We thank Reviewer1 for the constructive comments and detailed scrutiny of our manuscript. We agree that there are points that can be further clarified and we are happy to address the main and specific points that the Reviewer has brought up.

General comments:

[…] This study goes one step further, as it has the underlying assumption that all climate records are indeed the same (equation in line 228). Spelled out, this means that NGRIP Ca2+ = Speleothem d18O, which is obviously not true as they are very different physical quantities, controlled by different processes. Admittedly, some of the controlling processes may be shared, but the assumption of the applied method is much stronger: It implies the existence of a linear function that relates NGRIP Ca2+ and speleothem d18O. […] the basic assumption underlying the approach presented here (i.e., NGRIP Ca2+ = speleothem d18O) is incorrect and should hence not be used.

We appreciate the Reviewer raising this point, as it allows us to further clarify details of the method. The notation in the equation indicates an approximation (≈) rather than an equality. It is also important to note that this assumption is performed after scaling both data sets to have a strict range of [-1,1], which means $\forall\, t \in T : -1 \leq u(t) \leq 1 \wedge -1 \leq g(\tau(t)) \leq 1$, where T is the window range of the target (this will be clearly explained in the revision).

We agree that a nonlinear alignment between the records is difficult to capture with a single linear function. However, as described in the manuscript, we employ a piecewise linear function to model the alignment between the NGRIP Ca2+ and speleothem $\delta^{18}O$ records. A piecewise linear function is fundamentally different than a simple linear function, as it works by dividing the domain into K separate intervals and fitting a different linear segment within each interval. The linear segments are connected at breakpoints, where the slope can change. This enables the piecewise function to locally approximate nonlinear patterns by using linear segments that capture the essence of more complex functional forms in each region of the domain.

In fact, using this segmented piecewise linear technique provides several key advantages compared to a simple global linear function:

- It allows flexible modelling of nonlinear shapes by adapting the slope in each interval.
- Computationally efficient optimization algorithms can be applied by leveraging the linear segments. A purely nonlinear function would be more challenging to optimize.
- Accuracy can be systematically improved by adding more segments. The nonlinearity is approximated to any desired level.
- The approach balances accuracy with efficiency. More complex nonlinear functions could overfit given uncertainties in the data.

In summary, the piecewise linear methodology enables tractably approximating the nonlinear alignment relationship while facilitating optimization. We understand how these nuanced distinctions may have been unclear in the original manuscript, and we welcome this opportunity to explain our innovative approach more fully in the revised version of the manuscript (new lines 217-225, and 244-246).

If the authors nonetheless want to follow this approach, they need to i) clearly state that their model assumptions are not fulfilled…

Please see our detailed reply above. We discussed our approach in more detail in the revised version of the manuscript.

…and ii) discuss these drawbacks and provide additional tests to demonstrate the robustness of the results.

This is a good idea and we are happy to provide two new Δt transfer functions based on NGRIP and GRIP $\delta^{18}O$ records, respectively. The new results are internally coherent and support the findings obtained using NGRIP Ca2+, which ultimately lend strength to our conclusions. These results are incorporated and discussed in the new version of the manuscript (new Figures 2, 5, 6, 7, 8 and new Section 2.1 and 3).

1. What determines the inferred timescale shift during the LGM, when there is little co-variability between NGRIP Ca and EASM PC1 (see figure 4)?

Between 18-24ka there is little align-able structure in the timeseries and in fact the model sticks to the information obtained outside this interval, i.e. effectively the Δt does not move much, and the alignment uncertainty grows accordingly. This is to be expected and in line with the design of our alignment model.

2. Which timescale offset is inferred when only the period between 15 – 22 kaBP (or other subsections) is synchronized (and both records are standardized only for this period)?

Unfortunately, the method is not built to align very short timeseries with little structure and low signal-to-noise ratios. We hope that this issue is resolved by providing additional synchronization tests and targets using NGRIP and GRIP $\delta^{18}O$, which demonstrate that the Δt estimates are overall robust during this period. In addition, the Δt during this critical interval is corroborated by independent estimates published by Sinnl et al. (2023) and Dong et al. (2022) (see also R2's suggestions/comments). This is highlighted in the new version of the manuscript (new lines 349-351, and 393-395).

