring-width chronology (unpublished data).

LSTs in the eastern Canadian boreal forest are frequently stained to various blue-gray intensities (Fig. S1). When dry, these stains correspond well to the Munsell Soil Color Chart 2009-5Y-Chroma I. In total, 78% and 79% of the cross-dated LSTs at L20 and L105, were stained to some degrees of gray, respectively (Fig. 1b, c). Very few LSTs displayed additional colors, for example due to fungal discoloration. Stained LSTs are distributed throughout the timespan of the two millennial chronologies with increasing proportions back in time, particularly before the year 1800 CE (Fig. 1b, c), suggesting that staining issues are unavoidable when using BI measurements from this material.

2.2 De-staining experiment

We selected stem cross sections from five (evenly and heavily) stained LSTs and five unstained lakeshore living trees from each of the two studied lakes. Using a twin-blade saw (DendroCut, Walesch Electronic), we transversally cut sixteen 1 mm-thick laths along the radius of each subfossil tree (total of 160 laths), and 2 laths from each living tree (total of 20 laths). All laths were pretreated using 95% ethanol in Soxhlet extractors for 48 hours to remove resins and then air-dried and weighed. Sixteen pretreated laths from each LST were divided into eight pairs. Seven pairs were immersed in 50 mL of one of the seven chemical solutions (Table 1) in Falcon® 50mL tubes (see an example in Fig. 2), then placed on an electronic shaker (SK-600, Montreal Biotech Inc.) with a speed of 133 rpm at room temperature (c.a. 20 ºC) for 24 hours (MixC) and 48 hours (other reagents). Treated laths were rinsed 4–5 times, immersed in de-ionized water for 2 hours to remove dissolved elements absorbed by wood tissues, and then air-dried for subsequent analysis. The eighth pair of subfossil laths and all the living-tree laths were not treated with de-staining reagents and considered as control samples. Design of the experiments is illustrated in Fig. S2.

During the de-staining treatments, we sequentially sampled reaction solutions to construct temporal Fe dissolution curves (the most abundant metal element detected in preliminary tests) for MixA, MixB, and MixC, which were the most effective de-staining reagents. 1 mL of solution was sampled at 0, 1, 3, 6, 12, and 24 hours for MixA and MixB, and at 0, 0.5, 1, 3, 6, and 12 hours for MixC (Fig. 2). Treatment and sampling times were set shorter for MixC since it was the most active reagent. 1 mL extracted solutions were then diluted in 5 mL of 5% (v/v) hydrochloric acid in Falcon® 15 mL tubes in order to avoid Fe(II) precipitation prior to the chemical analysis of Fe concentrations. After de-staining, one lath from each pair of laths was used to quantify the amount of residual Fe, and was digested using 5 mL nitric acid and 1 mL hydrogen peroxide in a MARS-Xpress microwave digestion system (CEM Corporation) at 150 ºC for 30 minutes (Fig. 2). Digested solutions were diluted to 25 mL in volumetric flasks using de-ionized water. Fe concentrations were measured using the microwave plasma-atomic emission spectrometer (MP-AES, Agilent 4200; the limit of detection is ~4.6 ppb for dissolved Fe) and data were adjusted to “milligram of Fe per gram of wood” according to the dilution and weight of the corresponding wood lath. Fe concentrations in this study represent the total amount of ferrous and ferric Fe, because MP-AES does not distinguish the type of Fe ions.

The second lath of each pair was air-dried, finely sanded (to 1000 grit), and scanned using the SilverFast 8.0 software (LaserSoft Imaging) and an Epson V800 flatbed scanner. In order to obtain optimal calibration results, the sanded laths were scanned to RGB images of 3200 dpi along with a color IT8.7/2 calibration target (LaserSoft Imaging). The scanner was covered by a black plastic box to avoid interference of external light (Rydval et al., 2014). It should be noted that the actual image resolution is approximately 2580 dpi (horizontal) by 1825 dpi (vertical) according to the USAF-1951 resolution target. Wood RGB intensities (definition 1–3 in Table 2), were then measured using the CooRecorder 8.1 software (Cybis Dendrochronology). Because high RGB values represent light colors (i.e. high brightness), they were subtracted from a value of 256 such that smaller RGB values are associated with lighter colors. RGB intensities were compared among treatments (seven treatments plus two controls) to assess the efficiency of the de-staining reagents.

2.3 Dendroclimatic assessment

In order to perform dendroclimatic assessments of the most effective de-staining treatments, in a separate experiment we selected 57 trees of different types (stained and unstained LSTs as well as living trees) which were cross-dated after the year 1600 CE from L20 and L105 (28 and 29 trees, respectively; Fig. S3). We cut eight 1 mm-thick laths from two radii of each tree to acquire four pairs of laths (Fig. S4). Laths were pretreated using 95% ethanol in Soxhlet extractors for 48 hours. We then treated three pairs of laths per tree using the MixA, MixB, and MixC, respectively, while keeping the fourth lath as an untreated control. The conditions of de-staining treatments were the same as explained above, except that the treatments lasted for six hours, which is the time required for optimal de-staining according to the Fe dissolution curves (Fig. S5). Treated laths were then air-dried for LBI, DBI, and MXD measurements.

Measured tree-ring parameters are explained in Table 2. The wood laths for dendrochronological assessments were firstly X-rayed to generate MXD data prior to being sanded and scanned for BI measurements. X-ray densitometry experiments were conducted in a controlled environment with relative humidity of 50% and room temperature of 20ºC. X-ray films were developed using the DendroXray2 system (Walesch Electronic) and MXD series were measured using the Dendro2003 system (Walesch Electronic) with a resolution of 10µm along the measured density profile. LBI and DBI were measured using the same procedure as the measurement of wood RGB intensities. We subtracted the raw LBI values from a value of 256 in order to make the LBI positively correlated with MXD data, according to Rydval et al. (2014) and Wilson et al. (2019). Before data analysis, LBI, DBI, and MXD data were averaged by tree (i.e. each lath pair) for each treatment (MixA, MixB, MixC, and Control).

We used regional curve standardization (RCS) to remove the biological trends from tree-ring series in order to retain low-frequency (decadal–centennial) climatic variations (Briffa and Melvin, 2011; Helama et al., 2017). An age-dependent spline with an initial stiffness of 2 years was used to estimate the regional curve. Standardized tree-ring series were computed as ratios between the raw data and the smoothed regional curve. In total, we standardized 24 groups of tree-ring data by site (L105 and L20), parameter (LBI, DBI, and MXD) and treatment (MixA, MixB, MixC, and Control). We excluded data from chemically treated, unstained trees (unstained LSTs plus living trees) because BI data of unstained trees tended to be altered by reagents, mostly the MixC (Fig. S7). Consequently, we pooled the data of stained LSTs that were chemically treated, plus the data of the untreated, unstained trees for each standardization. In addition, living-tree data of L20 after the year 1950 CE were excluded as BI diverged from the MXD data (Fig. S3 and S8), which was likely due to the sapwood-heartwood color issue and narrow rings caused by poor tree health (see discussions below).

Regional chronologies for each tree-ring parameter (LBI, DBI, and MXD) and treatment (MixA, MixB, MixC, and Control) were generated by pooling standardized series from both sites using the Tukey's bi-weight robust mean. This approach was used because of the limited tree replication per site (Fig. S3). Regional MXD chronologies from the four treatments were similar (Fig. S9) and thus averaged into one reference chronology. All regional chronologies were truncated at the year 1655 CE to ensure a minimum replication of five trees (Fig. S3c).

