Dear Alberto Reyes,

I have reviewed the revised manuscript with interest and acknowledge the great work that the author put in to take into account my comments and suggestions. The author presents varied evidence and a stimulating discussion of centennial to millennial scale variability that I think should be published in Climate of the Past once the minor points I raised below have been considered.

Additionally, I wish to apologize to the authors for the delay in my review as I was off academic duties the last few months (and am generally slow to review in addition).

Cordially,
Raphaël Hébert

Minor Points:
Line 170: Weird sentence bit “such as dominates”
Line 188: I don’t think smoothing is the correct term, because smoothing inter-annual variability can only lead to less variability and not create low-frequency one. Something like “integrate” or “accumulate” would be more appropriate. I would also avoid characterizing it as “white noise” since inter-annual variability is not generally white noise, albeit it can in fact be quite close to it especially in more continental regions.
Line 190: In this case I agree smoothing processes can generate artificial low-frequency variability by interaction with proxy processes, just not real one.
Line 212: I’m confused how the regional comparison is spanning eastern and central NA, whereas the sub-regional, which should be smaller, compares central and coastal NA, which appears similar to me as we are comparing central NA with either eastern NA or coastal NA.
Line 228: Typo “an potential”
Line 275: I’m not convinced that this is sufficient to avoid all biases, but I guess it should decrease potential biases.
Line 320: Provide reference for the “previously published confidence interval”.
Line 368: Unclear to me what is done here. We have data that is already binned/interpolated at 50- or 100-year resolution, so are the 50-year ones re-binned at 100-year by averaging two nearby points? Or are we using a Gaussian smoother to reinterpolate? I guess I don’t understand what is meant by we smooth and detrend at 100-year resolution.
Line 375: Were the surrogates also detrended with the Gaussian filter? I’m just thinking that this might bring down the correlations in the surrogate on the longest timescales on Figure 5 and potentially make some of the results more significant, e.g. for Figure 5B, the drop in correlation in the midlatitude series could be because the low-frequencies have been detrended whereas the surrogates were not and keep increasing. I’m not sure it would make a difference though as the filter is 6000 years and so maybe it doesn’t actually impact 2000-year timescales, but might be worth checking.
Section 1.2 I don’t see the difference between null distributions and null expectations, so maybe this section could be integrated with the previous discussion of null distribution, although I see that the two are serving different purposes, this section is rather an example that isn’t used for hypothesis testing but for a pedagogical demonstration, so it could also be titled accordingly and kept separate.

Line 385: Do null expectations aid detection? Or rather they allow to assess whether the detection is significant or not.
Line 596: Maybe it would be good to give the reconstruction uncertainties. It makes one wonder though what is the significance of this variability if the reconstruction uncertainties are higher. However, the errors on the reconstructions are generally not independent from each other in a given series. In the case of pollen reconstructions for example, we may take the RMSE of the calibration database as a measure of the errors related to the transfer function, but this will include a component that will be related to predicting the absolute temperature value and likely the same for all the samples, such that it will be taken out when detrending and looking at anomalies with respect to the trend. I don't know how we can separate the two types of errors though, I just wanted to note that the part of the reconstruction errors that is independent between the sample might not be so big compared to the anomalies, but again, I'm not sure how we can begin to separate them formally.
Line 629: I'm just wondering here whether the two datasets are truly independent, or whether there are records extracted from the same lakes or cores.
Table 1: Formatting could be improved, there is a lot of blank space in the table. I could see it reduced to 4 rows.
Figure 5: Typo in "Raeske", should be "Reschke"
Figure 6: I assume the slopes were fit over the entire range? One thing to think about is that the detrending is going to remove low-frequency variability, so it could be an idea to compare the detrended and undetrended spectra and see where the two diverge in order to see when the power loss occur and maybe not fit it. It depends what we want to measure with this, it the long-term trend is say a forced component that is removed with independent information, then we might say that the residuals represent the internal variability and that the result is the slope of the internal variability. In this case however, the trend removes everything, so the result is a fit of the real variability with a bias low from the long timescales dominated by the power loss.
Section 3.5 Is there no point in giving the correlation between the two groups of series?
Line 958: This is an interesting idea that NAO would manifest on longer timescales. Just a question of wording here, but I think it would be better to say "millennial-length phases dominated by either negative or positive NAO states" to make clear that it just means that one phase may occur more frequently than the other, but they would still both occur on inter-annual timescales right.
Line 1117: And here it could be written "negative- to positive-dominated NAO regimes" or something like that. I'm not sure what's the best way, but I think it would be useful to make a distinction given the different timescales involved.