3. How would the results differ if NGRIP d18O was used instead of Ca2+?

Thanks for this suggestion. This is a sensible request and more in line with the premise of our manuscript, i.e. we show the physical relationship between Greenland air temperature and precipitation in the EASM region (see Fig. 1). The results are qualitatively consistent with those based on NGRIP Ca2+, which, again, demonstrates that our method is overall robust (please see our replies above). The new results are shown in Figures 5, 6, 7, 8 and discussed in the new sections 2.1 and 3.

4. How would the results (and uncertainties) differ if the uncertainty sigma_ui in the model was increased sufficiently to fulfil the model assumption (NGRIP Ca2+ = speleothem d18O within error).

Thanks for bringing this up. It should be noted that the input and target are scaled between -1 and 1, so by using a 0.1 stdev we are effectively covering 30% of the observable window, and on top of that we are using a heavy tailed distribution (t-distro) which means that we are assuming an uncertainty that fulfils the NGRIP Ca2+ = speleothem $\delta^{18}O$ assumption. We should also mention that we are employing overly conservative estimates for sigma using a multiplying factor of 2 (this is now clarified in the revised version of the manuscript, new line 253-257).

Further, the results need to be evaluated more critically with respect to previous studies:

1. Please include the timescale differences by Corrick et al. into figures 4 & 5.

We have included Corrick's Δt estimates in the new Figure 7.

2. It appears that most other studies (Buizert et al. / Corrick et al. / Martin et al) found systematically smaller timescale differences then the results presented here. Why?

The difference is marginal and Corrick's estimates are simply too uncertain to ascertain whether the offset is meaningful/systematic (see new Figure 7). As to the other studies, the small differences may stem from the fact that previous work used only Hulu Cave $\delta^{18}O$ data, whereas here we use a more comprehensive approach that incorporates several EASM spelothem records. In addition, the new synchronizations based on NGRIP and GRIP d18O suggest that the bias may be smaller than previously estimated, i.e. possibly 0.75% (new lines 338-341).

Specific comments:

L205 (eq. 1): Maybe I got this wrong but looking at this equations and trying to put in [units]: m must be [years/m]; so t must be [m] not time; so tau is defined on depth? If so, please use a different symbol as t is time later on.

We appreciate the reviewer raising this insightful question about the units in equation 1. It allows us to clarify that in our case, m is dimensionless, representing an expansion/compaction parameter in units of years/years. Meanwhile, $\tau 0$, δ, and ci have units of years. This is now clarified in the revised version of the manuscript (new lines 187, 208-209).

L180: "in response to changes in accumulation" since you're not modelling accumulation, maybe better "in response to miscounting"? L213-214: "…distinct depositional environments…" But you only model the ice core alignement and their accumulation rates are certainly autocorrelated? It is ok to do it like that, but I am not sure I agree with the explanation. L214-216: Isn't it that: You are not modelling the timescale (or ice accumulation) but only minor modifications of it (counting errors), which do not need to be autorcorrelated.

This is correct. Thank you for point this out. Instead of using the term "accumulation rates" we now discuss the model results in terms of compaction/expansion of the original timescale (new lines 189, 260).

L228: "u(ti) = g(tau(ti))" See major comments. This is obviously not true and should be discussed.

As we mentioned in our response to the previous question regarding equation 1, we agree it is crucial to use clear and consistent notation to avoid confusion. The reviewer is correct that "u(ti) = g(tau(ti))" is imprecise shorthand and could be misinterpreted. Nevertheless, we use the notation $u(ti) \approx g(\tau(ti))$, which means $\exists\ \tau(z) \mid u(zi) \approx g(\tau(zi))$. Note that we use ≈ and not =.

L244 (Eq. 3): This was defined for comparing 14C-dates to a 14C-calibration. I.e., similar physical quantities. Because your sigma_ui is too small to fulfil the model (u=g) the vast majority of the data is essentially treated as outliers in the gamma-distribution. See major comments.

It is important to note that $u(ti)$ and $g(\tau(ti))$ represent the NGRIP Ca^{2+} and aligned speleothem $\delta^{18}O$ records, respectively, after rescaling the data to the interval [-1,1]. This rescaling means that the uncertainty σ_{ui} used in the t-distribution becomes a conservative estimate around the rescaled record $u(zi)$. However, the use of the t-distribution has proven robust against outliers. Therefore, any data points that become outliers due to the rescaling assumptions do not significantly affect the resulting inferences of the alignment function $\tau(z)$. The t-distribution's heavy tails downweigh the influence of extreme values. In summary, rescaling the records to [-1,1] provides a simple standardized domain for comparing the data, while the t-distribution likelihood protects against artifacts from this transformation when inferring the optimal $\tau(z)$. This is now clarified in the new version of the manuscript (new lines 253-257).