We performed temperature reconstructions using the regional LBI, DBI, and MXD chronologies to further quantify the influence of the de-staining protocols. Instrumental summer (May to August) temperature data were obtained from the CRU TS 4.02 0.5° gridded monthly mean temperature dataset (Harris et al., 2014) and averaged from the four grid cells closest to each lake in order to generate a regional temperature target. The reconstructions were based on the scaling method (Esper et al., 2005; Rydval et al., 2017) by adjusting means and standard deviations of the chronologies to those of the temperature target over the 1901–2015 time interval.

2.4 Data analysis

Data were analyzed using the R program (R Core Team, 2018). We conducted linear regressions between wood RGB intensities and logarithmic residual Fe of both treated and untreated LSTs in order to assess the roles of Fe in the staining issue. For chronology assessments, we generated several high-pass and low-pass LBI, DBI, and MXD series using the Butterworth filter available in the "dplR" R package (Bunn, 2008). Pearson correlation coefficients were used to assess the degrees of agreement among all the time series. Performance of reconstructions was assessed following a regression-based calibration-verification procedure using the "treeclim" R package (Zang and Biondi, 2015). Since our chronologies showed higher replication during 1901–1960 (Fig. S3), this time interval was used for calibration, while the 1961–2015 period was used for verification. We also calculated 1-year lagged 31-year moving expressed population signal (EPS; Wigley et a., 1984) to assess the temporal robustness of chronologies used for reconstruction.

3. Results

3.1 Effects of chemical de-staining

LSTs displayed a variety of color changes after the seven de-staining treatments (Fig. 3a). NaAsc resulted in very similar colors to that of the untreated stained samples, representing the weakest de-staining effect. Conversely, MixA, MixB, and MixC showed dramatic effects and almost completely removed the gray stain. MixC was the most effective de-staining solution based on wood RGB intensities, although the resultant colors still slightly differed from the living-tree standards (Fig. 3a, b). BI was less variable with varying degrees of post-treatment stains in comparison to red and green intensities (Fig. 3b), suggesting a potentially weaker influence of wood stain on the BI data. MixC, the only reagent with bleaching function (Table 1), could have an additional bleaching effect on the wood, resulting in smaller LBI and DBI values in living and unstained trees compared to untreated control (Fig. S7). However, DBI of the stained LSTs was only slightly modified by the MixC treatment (Fig. S7), indicating that the bleaching effect of MixC is weak for the stained samples.

3.2 Stains versus iron

Chemical analyses showed strong links between Fe concentrations and color intensities of wood, especially for green and red intensities (Fig. 3a, c). Total Fe concentration was the highest for untreated stained LSTs and near zero for living trees (Fig. 3c). MixA, MixB, and MixC could remove 94.1%, 92.5%, and 96.2% of Fe relative to the amount measured in the untreated stained LSTs, respectively. Although Fe dissolution curves stabilized after 6–12 hours (Fig. S5), minor quantities of residual Fe after the treatments (24–48 hours) indicated that all de-staining reactions are incomplete. Significant (p<0.001) linear relationships existed between the log of residual Fe and post-treatment color (RGB) intensities of treated earlywood and latewood (Fig. 3d, e). However, such linearity markedly weakened for the delta RGB intensities, especially for the delta BI (p=0.087, Fig 3f), indicating that DBI of LSTs is insignificantly affected by the staining issue. Note that wood delta BI and DBI were not calculated exactly in the same way (see Table 2; delta BI is the averaged difference between BI of entire latewood and earlywood from all tree rings in a sample, while DBI is a tree-ring parameter which presents the difference between LBI and BI of entire earlywood for each tree ring).

3.3 Comparison of LBI and DBI against MXD chronology

LBI chronologies of the four retained treatments (MixA, MixB, MixC, and Control) diverged relative to the reference MXD chronology (Fig. 4a) prior to the year 1900 CE when the stained LSTs dominated the chronologies (Fig. S3c). Correlation analyses showed that coherence between LBI and MXD chronologies was only robust for the 10-year high-pass filtered data (r>0.89) and decreased at longer timescales (Fig. 4c). In contrast, DBI chronologies were very similar to the reference MXD chronology for all the four treatments (Fig. 4b). Correlations with MXD data were strong and stable among all frequencies tested (r>0.82) (Fig. 4d). In addition, few differences were found between the control DBI series and the chemically treated DBI data, although the colors of wood samples were visually distinct (Fig. 3a).

4. Discussion

4.1 Causes of stain

The significant relationships between post-treatment wood RGB intensities and residual Fe (Fig. 3), along with the rapid post-sampling staining (Fig. 5b), support our hypothesis that oxidation of Fe is a major cause of stain in our lake subfossil material. Fe is abundant in natural aquatic systems as dissolved and particulate fractions (Bortleson and Lee, 1974; Davison, 1993; Nürnberg and Dillon, 1993). Briefly, dissolved Fe, mainly in Fe(II) state, is reduced and mobilized in porewater in the anoxic sediments. Dissolved Fe can migrate upward to oxic bottom water to form particulate Fe(III)-oxides (Davison, 1993; Davison et al., 1982). This cycle results in much higher concentrations of dissolved Fe in the anoxic sediments compared to the oxygenated freshwaters (Zaw and Chiswell, 1999). Soluble forms of Fe in the anoxic sediments can readily penetrate into the buried wood tissues. When buried LSTs are extracted from lakes, cut and exposed to air, dissolved Fe is rapidly oxidized to colored Fe-oxides (oxyhydroxides, hydroxides and more crystalline Fe(III)-oxides) which may bind to wood (Pelé et al., 2015). This process is also supported by the fact that fresh cuts from buried portions were heavily stained while exposed (but submerged) portions of the same LST were not (Fig. 5a). Photo-oxidation is assumed to be less likely, yet not improbable, because the stain contaminated both surface and inner portions of LSTs.

Furthermore, we found that amorphous and crystalline Fe-oxides are likely produced during the oxidation of dissolved Fe. About 63.4% of Fe was removed by the NaAsc reagent, although post-treatment colors of LSTs only slightly lightened (Fig. 3a, c). The neutral NaAsc, similar to buffered ascorbates, only extracts the most reactive amorphous Fe-oxides (Anschutz et al., 2005; Hyacinthe et al., 2006; Kostka and Luther, 1994). The less reactive crystalline phases can be removed by EDTA, HAsc, MixA, MixB, and MixC, each of which removed at least 25% more Fe than NaAsc. In fact, EDTA, ascorbic acid, and sodium dithionite, which are the active chemicals in these solutions, are known as useful extractants of both crystalline and amorphous Fe-oxides (Borggaard, 1982; Hyacinthe et al., 2006; Kostka and Luther, 1994; Tessier et al., 1979). The notable de-staining effect of these solutions (Fig. 4a) also implies that crystalline Fe-oxides are more color-reflective than amorphous ones. In our experiment, HAc extracted less Fe than NaAsc but with a better de-staining effect (Fig. 3). A probable explanation is that acid-soluble Fe-oxides extracted by HAc (Chester and Hughes, 1969; Gupta and Chen, 1975; Tessier et al., 1979) are in amorphous and crystalline phases which have stronger color reflectivity, whereas NaAsc only extracted less color-reflective amorphous Fe-oxides.

