We would like to thank you for the considerate and helpful review. It has been very helpful in showing us which sections really spoke to the reader and the areas that need further clarification. Your original comments are reproduced below in bold to aid the reader, with our responses to each numbered bullet point in plaintext.

The manuscript submitted to Climate of the Past Discussions by Cecile et al. presents a novel method to consider and remove biases/unwanted noise from both age trends Sciences and tree productivity in tree-ring data by a technique dubbed fixed effects standardization. The age trend in tree-ring data represents a combined biological, geometric, and ecological noise that must be removed prior to using tree-ring data to reconstruct Ocean Science past climate variability. How this age-related signal is removed is arguably one of the Ocean Science Discussions larger sources of uncertainty in using climatically-sensitive tree-ring data to estimate a wide variety of environmental phenomena. Efforts to reduce these uncertainties and improve estimation of the age-related noise (and also tree specific noise) are significant research priorities. The manuscript and efforts by Cecile et al. are thus welcome and important contribution in this direction.

I find the manuscript is generally well written, but for better or worse employs unconventional structure (e.g. placement of the third section). The motivation, the scientific background for the proposed methods, their implementation (and some challenges thereof) are well described. Similarly, the ecological context is given a rather extensive treatment. Nevertheless, there are several aspects of the manuscript that I feel should be addressed prior to publication. In my opinion, after these major revisions this manuscript could make a nice contribution to Climate of the Past.

1. The results and conclusions regarding the sign tendencies of the sample bias in the large fraction of the ITRDB dataset require further support, particularly given this result contradicts essentially all past work on this topic. It seems that to reach this conclusion, the authors should demonstrate that the difference in the G=ITA versus G=TA models which they use to attribute changes in I, the individual tree growth, has not resulted in offsetting modifications of the T and A terms. Furthermore, I would have expected more discussion and exploration supporting these unexpected findings. The ~2 short paragraphs in the discussion does not seem to be appropriately balanced with the 4 pages of text in section three providing background on the likely source of this biases and its (now debated?) sign.

The lack of metadata and ecological knowledge about each chronology mean that it's challenging to explain the patterns shown by the ITRDB analysis in a way that moves beyond speculation. Our main purpose in this paper was to illustrate the new technique, including its ability to quantify modern sample bias. Hopefully the use of case studies (including those in the next version of the manuscript), along with future well-controlled meta-analyses will clarify its causes.

In general, it's hard to compare the results we obtained with previous studies as they used very different datasets. Notably, many of the ITRDB chronologies are poorly suited for regional curve-style standardization due to either variable age-dynamics or missing pith rings. Note though that Voelker et al. (2006) found occasional cases that exhibited similarly counter-intuitive trends.

2. The emphasis on statistical metrics to determine the applied model, while perhaps appropriate in theory, does not yet appear to be well guided in practice here. E.g., it seems odd to me that based upon these metrics, the age related trend, classically described as a age-dependent (e.g., negative exponential model), should be neglected in ~85% of cases based upon the AIC and in ~99% of the ITRDB datasets analyzed based upon the BIC (text on page 4519). Should we allow the BIC to give us "strange time-insensitive" tree-ring chronology? Should we defer to these statistical metrics to justify not removing age-related trends anymore? In the abstract the authors note "we can use powerful and transparent tools such as R2 and Akaike's Information Criteria to assess the quality of tree ring standardization allowing for objective decisions between competing techniques." In the text, e.g., in pages 4512-4513 and 4518-2519 the authors somewhat subjectively (?) describe limitations and relative merits of these tools. More consideration is necessary.

Choosing an appropriate standardization technique is always challenging, as we discuss at length in our comments to Dr. Melvin. The challenge is that model fit statistics only describe in-sample predictive / explanatory power when our main objective is largely the reliability of the time signal. Of course, as we don't know what this time signal should look like ahead of time (for real data), it becomes difficult to assess how well our choices perform on that basis.

The metrics put forward are intended to supplement, rather than supplant, researcher expertise and judgement. In general, poor model fit statistics should be a warning that the data quality is poor or the model is inappropriate but ultimately the tools must be used to answer the scientific questions intended. We hope that our admittedly subjective description of their strengths and weaknesses, along with the "typical" values of published chronologies will help future dendrochronologists gain some perspective on the characteristics of their own data sets.