L331-332: The agreement between the records is not very convincing. Please discuss critically. What is the correlation coefficient? What is the error of the model (u=g) after alignment?

We appreciate the reviewer's feedback on discussing agreement between the aligned records. However, we deem applying traditional metrics like a correlation coefficient unnecessary in our Bayesian alignment framework. The optimization process inherently identifies the maximum likelihood alignment given the uncertainties around each data point $u(ti)$ and $g(\tau(ti))$.

Specifically, in each step of inferring the posterior distribution for the piecewise linear function $\tau(t)$, the likelihood is calculated based on the t-distribution residuals between $u(ti)$ and $g(\tau(ti))$. The Bayesian approach thus quantitatively determines the optimal nonlinear alignment that maximizes the joint likelihood. Therefore, rather than introducing additional metrics, we believe the optimal uncertainties around $\tau(z)$ themselves demonstrate the credible alignment between the records.

Other specific comments:

L13-14: "which are currently not detectable…" Why wouldn't they be?
This has been re-worded.

L19-20: "a bias attributable…" The paper provides a reasonable discussion around this, but it is not conclusive. Please add "possibly" or similar.
This has been revised accordingly.

L37-38: "much smaller uncertainty in the absolute ages". During the glacial.
This has been revised accordingly.

L95-96: Please also mention the advantage, that this is a relatively low-level assumption, that only requires synchroneity and not a linear relationship as assumed by the model applied here. Further, the discrete tie-points have typically a high signal to noise ratio, while the method applied here, employs also low signal to noise variations for matching. Please be critical with the assumptions of your method.
We rephrased this paragraph and now acknowledge the high signal to noise ratio of abrupt proxy transitions (new lines 96-99).

L103-111: In principle, I agree with the problems of the alignment technique, but I am not sure what this has to do with (which?) autocorrelation. 10Be is autocorrelated over time - so are most climate and forcing records. The autocorrelation argument would also be true for pure 14C-wiggle match-dating. In my opinion, the crux is the window-length: If there is one large

peak within a window, it will dominate the obtained pdf as long as it is in the window. Hence, we used only non-overlappting windows in Adolphi & Muscheler 2016. But that obviously affects the resolution we can obtain, as we need a certain window length to have a signal.
We deem this sufficiently clear but we are happy to take on specific suggestions from the Reviewer as to how we can edit the text.

L117-119: This is not an issue of the alignment technique, but of the lack of convincing tie-points. In Adolphi et al. 2018 we only chose one tie-point around 21 kaBP which we called "tentative" but which forms the basis of much of what is discussed here. The lack of tie-points (or co-variability) is similar in this study. Looking at figure 4 there seems little agreement between the records. See major comments.
We deleted this comment accordingly.

L143-145: There are many processes that contribute to Ca deposition in Greenland. Please discuss in more detail.
We now acknowledge these additional processes this (new line 144).

L167-169: Maybe point out the advantages too: The assumption that the timing of a major climate transition synchronous is much more conservative than assuming a linear relationship between the proxies on all timescales which is clearly proven wrong during the LGM.
We edited the text accordingly (new line 173-174).

L180: "in response to changes in accumulation" since you're not modelling accumulation, maybe better "in response to miscounting"?
Thanks. This has been changed throughout the manuscript.

L181-182: "simulated ice core depositional history". See above, you're not modelling deposition but only the timescale.
This has been changed accordingly (new lines 189).

L205 (eq. 1): Maybe I got this wrong but looking at this equations and trying to put in [units]: m must be [years/m]; so t must be [m] not time; so tau is defined on depth? If so, please use a different symbol as t is time later on.
This has been addressed above (see specific comments).

L213-214: "…distinct depositional environments…" But you only model the ice core alignement and their accumulation rates are certainly autorcorrelated? It is ok to do it like that, but I am not sure I agree with the explanation.
Thanks. We removed this line to avoid confusion.

L214-216: Isn't it that: You are not modelling the timescale (or ice accumulation) but only minor modifications of it (counting errors), which do not need to be autorcorrelated.
That is correct. Thank you. We changed the wording throughout the manuscript to stress this out.

L228: "u(ti) = g(tau(ti))" See major comments. This is obviously not true and should be discussed.
This has been addressed. Please see our response above.