Other metal elements most likely have a negligible staining effect on LSTs compared to Fe. Our preliminary analyses demonstrated that, among 15 potential metal elements (including iron, manganese, chromium, cobalt, copper, lead, etc.), Fe was the only element present at high concentrations. Although manganese was relatively abundant in the samples (several times higher than other metals), its concentration was still approximately 20 times lower than Fe. In addition, we did not detect any copper and lead in our samples. On the other hand, Fe complexes bound to sulfur and phosphorus might also be responsible for the staining of LSTs in addition to the Fe-oxides. However, the MP-AES instrument is not sufficiently sensitive to verify our hypothesis regarding those Fe complexes (detection limits are ~6500 ppb and 125 ppb for dissolved sulfur and phosphorus, respectively, compared to ~4.6 ppb for dissolved Fe).

4.2 Chemical de-staining versus delta correction

The divergent trends of the three chemically treated LBI chronologies (MixA, MixB, and MixC) compared to the reference MXD chronology demonstrate that none of the de-staining treatments can generate satisfactory and robust LBI data (Fig. 4a). Although MixC is the most effective protocol resulting in very little residual Fe (<5% relative to the stained LSTs), the corresponding LBI chronology still displayed a significant long-term bias before the year 1830 CE (Fig. 4a), leading to a discrepancy of about +2 ºC compared to the MXD-based reconstruction (Fig. 6a). If no chemical treatment had been applied, the temperature discrepancy would be amplified to about +4 ºC (not shown). These errors are caused by the extreme color sensitivity of LBI values as direct measures of blue light reflectance. Therefore, any stain contributing to the wood color will strongly contaminate the LBI data, especially in the low-frequency domain (Björklund et al., 2014). These results discourage the use of LBI when staining is present in subfossils woods. By contrast, the high-frequency variability of LBI data seems unaffected by the Fe stain (Fig. 4b).

Unlike LBI data, DBI is unaffected by the Fe stain from annual to centennial timescales, which is shown by the high and stable coherence between DBI and MXD chronologies (Fig. 4b, d) and the non-significant linear relationships between the log of residual Fe concentrations and delta BI data (Fig. 3f). The linearity between DBI and MXD remained almost unaffected by de-staining treatments, for example as with the most efficient MixC protocol (Fig. S10). Furthermore, compared to the Control DBI chronology, no chemical treatments substantially improved the correlation of DBI with MXD data (Fig. 4d), resulting in nearly identical trends in the corresponding temperature reconstructions in comparison with MXD, except for some periods where tree replication is less than 10 (Fig. 6c, d). This evidence suggests that DBI is not only excellent to resolve the sapwood-heartwood color biases, but also efficiently resolves the low-frequency biases caused by the Fe stain in black spruce LSTs from the eastern Canadian boreal forest.

We observed that LBI as well as DBI diverged from the MXD data after the year 1950 CE at L20 (Fig. S8). The divergence of LBI is due to slight color differences between heartwood and sapwood of selected living trees although this issue is generally not serious for black spruce compared to pine or larch species (Rydval et al., 2014; Sheppard, 1999; Yang, 2007). Although DBI is theoretically sufficient to solve the sapwood-heartwood color issue (Björklund et al., 2014), in our case it could only partially correct this problem (Fig. S8). Old living trees were collected from lakeshore forests at the L20 site and they often displayed declining ring widths compared to healthy trees sampled later at the same site (not shown). DBI of L20 is likely influenced by these narrow tree rings (Björklund et al., 2019) because DBI of black spruce is not only correlated to MXD but also to the ring-width data (Wang et al., submitted). We thus speculate the divergence of DBI reflects mostly a specific issue related to the declining growth of unhealthy trees.

This study confirms the robustness of DBI data from stained black spruce LSTs. Yet, two points need to be considered for future DBI-based climate reconstructions. Firstly, a higher tree replication is often needed for DBI than MXD data, in order to obtain a robust chronology (Rydval et al., 2014; Wilson et al., 2019). Thus, it is not surprising that the control DBI-based reconstruction showed weaker verification statistics against instrumental temperature (Table S1) and some slight instability during poorly replicated periods with more variable EPS values (Fig. 6e). When replication was above 15 trees during 1901–1960 (Fig. S3c), the calibration r_2 was similar for DBI and MXD data against temperature (Table S1). Secondly, Björklund et al. (2014, 2015) suggested some multi-centennial biases in DBI data due to the heterogeneous wood color. Although this phenomenon is not obvious in our case regardless of the Fe stain, our reconstruction only spanned the last three centuries and further attention is needed to verify this potential bias.

5. Conclusion

Our study indicates that the simple delta correction of differentiating latewood and earlywood BI values is more effective to resolve the staining biases of BI data from LSTs than the much more complex and time-consuming chemical de-staining protocols tested here. DBI of black spruce LSTs is unaffected by the Fe stain from annual to centennial timescales and allows robust temperature reconstructions similar to MXD data. Consequently, DBI from stained black spruce LSTs is a promising proxy for developing millennial temperature reconstructions in the eastern Canadian boreal forests, a region with very few long MXD series (Wang et al., 2001). On the contrary, LBI is very color-sensitive and appears problematic in retaining the low-frequency climatic signal.

The chemical de-staining experiments, though not satisfactory regarding the robustness of LBI data, suggest that post-sampling chemical Fe oxidation most likely result in the staining issue. Since Fe is so abundant in the Earth's systems, our results may be representative of much wider regions. On the other hand, the excellent Fe extraction abilities (removal of >90% Fe) of three chemical mixtures also suggest that they, in particular the MixC, can be further used as part of Fe extraction protocols for waterlogged archeological artifacts which also face Fe-staining issues (Fors et al., 2014, 2012; Pelé et al., 2015; Zhang and Xi, 2003).

Data and sample availability. Instrumental temperature data is available at Climate Exploror: https://climexp.knmi.nl/start.cgi. Other data used for analysis are available at https://doi.org/10.5281/zenodo.3930493. The wood samples are stored at the University of Quebec in Rimouski.

Author Contributions. FW, DA, ÉB, SO and GC designed the experiments. SY, GC, AD and LW supported facilities and guided the experiments. FW, SY, SO and GC performed the experiments. FW and DA analyzed the data. FW and DA prepared the original manuscript with inputs from ÉB, GC, SO and AD.

Competing interests. The authors declare that they have no conflict of interest

Acknowledgments. This work was supported by the “PERSISTENCE” project funded by the Natural Sciences and Engineering Research Council of Canada (the NSERC-DRC program), Hydro-Québec, Manitoba Hydro and the Ouranos Consortium. FW was also funded by the China Scholarship Council. SY was supported by Key Laboratory Opening Subject of Xinjiang Uyghur Autonomous Region (2016D03005) and Basic Research Operating Expenses of the Central-level Non-profit Research Institutes (IDM201202). LW was supported by National Natural Science Foundation of China (41571094). We greatly thank Florent Vignola for his assistance in cutting wood blocks. Leila Jolicoeur, Nadège Trou-Kechout and Pauline Balducci helped during laboratory and fieldworks. We also thank professor Feng Chen and Tongwen Zhang for their permissions to perform tree-ring density measurements at Institute of Desert and Meteorology, China Meteorological Administration.

References

Anchukaitis, K. J., Wilson, R., Briffa, K. R., Büntgen, U., Cook, E. R., D’Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B. E., Hegerl, G., Helama, S., Klesse, S., Krusic, P. J., Linderholm, H. W., Myglan, V., Osborn, T. J., Zhang, P., Rydval, M., Schneider, L., Schurer, A., Wiles, G. and Zorita, E.: Last millennium Northern Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions, Quat. Sci. Rev., 163, 1–22, doi:10.1016/j.quascirev.2017.02.020, 2017.