Part A: age-dependent effects
The age effect in our model is fundamentally akin to the regional curve, in that it describes a single common pattern of growth changes by age. Rejecting this model effectively suggests that regional curve standardization is less appropriate than the raw chronology. This is probably unsurprising; most of the chronologies in the ITRDB were not designed for use with RCS (due to issues with stand dynamics, extremely old / stressed trees, missing pith rings etc.) but instead used individual series standardization to remove the age trend. To confirm that the raw, undetrended chronology is a better choice, we would need to compare the relevant models for individual series standardization to the time-only fixed effects model.

In most cases, the time effect is essential and the decrease in bias from removing age and/or individual effects will be well worth the increased variance (random noise) in our time signal.

Part B: time-insensitivity
In the rare case that model selection suggests that time is not a useful explanatory variable in modelling the growth of the trees in your chronology, it prompts further questions about the nature of the data set and its suitability to dendrochronology. It may be under-replicated, poorly cross-dated, largely insensitive to climatic factors or contain a mixture of temporal signals. Note that this was a very rare occurrence, using only the most aggressive of model-selection criteria, suggesting that it will rarely be an issue in practice.

3. Overall, I find the selection of figures could be improved, and should be expanded upon to provide deeper insights into the proposed methods. My take on the figures is: figure 1 is fine but could also be moved to the supplement or later in the text; figures 2 and 3 are useful; figure 4 is trivial; figures 5 and 6 do not really depict well (at least on my computer screen and printouts) the differences in selection percentage (but see point 2 above); figure 7 and particularly 8 are useful and interesting (but see point 1 above). I look forward to seeing some more specific examples as already indicated by the authors in the online discussion, and hope that these and additional new illustrations/analyses (e.g., testing methods on datasets composed only of living trees, providing a more detailed assessment of possible growth rate biases) will provide insights on the methods and conclusions.

This is an incredibly helpful comment. We will clarify figures 5 + 6, likely by changing the scale / legend or adding color and remove Figure 4 (which is really better described using text alone).

4. The authors do not really address the abilities to retain the long-term climate signal with this method. It has been traditionally (and by traditionally I mean following the widespread use of RCS) viewed that it is necessary to preserve information about the absolute growth rates of trees growing in different times to fully retain long-term climate variation. Can the authors please assess if these techniques overcome, in practice, the "segment length curse"? This may come down to the vagaries in how well the I, T, and A terms can be faithfully separated with the newly described methods.

Traditionally, the segment length curse only applies to individual series standardization approaches. As such, models that only include time and/or age effects shouldn't encounter any difficulty at all. The individual effect though, being analogous to flat detrending, should be susceptible.

This actually leads into some of our future work surrounding the segment length curse itself. Your intuition that the separation of effects is core to the segment length curse problem is correct. It turns out that when we remove effects sequentially, the segment length curse is present but it disappears completely when we fit each effect in our model simultaneously (using either fixed effects standardization / regression or signal-free standardization). The problem is actually equivalent to the "trend-in-signal" problem described by Melvin and Briffa (2008), with an unbalanced design in terms of trees vs. time causing the misfit, rather than age vs. time.

This is a fairly substantial claim, and will be addressed in detail (including applications to individual series standardization) in upcoming work. For now, the revised paper will simply include a basic sketch of the proof and then demonstrate the absence of segment-length curse using the revised techniques in the case studies already included.

5. In the discussion of signal free standardization, it would be helpful to clarify applicability to signal-free RCS versus signal free individual tree detrending.

Agreed, we need to make this distinction clearer. As presented, our techniques only truly apply to signal-free regional curve standardization. Individual tree detrending techniques could be incorporated into a regression / model framework if you could find and then fit the equivalent growth model.

6. I would appreciate if the authors would compare their standardization approach to “classical” detrending e.g., by negative exponential or spline fits to all series. The classical methods do in fact consider both the productivity of the individual tree and the age trend simultaneously. This seems to be rather close to what full mixed model (G=ITA) does?

We will show this in our case studies. As we will clarify in our next draft, the difficulty with “classical” detrending is that they while they purport to remove age and productivity effects, they can’t reliably differentiate them from true low-frequency climate variation. The strength of regional curve standardization, and thus fixed effects standardization, is that the age effect is not strictly collinear with the time effect, allowing for better separation.

