Re-Review of Muschitiello & Aquino-Lopez CPD

I'd like to thank the authors for their replies to my comments, which cleared up some of my questions but partly also reinforce some of my concerns. The additional analyses lend some support to the robustness of the results, yet, in the current version of the manuscript it is not clear how the synchronization is done exactly and whether it may be prone to biases.

A part of my confusion is stemming from the response of the authors to my previous comment on the underlying assumption of a linear relationship between speleothem and ice core data. I remarked, that equation 3 requires a linear relationship between the proxies. The authors argue in their response that this is overcome by standardizing the data to [-1,1] and assigning large uncertainties to the signal. However, the applied changes in the manuscript (L217-221) only outline that a non-linear adjustment of the timescales is facilitated by the method. This is of course obvious (and could be removed from the manuscript) but does not address my original question.

From the author's reply, I understand (but I am not sure because the manuscript and the reply are incoherent in this respect) that the standardization of the data is done separately for each 180-year segment. If so, this is not clearly stated in the manuscript (compare L253-257). Further, this raises additional questions.

1. A standardization for each segment would allow drastic changes in the relationship (regression slope) between ice core and speleothem data. It could lead to the large peak in the speleothem d18O data around 18-20 kaBP, which has the magnitude of a DO-transition, to be matched to some minor structure in the ice core record, which has no equivalent change there. What would be the reason for such a drastic change in the relationship? How valid is the assumption that this still reflects the same physical driver in both proxies? Allowing for a freely varying relationship between the proxies increases the likelihood of erroneously aligning signals that have no physical connection. I also disagree, that the standardization is better at handling non-linearity in the relationship between the proxies, since the full dynamic range of the proxies occurs at DO-onsets within decades which fits inside one standardization-segment and is thus, still treated as a linear relationship. Further, a standardization of each 180-year segment effectively corresponds to a 180-year high-pass filter, while the speleothem data has very little variability in this frequency band, especially during MIS-3.
2. A standardization of the record as a whole (as shown in the figures 4-6) on the other hand, leads to long periods of systematic differences between the records (MIS-2 but also during stadials of MIS-3). Because the method evaluates squared differences between the records (equation 3) this leaves the method prone to minimizing those differences instead of matching structures. To test this, I ran a test on artificial data (figure below), which are composed of a AR(1)-process (the exact same in both datasets) and different linear trends in both datasets, as seen in the real data. Standardizing both datasets (as a whole) to [1,1] and allowing for a linear timescale compression/stretching of one dataset by +/- 5% clearly leads to an erroneous compression of the timescale for the investigated segment. Increasing the uncertainty of the records does not fix this problem in contrast to the statement in the manuscript (L253-257). I understand that this would obviously be different when all segments of the data are jointly evaluated in the MCMC, but it is not clear whether this really avoids the problem as a whole. This effect may for example bias the inference during MIS-3 since the scaling leads to large difference between the records during stadials and small differences during interstadials (see figure 4) and I wonder whether this can explain part of the difference to the results by Corrick et al. (see also comment below).

This leads to several request for clarification/revision in the manuscript:

1. Please clearly indicate whether the standardization is done for each segment or for the record as a whole.
    a. If done for each segment: Please include a supplementary figure where you show the records after standardization of each segment. As it is now, the upper two panels of figures 4-6 are misleading. Furthermore, please discuss (incl. figure in SI) how variable the ratio of the scaling factors is over time (i.e., how variable is the assumed relationship between the proxies) and whether it can still be assumed that this reflects a common climatic process in both proxies.
    b. If done for the record as a whole, please show that the issue outlined under point 2 above is not affecting the synchronization (for example by analysing subsections of the data and standardizing those).
2. In both cases, I repeat my request for additional panels in figures 4-6 that show the correlation per segment before and after synchronization, to allow the reader to evaluate which signals are really driving the synchronization, and by how much the fit between the records is improved by synchronization. It is for example surprising that the synchronizations for the different datasets (NGRIP d18O, GIRP d18O, NGRIP Ca) is so similar in MIS-2 when these records have been shown to diverge during this period (see figure 2 but also Rasmussen et al. 2008 fig. 3, 10.1016/j.quascirev.2007.01.016). I am aware that it is not the correlation of records that is being evaluated, but it is an intuitive measure for the readers, and ultimately, a correlation of the signals is the fundamental reason why climate-wiggle matching is considered a valid tool in paleoclimatology. Further, this would illustrate how "continuous" the synchronization really is and where the transfer function is driven by the priors. If the synchronization is hinging on relatively few tie-points, then the title and main text need to be adjusted accordingly.

Another aspect that may need to be revised is the inclusion of the results by Corrick et al. Comparing the way how the authors present the results by Corrick et al. in figure 7 to the equivalent plot in the original publication (Corrick et al. figure S6) it appears that the authors exaggerate the uncertainties by Corrick et al. It is my impression, that they included the GICC05 uncertainty into this figure, which is irrelevant for this comparison. If this is done correctly, it becomes apparent that the results presented here, significantly disagree with those by Corrick et al. between 30-38 kaBP. This should be discussed as this is also the period where there seems to be a systematic disagreement to the match by Buizert et al. 2015, which the authors attribute to Buizert et al's use of Hulu-cave only. However, this argument would not hold for the results by Corrick et al. Further, as the authors state in L365ff, their synchronization of this period is largely driven by DO-onsets, which is similar to the estimates by Corrick et al. This difference may stem from a bias mentioned above (small differences between the records during interstadials, large differences during stadials). Alternatively, this disagreement may arise from the method trying to find a compromise between aligning GS-GI and GI-GS transitions, while not violating the counting error constraints? Please discuss. Again, additional panels with running correlations of similar would help evaluating this.