Anschutz, P., Dedieu, K., Desmazes, F. and Chaillou, G.: Speciation, oxidation state, and reactivity of particulate manganese in marine sediments, Chem. Geol., 218, 265–279, doi:10.1016/j.chemgeo.2005.01.008, 2005.

Arseneault, D., Dy, B., Gennaretti, F., Autin, J. and Bégin, Y.: Developing millennial tree ring chronologies in the fire-prone North American boreal forest, J. Quat. Sci., 28, 283–292, doi:10.1002/jqs.2612, 2013.

Björklund, J., Gunnarson, B. E., Seftigen, K., Zhang, P. and Linderholm, H. W.: Using adjusted Blue Intensity data to attain high-quality summer temperature information: A case study from Central Scandinavia, The Holocene, 25, 547–556, doi:10.1177/0959683614562434, 2015.

Björklund, J., von Arx, G., Nievergelt, D., Wilson, R., Van den Bulcke, J., Günther, B., Loader, N. J., Rydval, M., Fonti, P., Scharnweber, T., Andreu-Hayles, L., Büntgen, U., D’Arrigo, R., Davi, N., De Mil, T., Esper, J., Gärtner, H., Geary, J., Gunnarson, B. E., Hartl, C., Hevia, A., Song, H., Janecka, K., Kaczka, R. J., Kirdyanov, A. V., Kochbeck, M., Liu, Y., Meko, M., Mundo, I., Nicolussi, K., Oelkers, R., Pichler, T., Sánchez-Salguero, R., Schneider, L., Schweingruber, F., Timonen, M., Trouet, V., Van Acker, J., Verstege, A., Villalba, R., Wilmking, M. and Frank, D.: Scientific Merits and Analytical Challenges of Tree-Ring Densitometry, Rev. Geophys., 57, 1224–1264, doi:10.1029/2019RG000642, 2019.

Björklund, J. A., Gunnarson, B. E., Seftigen, K., Esper, J. and Linderholm, H. W.: Blue intensity and density from northern Fennoscandian tree rings, exploring the potential to improve summer temperature reconstructions with earlywood information, Clim Past, 10, 877–885, doi:10.5194/cp-10-877-2014, 2014.

Borggaard, O. K.: Selective extraction of amorphous iron oxides by EDTA from selected silicates and mixtures of amorphous and crystallline iron oxides, Clay Miner., 17, 365–368, doi:10.1180/claymin.1982.017.3.09, 1982.

Bortleson, G. C. and Lee, G. F.: Phosphorus, iron, and manganese distribution in sediment cores of six Wisconsin lakes, Limnol. Oceanogr., 19, 794–801, doi:10.4319/lo.1974.19.5.0794, 1974.

Briffa, K. R. and Melvin, T. M.: A Closer Look at Regional Curve Standardization of Tree-Ring Records: Justification of the Need, a Warning of Some Pitfalls, and Suggested Improvements in Its Application, in Dendroclimatology: Progress and Prospects, edited by M. K. Hughes, T. W. Swetnam, and H. F. Diaz, pp. 113–145, Springer Netherlands, Dordrecht., 2011.

Bunn, A. G.: A dendrochronology program library in R (dplR), Dendrochronologia, 26, 115–124, doi:10.1016/j.dendro.2008.01.002, 2008.

Campbell, R., McCarroll, D., Loader, N. J., Grudd, H., Robertson, I. and Jalkanen, R.: Blue intensity in Pinus sylvestris tree-rings: developing a new palaeoclimate proxy, The Holocene, 17, 821–828, doi:10.1177/0959683607080523, 2007.

Chester, R. and Hughes, M. J.: The trace element geochemistry of a North Pacific pelagic clay core, Deep Sea Res. Oceanogr. Abstr., 16, 639–654, doi:10.1016/0011-7471(69)90064-3, 1969.

Davison, W.: Iron and manganese in lakes, Earth-Sci. Rev., 34, 119–163, doi:10.1016/0012-8252(93)90029-7, 1993.

Davison, W., Woof, C. and Rigg, E.: The dynamics of iron and manganese in a seasonally anoxic lake; direct measurement of fluxes using sediment traps, Limnol. Oceanogr., 27, 987–1003, doi:10.4319/lo.1982.27.6.0987, 1982.

Environment Canada: Canadian climate normals or averages 1981-2010, available at http://climate.weather.gc.ca/climate_normals/index_e.html, last access: 14 March, 2020.

Esper, J., Frank, D. C., Wilson, R. J. S. and Briffa, K. R.: Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, Geophys. Res. Lett., 32, doi:10.1029/2004GL021236, 2005.

Esper, J., Düthorn, E., Krusic, P. J., Timonen, M. and Büntgen, U.: Northern European summer temperature variations over the Common Era from integrated tree-ring density records, J. Quat. Sci., 29, 487–494, doi:10.1002/jqs.2726, 2014.

Fors, Y., Jalilehvand, F., Damian Risberg, E., Björdal, C., Phillips, E. and Sandström, M.: Sulfur and iron analyses of marine archaeological wood in shipwrecks from the Baltic Sea and Scandinavian waters, J. Archaeol. Sci., 39, 2521–2532, doi:10.1016/j.jas.2012.03.006, 2012.

Fors, Y., Grudd, H., Rindby, A., Jalilehvand, F., Sandström, M., Cato, I. and Bornmalm, L.: Sulfur and iron accumulation in three marine-archaeological shipwrecks in the Baltic Sea: The Ghost, the Crown and the Sword, Sci. Rep., 4, 4222, doi:10.1038/srep04222, 2014.

Frank, D. and Esper, J.: Characterization and climate response patterns of a high-elevation, multi-species tree-ring network in the European Alps, Dendrochronologia, 22, 107–121, doi:10.1016/j.dendro.2005.02.004, 2005.

Gennaretti, F., Arseneault, D. and Bégin, Y.: Millennial disturbance-driven forest stand dynamics in the Eastern Canadian taiga reconstructed from subfossil logs, J. Ecol., 102, 1612–1622, doi:10.1111/1365-2745.12315, 2014a.

Gennaretti, F., Arseneault, D. and Bégin, Y.: Millennial stocks and fluxes of large woody debris in lakes of the North American taiga, J. Ecol., 102, 367–380, doi:10.1111/1365-2745.12198, 2014b.

Gennaretti, F., Arseneault, D., Nicault, A., Perreault, L. and Bégin, Y.: Volcano-induced regime shifts in millennial tree-ring chronologies from northeastern North America, Proc. Natl. Acad. Sci., 111, 10077, doi:10.1073/pnas.1324220111, 2014c.

Grudd, H., Briffa, K. R., Karlén, W., Bartholin, T. S., Jones, P. D. and Kromer, B.: A 7400-year tree-ring chronology in northern Swedish Lapland: natural climatic variability expressed on annual to millennial timescales, The Holocene, 12, 657–665, doi:10.1191/0959683602hl578rp, 2002.

Gupta, S. K. and Chen, K. Y.: Partitioning of Trace Metals in Selective Chemical Fractions of Nearshore Sediments, Environ. Lett., 10, 129–158, doi:10.1080/00139307509435816, 1975.

Harris, I., Jones, P. D., Osborn, T. J. and Lister, D. H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623–642, doi:10.1002/joc.3711, 2014.

Helama, S., Melvin, T. M. and Briffa, K. R.: Regional curve standardization: State of the art, The Holocene, 27, 172–177, doi:10.1177/0959683616652709, 2017.