7. How do the generalized cross-validation spline specifications compare with splines typically (flexible enough to remove the age-trend and rigid enough to retain longer term climate trends) applied in dendroclimatology?

Statistical literature suggests that generalized cross-validation (thin-plate) splines (GCV splines) tend to perform slightly better than other smoothing techniques (in terms of interpolation ability) when compared to traditional splines given a similar number of degrees of freedom (Hutchinson, 1995; Wood, 2001; Wood, 2003). Their major benefit comes from their relative insensitivity to starting conditions and explicit penalization of excessive flexibility. In general, the GCV splines tend to adjust to the characteristics of the data provided. A large sample size in each year, low intraannual variability and consistent changes in intra-annual growth produce a more locally flexible curve. The converse is also true, leading to a spline that is highly flexible in the early years of the “regional curve” and stiff towards the end as sample size declines. In individual series standardization, dendrochronologists must choose balance the removal of the age trend and the retention of the long-term climate trend, as you stated. If we chose to apply the GCV splines to individual series, they would be stiffer than in the case of the regional curve (no replication, more very high-frequency noise) but likely too stiff to retain much long-term climate trend at all.

Fortunately, our approach aggregates data about age effects across series, like in regional curve standardization. In general, we can extract good long-term climate trends using regional curve standardization style techniques without the use of smoothing splines at all. Smoothing the regional curve is useful in reducing overfitting the age trend, slightly improving our reconstruction of high-frequency temporal effects and increasing the reliability of sections with poor replication by age. For modern chronologies, a few very old trees will often provide the only data available on early periods of the chronology. But even in that case, the age effect for these early years is well-resolved; the benefits of the smoothed age trend are most pronounced in the modern / old data points where we typically have extensive replication in time using younger trees.

In principle, one could choose to use alternative forms of smoothing for the age trend such as more traditional cubic splines or local regression, all while retaining the basic regression model form. Such models are all forms of generalized additive models. While categorical effects models are simple enough that we could produce customized model-fitting code, we chose to adapt already constructed regression packages to fit the smoothed age effects model. We felt that GCV splines were a superior choice to cubic splines or local regression for this purpose, and decided to implement and present only one of the potential options as a proof of concept. Future studies should compare these alternatives to the unsmoothed model to guide their use in dendrochronological standardization.

8. In my opinion the text unnecessarily (inappropriately?) downplays, if not demeans, some prior work in what might be a natural effort to sell the results/methods presented in this study. The authors may wish to consider such aspects in their revisions. I feel the fixed effects standardization will promote interesting discussion and further innovation in this field. The authors may prefer to describe in greater detail the more general limitations of their proposed methods, rather than leave this to others. The techniques described here are a novel and thought provoking way to analyze tree-ring data, yet these techniques will likely still require refinement, if not more major evolution, in the next years.

Thank you for the comment, we certainly didn't mean to come across as downplaying existing work. The studies cited were genuinely informative and well-conducted and truly guided our thoughts on the subject. We'll do our best to highlight this. Please let us know if there are specific aspects of certain studies that you feel should be discussed in light of our findings after the next round of revisions.

Fixed effects standardization is by no means the final word in tree-ring standardization. As we will discuss in our conclusion, there are several outstanding concerns with our approach :

1. How do we decide which standardization approach to use?
2. How do we handle zero-width "missing rings"?
3. Are the fixed effects models correct? Does there need to be explicit consideration of size? Different patterns of growth by age?
4. Individual series standardization seems to work quite well in practice. How can we compare it to fixed effect standardization?

REFERENCES

Helama, S., Lindholm, M., Timonen, M., & Eronen, M. (2004). Detection of climate signal in dendrochronological data analysis: a comparison of tree-ring standardization methods. *Theoretical and Applied Climatology*, *79*(3-4), 239-254.

Hutchinson, M. F. (1995). Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information Systems*, *9*(4), 385-403.

Melvin, T. M., & Briffa, K. R. (2008). A “signal-free” approach to dendroclimatic standardisation. Dendrochronologia, *26*(2), 71-86.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366.

Voelker, S. L., Muzika, R. M., Guyette, R. P., & Stambaugh, M. C. (2006). Historical CO2 growth enhancement declines with age in Quercus and Pinus. *Ecological Monographs*, *76*(4), 549-564.

Wood, S. N. (2001). mgcv: GAMs and generalized ridge regression for R. *R News*, *1*(2), 20-25.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95-114.

.