L244 (Eq. 3): This was defined for comparing 14C-dates to a 14C-calibration. I.e., similar physical quantities. Because your sigma_ui is too small to fulfil the model (u=g) the vast majority of the data is essentially treated as outliers in the gamma-distribution. See major comments.
This has been addressed. Please see our response above.

L257: “integrates” better “reflects” as this is the derivative of the MCE which may cause confusion.
Thanks. This has been changed.

L258-259: but that shift is absolute? Why is the RCE a good measure here?
Apologies for the confusion. The initial shift is in fact absolute. We edited the text accordingly.

L268-269: “exceeds the range allowed by the MCE (as is generally the case for the Holocene)”. This is not true. We also discuss, that the RCE is only exceeded very briefly. The exceeded MCE is inhereted from this early mistake. See figure 12 in Adolphi and Muscheler 2016.
Thank you. This statement has been deleted.

L318: 0.97 is 50% more than 0.63! Is that “comparable”?
The new results based on ice-core d18O are more in line with Buizert’s findings (now discussed in lines 338-339).

L321: Please compare the Delta T to Muscheler et al. 2008
This is now discussed in the main text (lines 345-346).

L322: “younger” within error this is consistent?
Apologies, we don’t understand this comment.

L331-332: See major comments. The agreement between the records is not very convincing. Please discuss critically. What is the correlation coefficient? What is the error of the model (u=g) after alignment?
This has been discussed in our reply above.

L334-335: “the error is large”. It appears that the error is actually smaller than during MIS-3?
This line has been edited accordingly.

L345: There seems to be quite some disagreement with Martin et al. 2023. Please discuss.
To avoid confusion due to comparing multiple independent timescales, we decided to remove these data entirely and only focus on the Δt between the U-Th and GICC05 timescales.

L346: Please include the re-assessment of the LGM tie-point by Sinnl et al., (2023) into the figures
This information has been added in the main text and on new Figure 7.

L361: Please also cite Sinnl et a. (2023)
This has been addressed.

Figure 4: Please include Corrick et al. 2023 tie-points
This has been addressed (see new Fig. 7).

##
##

We thank Reviewer2 for the supportive review and the constructive comments. We are happy to accept R2’s suggestions and meet all their requests.

agree that there are points that can be further clarified and we are happy to address the main and specific points that the Reviewer has brought up.

General comments:

[…] I generally agree to the finding of this and previous studies that there is some quite strong bias in the GICC05 layer counting for the 15-28 ka section that was fairly unconstrained at the time. In some sections, the bias appears larger than the stated MCE, and quite likely, the bias goes in both directions for different periods ending up at a close-to-correct absolute age for much of the 30-40 ka section. Still, I would think there is also the possibility that the U-Th stalagmite ages may sometimes have their accuracy issues although they are often published with very small error bars. Alone the observed scatter among different stalagmites covering the same events points in this direction. I think we have an example of this for the applied stalagmite records at around GI-10, where they 'exhibit some temporal inconsistencies' (Figure 4). Therefore, I would be careful to assume that all of the observed disagreement in absolute ages between the ice core and U-Th chronologies can be attributed issues related to the ice-core time scale(s). In any case, a long-term absolute error of about 1% is certainly much smaller than we thought it possible some 15-20 years ago, when GICC05 was put together.
This is a fair point and we agree with the Reviewer. We welcome this opportunity to tone down our claims and stress that the U-Th timescale (although absolute) may be problematic in certain intervals. We discuss this potential issue more openly in the revised version of the manuscript (new lines 383-386).

The following recent papers may be relevant to mention or discuss in the manuscript:

Dong et al., 2022, is concerned with GS-3 and introduces some accurately dated Asian stalagmites that allow for a detailed comparison of ice core and U-Th ages across that interval. The paper is supportive of the ice-core Ca/dust – Asian monsoon relationship for significant and abrupt climate events and it identifies biases of the ice-core chronologies in the same direction as the present manuscript although with somewhat smaller amplitudes.

Sinnl et al., 2023, identifies new 10Be bipolar links between G and A in the older part of the difficult GS-2 interval. The study is thus relevant for comparison in a similar way to that of Martin et al., 2023.
Many thanks for the suggestions. We now mention these studies in the revised version of the manuscript (new section 3.1 and 3.2) and present the data from Sinnl et al. in new Fig. 7. The offset estimate from Dong et al. (i.e. +320 years) is in good agreement with our new results. Our estimates integrated over the same 5-kyr period suggest a shift of +335 years for CLIM1, +255 years for CLIM2, and +240 years for CLIM3. The difference is mainly due to averaging the structure of our Δt transfer functions over five millennia. As for Sinnl et al., they estimated an offset of +375 years around 22 kyr b2k, which is remarkably similar to our Δt estimate of +390 years (mean of the three synchronizations; please see new Section 3.1).