Hyacinthe, C., Bonneville, S. and Van Cappellen, P.: Reactive iron(III) in sediments: Chemical versus microbial extractions, Geochim. Cosmochim. Acta, 70, 4166–4180, doi:10.1016/j.gca.2006.05.018, 2006.

Kaczka, R. J., Spyt, B., Janecka, K., Beil, I., Büntgen, U., Scharnweber, T., Nievergelt, D. and Wilmking, M.: Different maximum latewood density and blue intensity measurements techniques reveal similar results, Dendrochronologia, 49, 94–101, doi:10.1016/j.dendro.2018.03.005, 2018.

Kostka, J. E. and Luther, G. W.: Partitioning and speciation of solid phase iron in saltmarsh sediments, Geochim. Cosmochim. Acta, 58, 1701–1710, doi:10.1016/0016-7037(94)90531-2, 1994.

McCarroll, D., Loader, N. J., Jalkanen, R., Gagen, M. H., Grudd, H., Gunnarson, B. E., Kirchhefer, A. J., Friedrich, M., Linderholm, H. W., Lindholm, M., Boettger, T., Los, S. O., Remmele, S., Kononov, Y. M., Yamazaki, Y. H., Young, G. H. and Zorita, E.: A 1200-year multiproxy record of tree growth and summer temperature at the northern pine forest limit of Europe, The Holocene, 23(4), 471–484, doi:10.1177/0959683612467483, 2013.

McCarroll, D., Pettigrew, E., Luckman, A., Guibal, F. and Edouard, J.-L.: Blue Reflectance Provides a Surrogate for Latewood Density of High-latitude Pine Tree Rings, Arct. Antarct. Alp. Res., 34, 450–453, doi:10.1080/15230430.2002.12003516, 2002.

Nürnberg, G. K. and Dillon, P. J.: Iron budgets in temperate lakes, Can. J. Fish. Aquat. Sci., 50, 1728–1737, https://doi.org/10.1139/f93-194, 1993.

Österreicher, A., Weber, G., Leuenberger, M. and Nicolussi, K.: Exploring blue intensity-comparison of blue intensity and MXD data from Alpine spruce trees, TRACE–Tree Rings Archaeol. Climatol. Ecol., 13, 56–61, doi: 10.2312/GFZ.b103-15069, 2015.

Payette, S.: The range limit of boreal tree species in Québec-Labrador: an ecological and palaeoecological interpretation, Palaeoecol. Perspect. Veg. Can. - Festschr. Honour JC Ritchie, 79, 7–30, doi:10.1016/0034-6667(93)90036-T, 1993.

Pelé, C., Guilminot, E., Labroche, S., Lemoine, G. and Baron, G.: Iron removal from waterlogged wood: Extraction by electrophoresis and chemical treatments, Stud. Conserv., 60, 155–171, doi:10.1179/2047058413Y.0000000110, 2015.

R Core Team: R: A language and environment for statistical computing, available online at https://www.R-project.org/, 2018.

Rydval, M., Larsson, L.-Å., McGlynn, L., Gunnarson, B. E., Loader, N. J., Young, G. H. F. and Wilson, R.: Blue intensity for dendroclimatology: Should we have the blues? Experiments from Scotland, Dendrochronologia, 32, 191–204, doi:10.1016/j.dendro.2014.04.003, 2014.

Rydval, M., Loader, N. J., Gunnarson, B. E., Druckenbrod, D. L., Linderholm, H. W., Moreton, S. G., Wood, C. V. and Wilson, R.: Reconstructing 800 years of summer temperatures in Scotland from tree rings, Clim. Dyn., 49, 2951–2974, doi:10.1007/s00382-016-3478-8, 2017.

Sheppard, P. R.: Overcoming extraneous wood color variation during low-magnification reflected-light image analysis of conifer tree rings, Wood Fiber Sci., 31, 106–115, 1999.

Sheppard, P. R. and Wiedenhoeft, A.: An advancement in removing extraneous color from wood for low-magnification reflected-light image analysis of conifer tree rings, Wood Fiber Sci., 39, 173–183, 2007.

St. George, S. and Esper, J.: Concord and discord among Northern Hemisphere paleotemperature reconstructions from tree rings, Quat. Sci. Rev., 203, 278–281, doi:10.1016/j.quascirev.2018.11.013, 2019.

Tessier, A., Campbell, P. G. C. and Bisson, M.: Sequential extraction procedure for the speciation of particulate trace metals, Anal. Chem., 51, 844–851, doi:10.1021/ac50043a017, 1979.

Wang, F., Arseneault, D., Boucher, É., Galipaud Gloaguen, G., Deharte, A., Yu, S., Trou-kechout, N.: Temperature sensitivity of blue intensity, maximum latewood density, and ring width data of black spruce tree rings in the eastern Canadian taiga, submitted.

Wang, L., Payette, S. and Bégin, Y.: 1300-year tree-ring width and density series based on living, dead and subfossil black spruce at tree-line in Subarctic Quebec, Canada, The Holocene, 11(3), 333–341, doi:10.1191/095968301674769686, 2001.

Wigley, T. M. L., Briffa, K. R. and Jones, P. D.: On the Average Value of Correlated Time Series, with Applications in Dendroclimatology and Hydrometeorology, J. Clim. Appl. Meteorol., 23(2), 201–213, doi:10.1175/1520-0450(1984)023<0201:OTAVOC>2.0.CO;2, 1984.

Wilson, R., Rao, R., Rydval, M., Wood, C., Larsson, L.-Å. and Luckman, B. H.: Blue Intensity for dendroclimatology: The BC blues: A case study from British Columbia, Canada, The Holocene, 24, 1428–1438, doi:10.1177/0959683614544051, 2014.

Wilson, R., Anchukaitis, K., Andreu-Hayles, L., Cook, E., D'Arrigo, R., Davi, N., Haberbauer, L., Krusic, P., Luckman, B., Morimoto, D., Oelkers, R., Wiles, G. and Wood, C.: Improved dendroclimatic calibration using blue intensity in the southern Yukon, The Holocene, 29, 1817–1830, doi:10.1177/0959683619862037, 2019.

Wilson, R., Anchukaitis, K., Briffa, K., Büntgen, U., Cook, E., D'Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., Hegerl, G., Klesse, S., Krusic, P., Linderholm, H., Myglan, V., Peng, Z., Rydval, M., Schneider, L., Schurer, A., Wiles, G., Zorita, E.,: Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context, Quat. Sci. Rev. 134, 1–18, https://doi.org/10.1016/j.quascirev.2015.12.005, 2016.

Yang, K. C.: Growth ring contrast enhancement and the differentiation of sapwood and heartwood zones, Wood Fiber Sci., 19, 339–342, 2007.

Zang, C. and Biondi, F.: treeclim: an R package for the numerical calibration of proxy-climate relationships, Ecography, 38, 431–436, doi:10.1111/ecog.01335, 2015.

Zaw, M. and Chiswell, B.: Iron and manganese dynamics in lake water, Water Res., 33, 1900–1910, doi:10.1016/S0043-1354(98)00360-1, 1999.

Zhang, J., Xi, S.: Research on the decoloration of waterlogged bamboo slips. Sci. Conserv. Archaeol. 15, 37–42, 2003. In Chinese with English abstract.

Tables:

Table 1. Basic chemical properties of the seven chemical solutions. All solutions were diluted using de-ionized water with no trace of Fe concentration. Concentrations are in v/v for HAc and in w/v for other solutions. EDTA: Ethylenediaminetetraacetic acid. Values of pH were estimated using pH test papers.