Specific Comments:

L71: "14C concentrations" – change to D14C which is not a concentration

L72: "ocean carbon inventories" – change to "radiocarbon inventories"

L86: Please include Adolphi et al. 2018 into the reference list, since we the main point of our work was to test the synchroneity of DO-events in speleothems and ice cores.

L110-114: These issues have nothing to do with the autocorrelation of cosmogenic radionuclides (d18O and Ca are autocorrelated as well) but are an artefact of analysing overlapping windows. Rephrase or delete.

L121: “when timescales reach their largest offset” – please add “according to cosmogenic radionuclides (Adolphi et al. 2018, Sinnl et al. 2023)”

L123: “first continuous” – previous transfer functions where also continuous, albeit based on selected tie-points and various ways to interpolate in between. Given that this method is likely also only driven by a limited number of tie-points it is not that different. Please adjust the formulation and possibly the title.

L129: “improve precision and accuracy” – how do you determine that your transfer is more accurate than previous versions? Please elaborate or delete.

L135: “three independent synchronization” – synchronization should be plural. However, change to “three synchronizations based on independent Greenland ice core climate proxy records” or similar. The synchronizations are not independent (always the same speleothem data).

L253-257: See major comments. Is the standardization done per segment? If yes, please clearly indicate.

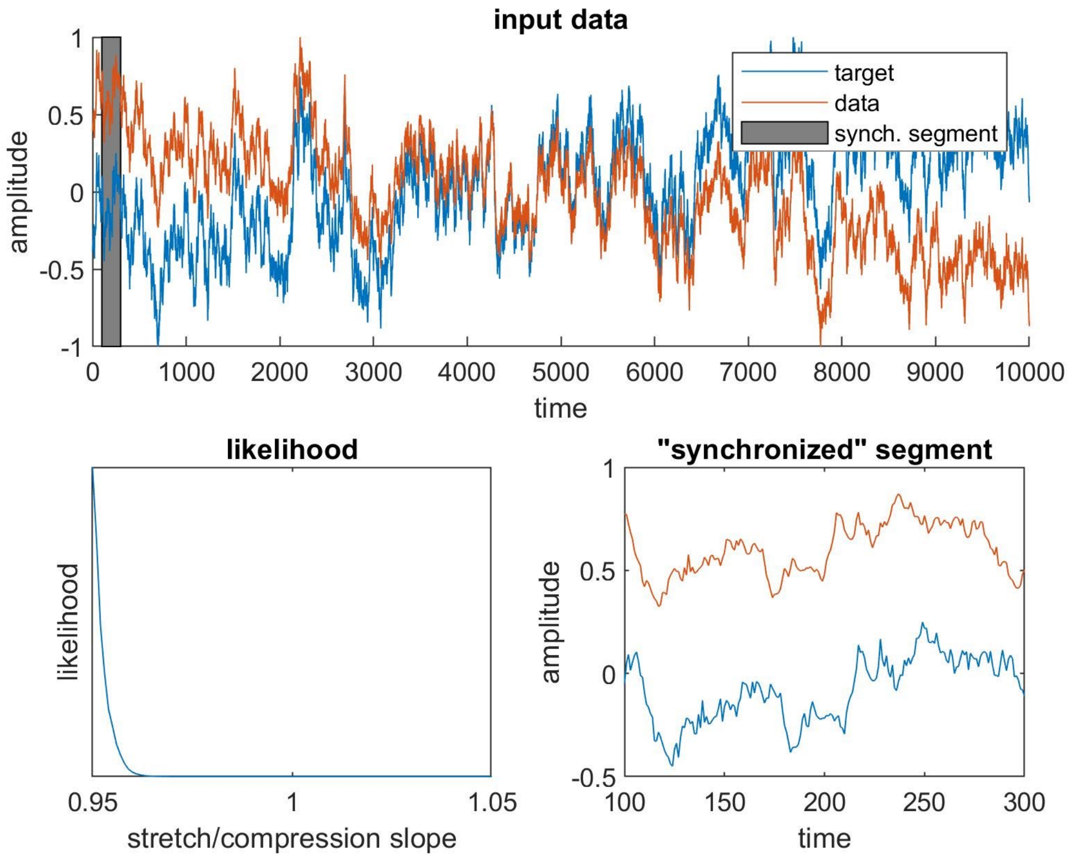
L265 (eq3): There is an error in this equation. The power of two only applies to the numerator of the last term of the equation (squared differences).

L269: “Any underestimation” – of what?

L277: See my previous comments. Shouldn’t tau0 be constrained by the MCE instead of the RCE?

L296: replace “uncertainties” with “credible intervals”

L346: Muscheler et al. 2008 inferred an offset of 65 not 55 years. Please change.


input data
target
data
synch. segment
amplitude
time
likelihood
likelihood
stretch/compression slope
"synchronized" segment
amplitude
time