

Dear Dr. Reyes,

I greatly appreciated the final review comments and have tried my best to make the appropriate changes throughout the manuscript.

Sincerely,

Bryan Shuman

Response to specific review comments:

Minor Points:

Line 170: Weird sentence bit “such as dominates”

- Changed to “such as is common on”

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.

- Replaced “smooth” with “integrate” and changed “white noise” to “stochastic”

Line 190: In this case I agree smoothing processes can generate artificial low-frequency variability by interaction with proxy processes, just not real one.

- I modified the wording to leave in “smoothing”, but also refer to “integrating”

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.

- I see how the wording was confusing here. I revised to refer to the ‘region’ as ‘mid-latitude North America’ and the sub-regions as “the central and northeastern (NE) subregions of mid-latitude North America”

Line 228: Typo “an potential”

- Corrected ‘an’ to ‘a’

Line 275: I’m not convinced that this is sufficient to avoid all biases, but I guess it should decrease potential biases.

- This makes sense. I tried to clarify some wording here, but I also don’t have a great way to fully avoid the biases. However, I have updated Fig. 7 to show that the patterns in the sub-sets developed using this approach usefully anticipate patterns in geochemical datasets from the two regions.

Line 320: Provide reference for the “previously published confidence interval”.

- The citations were added and the text clarified to explain that this refers to the 8 datasets examined throughout the paper.

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.

- Revised to clarify that the 100-yr intervals have been selected from all datasets for the smoothing and detrending analysis.

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.

- The surrogates were generated from the beta slopes of the detrended data, which I have clarified in the text.

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.

- I retitled the section "Demonstration of random effects"

Line 385: Do null expectations aid detection? Or rather they allow to assess whether the detection is significant or not.

- Revised the wording to state that “Null expectations can help assess whether analyses show significant patterns.”

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.

- I have struggled with this same issue and have not been clear on how to best address it. However, I wanted to be clear here that the RMSEs usually used as the reconstruction uncertainty are large compared to the apparent signals. I have revised and included a sentence trying to summarize the issue.

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.

- I have added a sentence to clarify that they are truly independent datasets, derived from different sets of lakes. (In an earlier related study, Shuman et al., 2019, the comparison was made in the same lake, but using different cores from different water depths with different age models).

Table 1: Formatting could be improved, there is a lot of blank space in the table. I could see it reduced to 4 rows.

- I agree. I will address this issue as part of typesetting if the manuscript is accepted.

Figure 5: Typo in “Raeske”, should be “Reschke”

- I corrected the caption.

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, if 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.

- I added a sentence to the caption clarifying that “Slopes were fit over the full range of periods from 5000-400 years.” I appreciate the additional suggestion for exploring the patterns here but am not able to fully evaluate this issue within the scope and timeframe of the revision.

Section 3.5 Is there no point in giving the correlation between the two groups of series?

- I revised section 3.5 to incorporate the geochemical records now shown in Fig. 7 and have added a sentence about the correlations.

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.

- I am uncertain whether the changes represent changes in the frequency of NAO modes or millennial-scale anomalies analogous to, but much slower than, the interannual oscillation. I added a sentence to make this point.

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.

- Good suggestion. I modified the wording to reflect the potential time scales involved.