Code	Chemical components	Chemical property	pH
NaAsc	2% sodium ascorbate	reduction	7
HAc	2% acetic acid	acidity	3
EDTA	2% disodium EDTA	chelation & acidity	5
HAsc	2% ascorbic acid	reduction & acidity	3
MixA	2% ascorbic acid + 2% disodium EDTA	reduction, chelation & acidity	4
MixB	2% sodium ascorbate + 2% disodium EDTA	reduction & chelation	7
MixC	2% sodium dithionite + 2% disodium EDTA	reduction, bleaching & chelation	7

Table 2. Definitions of wood color intensities and tree-ring parameters used in this study. The RGB intensities refer to the color intensities measured separately for the red (R), green (G), and blue (B) channels. The parameters No.1–3 are used to quantify wood colors while the parameters No. 4–6 are conventional dendrochronological parameters.

No.	Parameter	Definition
1	earlywood & latewood RGB intensities*	Mean R, G, and B intensities from all pixels of earlywood or latewood (see Fig. S6), averaged from all tree rings of each wood lath.
2	wood RGB intensities*	Mean RGB intensities averaged from earlywood and latewood RGB intensities.
3	delta RGB intensities	Earlywood RGB intensities subtracted from corresponding latewood RGB intensities.
4	LBI*	Mean blue intensity of 30% of the darkest pixels in latewood (Fig. S6).
5	DBI	Raw LBI (measured from 30% of the darkest pixels) subtracted from raw earlywood BI (measured from 100% of pixels), automatically derived from CooRecorder 8.0 for each tree ring.
6	MXD	The maximum value of measured tree-ring latewood density.

*: data were inverted by subtracting the raw data from a value of 256.

Figures:

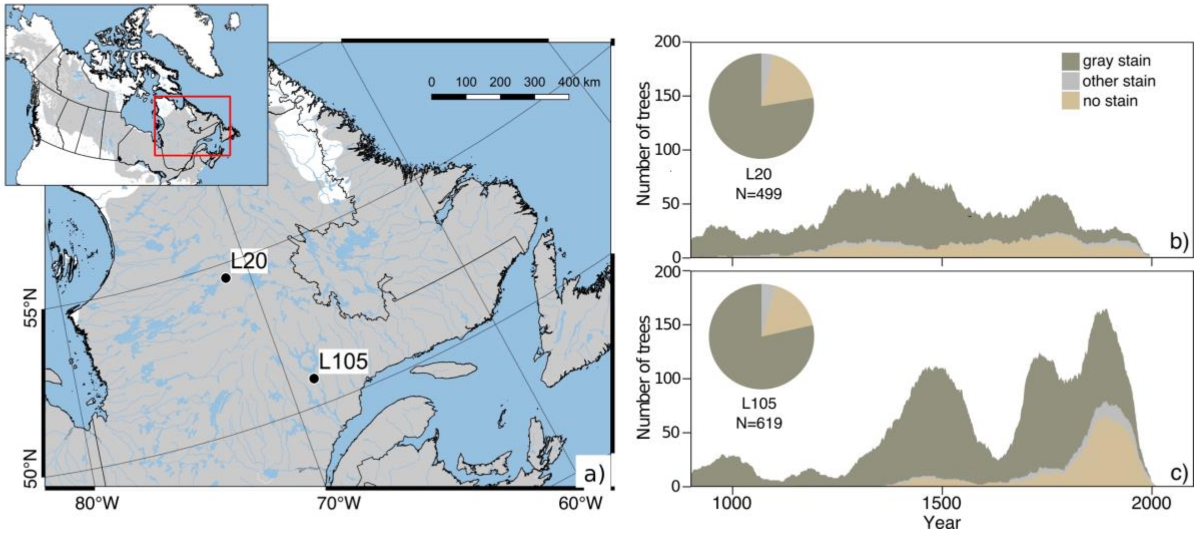


Figure 1. Location of the two studied lakes (a), and frequency of cross-dated LSTs according to staining at L20 (b) and L105 (c). The gray shading in (a) corresponds to the distribution range of black spruce, map source available at https://www.fs.fed.us/database/feis/pdfs/Little/aa_SupportingFiles/LittleMaps.html. The "other stain" category includes a variety of additional colors such as red or dark brown.

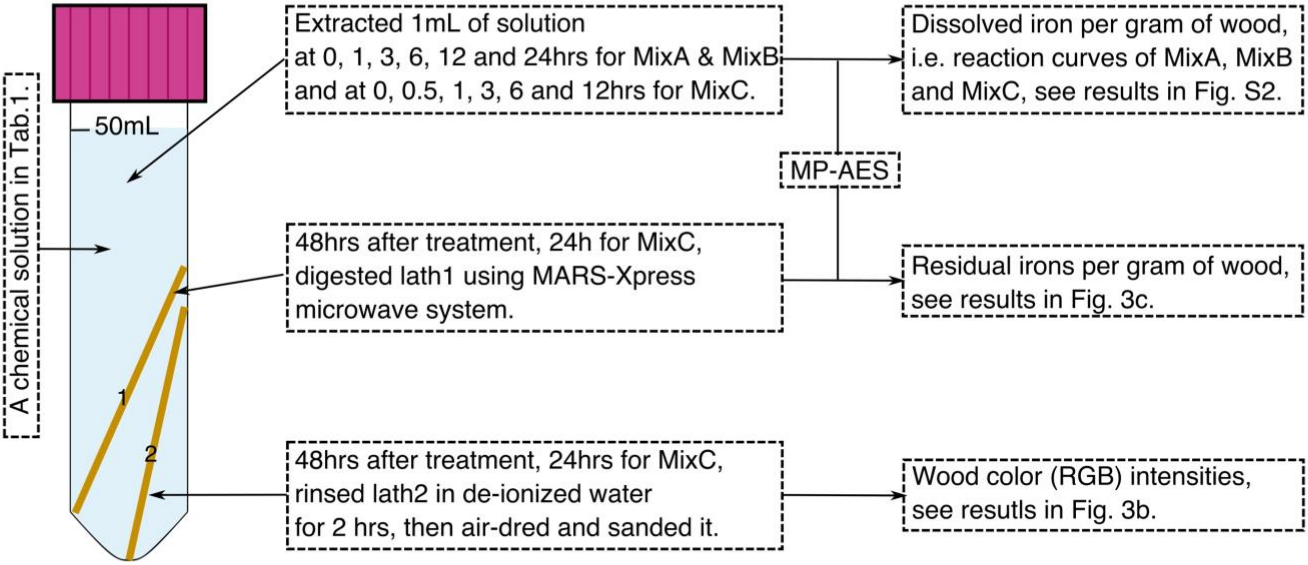


Figure 2. Diagram of one chemical de-staining experiment for one pair of wood laths from the same subfossil tree in a Falcon® 50mL tube. MP-AES: microwave plasma-atomic emission spectrometer, Agilent 4200.