Specific comments:

Lines 331-341: To test the robustness of the suggested similarity of the Greenland and East Asian records across GS-2 it may be an idea to apply a different Greenland record for the inversion algorithm. The NGRIP dust record is available in 5cm resolution, but it has rather poor quality when it comes to details. NEEM has available high-resolution records available for both Ca and dust concentrations. It may be worth trying to match the dust record and maybe the Ca using a log scale as the dust concentration varies exponentially with Greenland water isotopes (see attached figure).

We appreciate the Reviewer raising this point as it was brought up by R1 as well. This is a good suggestion and we will provide two new Δt transfer functions based on NGRIP and GRIP $\delta^{18}O$ records, respectively. As discussed in our replies to R1, the new transfer functions are consistent and overall support the results based on NGRIP Ca2+. These new findings are presented in the new version of the manuscript (please see replies to R1's comments). Also please note that the Ca2+ data have been log transformed before synchronization (see Fig. 2a).

Figure 6: In the attached figure, I compare the Sieben Hengste Cave (SHC) isotope record to the Ca and dust profiles of NGRIP and NEEM (all ice core records are on log scales). The SHC record is shown on its original time scale without application of the transfer function. Shown on those time scales, there appears to be a good correspondence between the ice core records and the SHC isotopes for the 22-28 ka period. In particular, the sharp transition associated with the onset of the younger of the Greenland dust spikes close to 24 ka and the adjacent structures seem to be well aligned between all records. Therefore, assuming there is a one-to-one relationship between ice-core dust/Ca and European stalagmite d18O, it appears that the transfer function makes things worse for this interval. If there are common events between the two records at around 18 ka, the transfer function may do a better job here?

We think that the SHC $\delta^{18}O$ is still marginally older than GICC05, although this is difficult to quantify with the naked eye. In particular, the structure around 28-30kaBP lends support to a systematically older U-Th timescale than GICC05. Estimating the offset using our methodology is an interesting suggestion but somewhat beyond the remit of this study. We believe this approach would better fit the scope of a follow-up project. Specifically, we are concerned that the SHC $\delta^{18}O$ data reflects a compound signal of atmospheric circulation that is not as physically straightforward (or, by all means sufficiently well understood) as for the EASM speleothem record.

Specific comments:

Lines 242-248: This may be a good place also to discuss the Sinnl et al., 2023, bipolar 10Be match points. Please also elaborate a bit on the relevance of the Martin et al, 2023, study. It may not be evident for the reader why the G – A synchronization is relevant in a context that is otherwise entirely NH.
Thank you for pointing this out. As explained in our response to R1, we decided to remove the data from Martin et al., to avoid confusion and comparing multiple independent timescales. Rather we prefer to focus entirely on the differences between the U-Th and GICC05 timescales.

The results from Sinnl et al. have been incorporated in the main text (please see our previous resplies) and presented in Fig. 7.

Line 361: Clearly, there is god agreement between the results of the present study and that of Martin et al., 2023, at around 18 ka in Figure 5, but for younger and particularly for older ages, there are large discrepancies, so what are the implications of this? Again, it may not be evident to the reader how Antarctica fits into the otherwise NH picture. The Dong et al., 2022, study could be relevant for this discussion.
Thank you. Please see our comment above. The results from Dong et al. have been incorporated in the main text.

Figure 3: A convincing comparison (although not surprising) but something must be wrong with figure titles or the caption. Should be right-hand figure be showing GS onsets? Not sure which reversed scale is referred to in the caption.
Thanks for pointing this out. We acknowledge that the time axis may be confusing. The figure has now been edited so that time plots right-to-left.

Figure 4 caption: Which blue line is referred to in caption of Figure 3c? In Figure 3d, I can also not distinguish the mentioned colors.
The figure and caption have been edited and we removed the MCMC chain.

Figure 6 caption: There seems to be some remains of previous versions of this figure in the caption? At least, I do not find the mentioned annual layer thickness profile in the figure.
Thank you. The caption has now been edited.