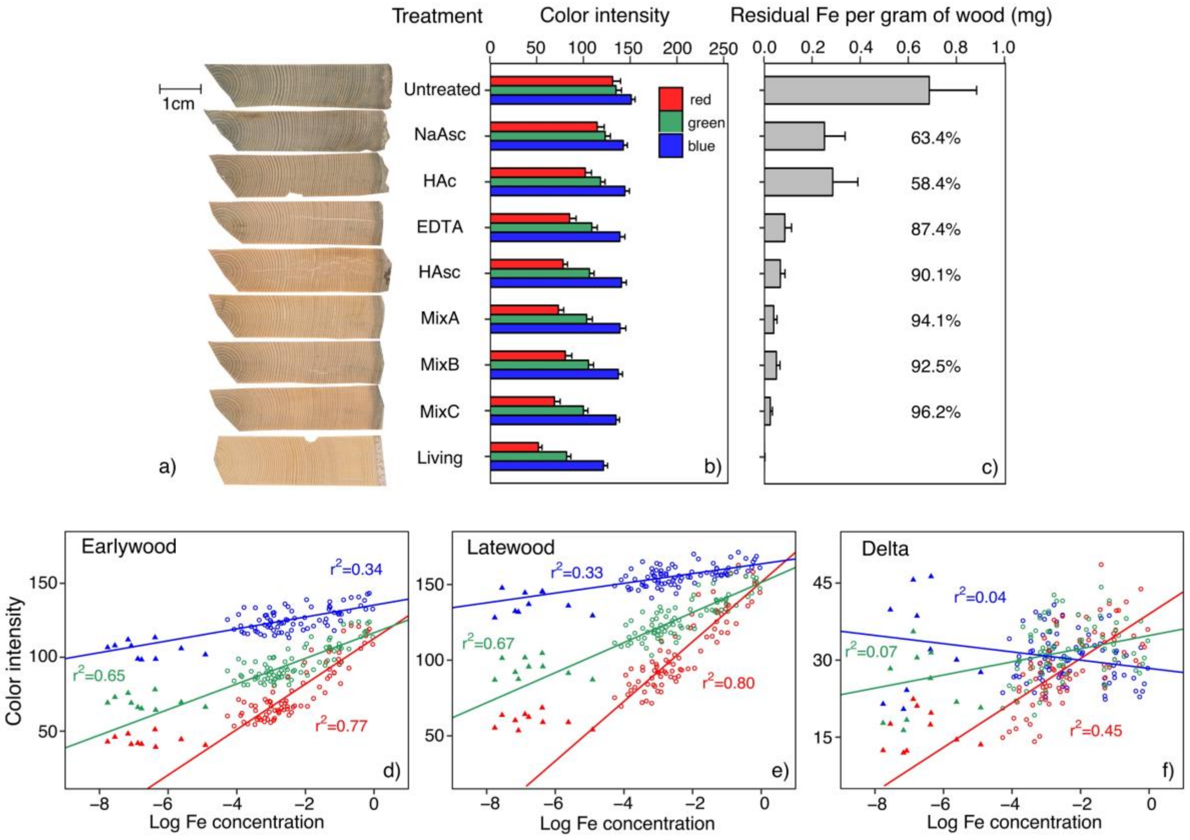


Figure 3. Residual iron (Fe) and wood RGB intensities (see definitions in Table 2) of LST laths treated with seven chemical reagents. (a) shows examples of treated laths from one LST sample and one living tree sample (last row). The gray outer part of the example LST is discolored due to decay. (b) shows the mean wood RGB intensities according to treatment. (c) shows the mean concentrations of residual Fe according to treatment. Error bars in (b) and (c) refer to standard deviations of corresponding group. Percentages in (c) refer to the Fe removed by de-staining treatments relative to the Fe concentrations of untreated stained LSTs. (d)–(f) show the linear regressions of earlywood, latewood and delta RGB intensities against the log of residual Fe. Regressions are based only on the LST data (circles). Living-tree data are plotted as triangles but are excluded from the regressions.

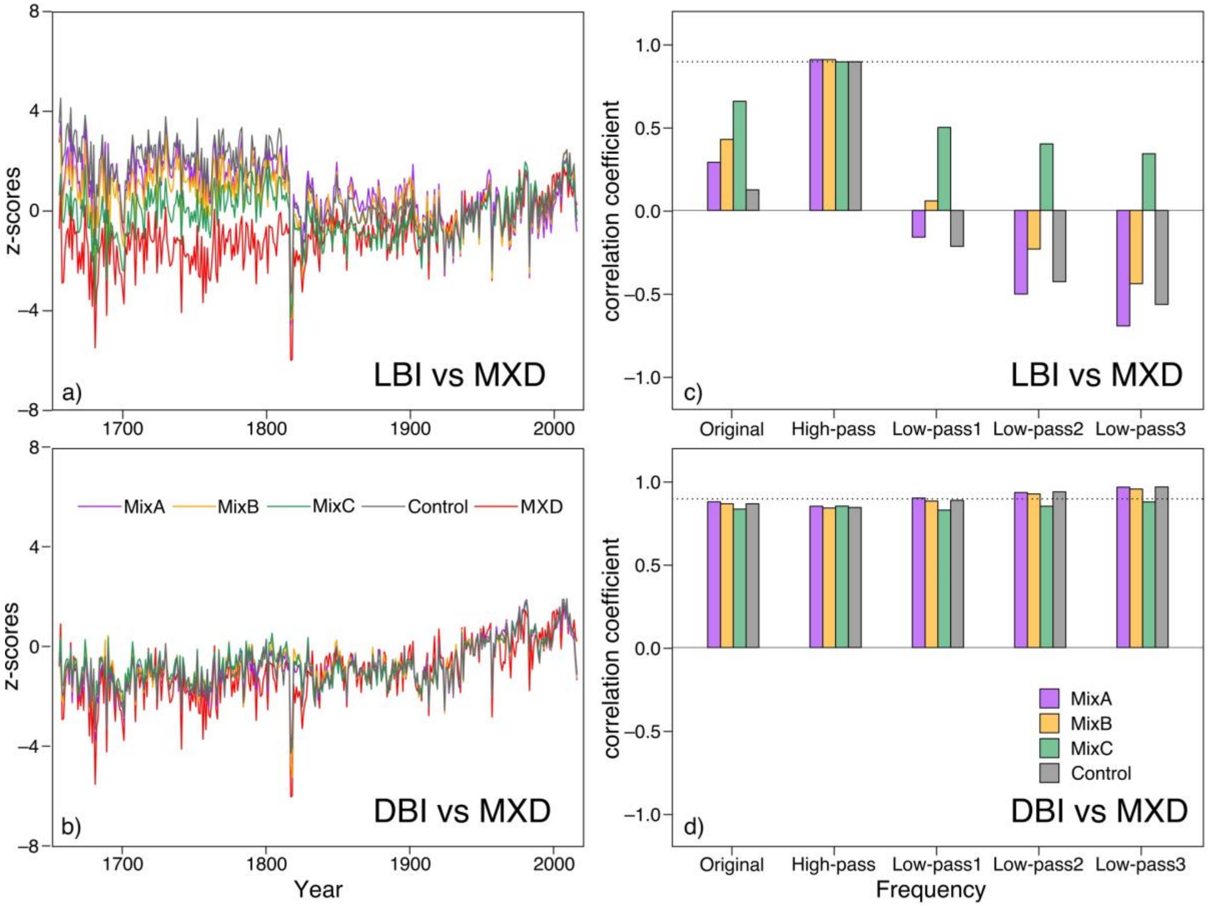


Figure 4. Comparisons of LBI (a, c) and DBI chronologies (b, d) for the MixA, MixB, MixC and Control treatments against the reference MXD chronology. In (a) and (b), chronologies are transformed to z-scores relative to the 1901–2015 time interval. (c) and (d) show correlations of LBI and DBI chronologies against MXD chronology at different timescales. Original: the original RCS standardized chronologies; High-pass: 10-year high-pass filtered series; Low-pass1, 2 and 3: 10-year, 50-year and 100-year low-pass series filtered using the Butterworth filter. Dotted horizontal lines in the right panel show the correlation (*r*=0.89) between the high-pass filtered LBI and MXD chronology.

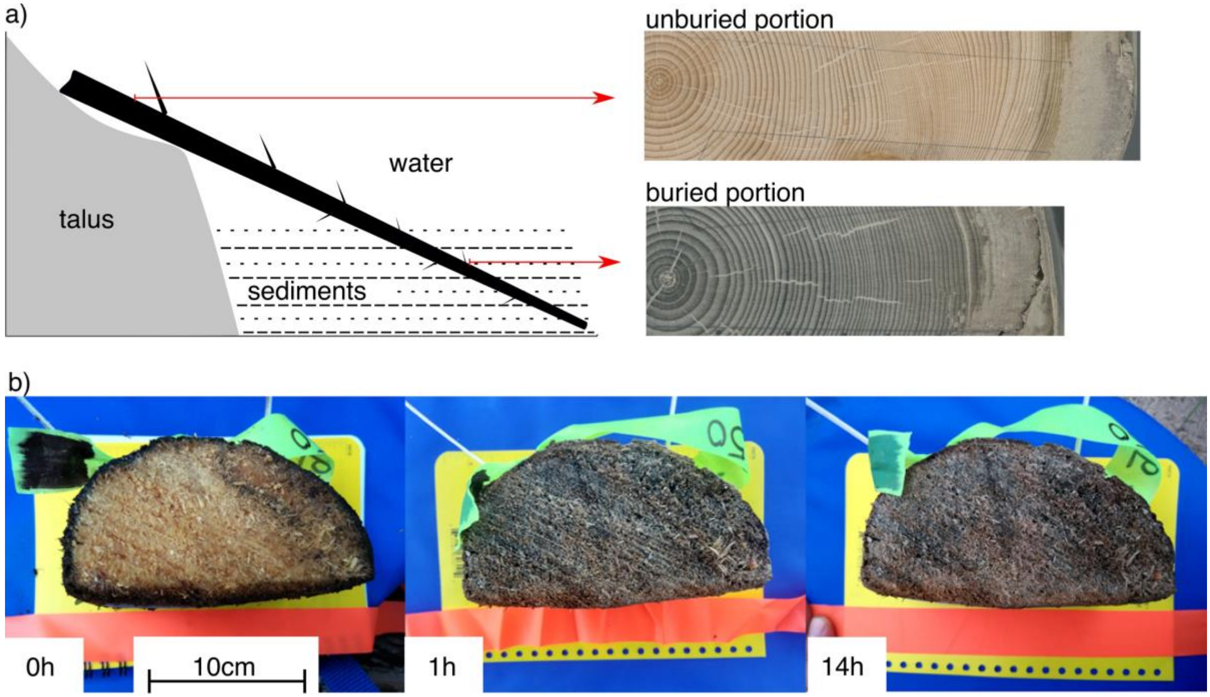
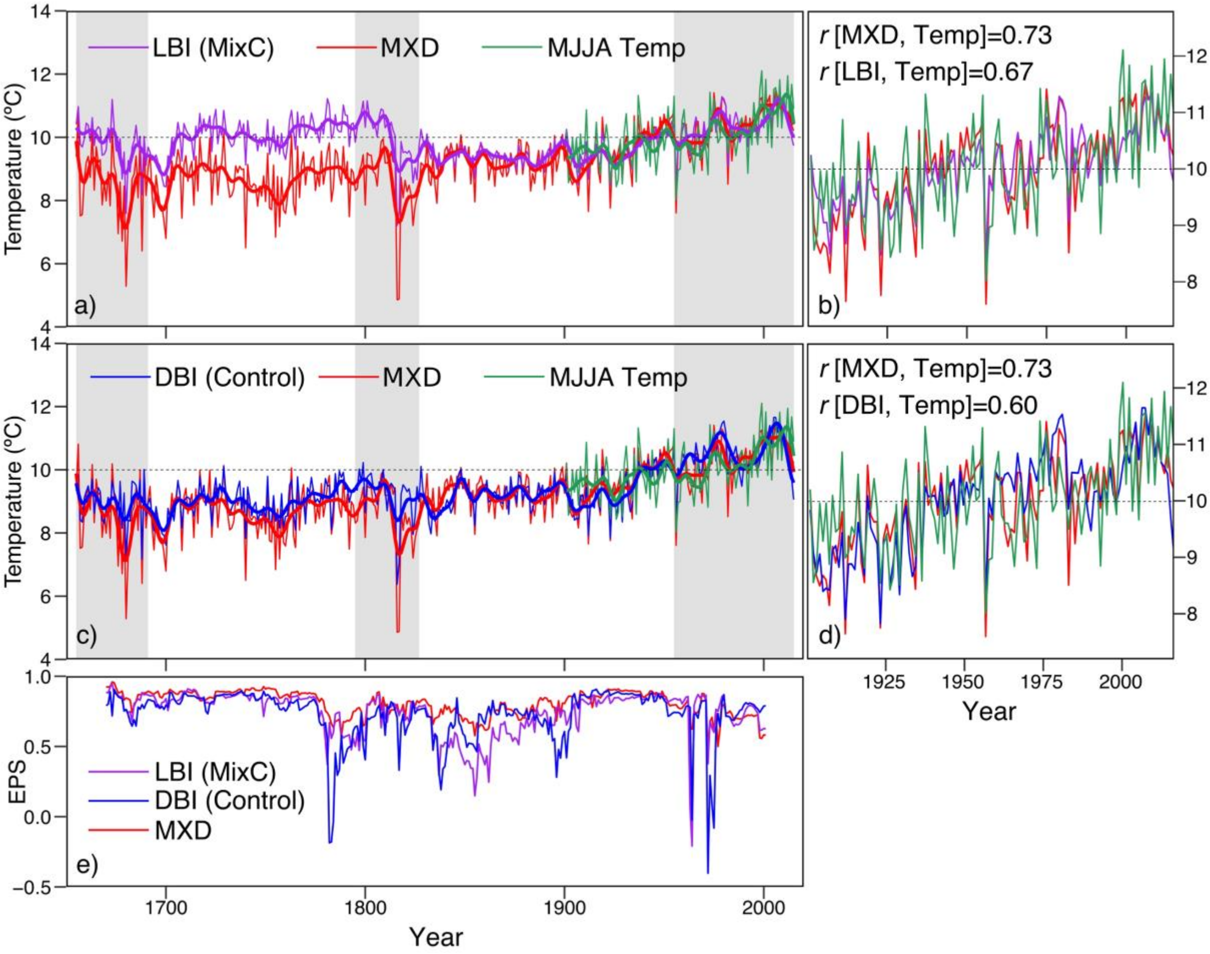


Figure 5. Schema of different cross-sectional colors from the buried and exposed cross sections of the same partially buried tree (a), and field observations of cross-sectional color changes after a fresh disc was cut from a buried tree and exposed to air (b).



490 **Figure 6. Temperature reconstructions using the LBI chronology for the MixC protocol (purple), the Control DBI chronology (blue), and the reference MXD chronology (red) for the 1655–2015 (a, c) and 1901–2015 (b, d) time intervals. (e) shows the 1-year-lag moving EPS computed in 31-year windows. EPS of the reference MXD was averaged from MXD chronologies of four different treatments (Fig. S9). Thick smooth lines denote the 10-year low-pass series filtered using the Butterworth filter. Vertical gray bars show the periods where tree replication is less than 10 (Fig. S3c).**