Authors' response accompanying minor revisions to Slattery et al. (2024) - CP/TC.

Referee #1 - Mathieu Casado

The authors would like to thank the referee for their insightful and helpful feedback throughout the review process. The referee's comments are in bold, and our responses follow in normal text.

This revision of the manuscript about simulating DO variability to evaluate leads/lags between climate parameters has been considerably improved. I suggest to accept the manuscript after taking into the following specific comments.

Specific comments:

Lines 61 to 65: "Whilst these model simulations are imperfect representations of real DO events, they nonetheless provide an invaluable means to help us investigate the question of whether it is possible to conclusively identify a trigger for rapid Dansgaard–Oeschger warming events. This is because, unlike ice-core proxies, they provide complete information about different components of the climate."

The second sentence seems unnecessary, if not inaccurate. First, there are other records to study past climate, including DO events than ice cores, which are ignored, which is surprising. Second, "complete information about difference components" is extremely vague. Third, GCM are not complete, otherwise there would be no need for observations. The first sentence was fine on its own, consider deleting the second one.

We have removed the second sentence quoted here.

Lines 66 to 68: "We build upon these recent advances, examining in detail the causes of uncertainty and the question of what can therefore be learned from paleo-archives about the onset time of DO events in different climate elements - or proxies for these - such as temperature, precipitation, atmospheric circulation, sea ice, and AMOC."

This sentence should be streamlined and simplified, or maybe even omitted.

We have streamlined this sentence to the following: “We build upon these recent advances, examining in detail the causes of uncertainty in the onset time of DO events in different paleoclimate proxies and model variables.”

Referee #2 - Sune O. Rasmussen

The authors would like to thank the referee for their insightful and helpful feedback throughout the review process. The referee’s comments are in bold, and our responses follow in normal text. Where the referee has quoted our previous responses / revisions, these are in bold italics.

The suggested revisions represent improvements that cover most of my concerns. I provide a few additional comments below (in normal typeface following the authors’ responses in bold), and in particular, I think the title should be revised to something closer connected to the conclusions. It is, for example, quite unclear what “reliably” means without the context provided in the text (which will hardly fit into the title): as discussed in the reviews and in the updated manuscript, the biases will often be too small to be of any practical importance unless data from many events are stacked or otherwise combined. Also, it’s not clear what the mentioned phasing refers to (between events, between proxies …), and “Rapid” is not needed as Dansgaard–Oeschger events always are rapid. Some suggestions to illustrate my point:
“Methodological biases hamper the detection of climatic leads and lags across Dansgaard–Oeschger events”
“Estimating biases during detection of leads and lags between climate elements/mechanisms across Dansgaard–Oeschger events”

We agree that the previous title does not reflect the conclusions well. We adopt a form of the referee’s second suggestion:

“Estimating Biases During Detection of Leads and Lags Between Climate Elements Across Dansgaard–Oeschger Events**”**

We also slightly adjust the abstract to reflect this change: “Dansgaard--Oeschger (DO) events occurred throughout the last glacial period. Greenland ice-cores show a rapid warming during each stadial to interstadial transition, alongside abrupt loss of sea ice and major reorganisation of the atmospheric circulation. Other records also indicate simultaneous abrupt changes to the oceanic circulation. Recently, an

advanced Bayesian ramp fitting method has been developed and used to investigate time lags between transitions in these different climate elements, with a view to determining the relative order of these changes. Here, we critically review this method in both its original implementation and a new, extended implementation. Using ice-core data, climate model output, and carefully synthesised data representing DO events, we demonstrate that both implementations of the method suffer from biases of up to 15 years. These biases mean that the method will tend to yield transition onsets that are too early. Further investigation of DO warming event records in climate models and ice-core data reveals that the biases are on the same order of magnitude as potential timing differences between the abrupt transitions of different climate elements. Additionally, we find that higher-resolution records would not reduce these biases. We conclude that decadal-scale leads and lags between climate elements across DO events cannot be reliably detected, as we cannot exclude the possibility that they result solely from the biases we present here."

Furthermore, for the NGRIP ice core proxies, our extended method finds slopes of comparable magnitude in the pre-ramp stadial and post-ramp interstadial. This will be made clear in our revised manuscript as all of the relevant parameters used to create the "analogous" synthetic parameters for the purpose of bias estimation will be listed in a table in the appendix. We most often find that the slope during the pre-ramp stadial is positive - that is to say in the same direction as the ramp itself. See also the mean parameters for each proxy (including slopes) shown on Additional Figure 2 (page 10). The proxies all show significant slopes, in the sense that these slopes lead to a significant bias, however we have not tested the statistical significance of these slopes in isolation.

Why not? The observation that allowing a pre-ramp slope changes the ramp location is not surprising in itself. If the observed pre-ramp slopes are not significant, it's not clear that (or how often) the updated model is an improvement. In particular, for non-model data, it remains to be demonstrated that there are significant slopes in the stadials and that the updated model is indeed an improvement and not 'just' another model.

We agree that this was not sufficiently demonstrated before. To put this beyond all doubt, we have explicitly tested the significance of the stadial slopes for the NGRIP data. We include the following as a supplementary figure in Appendix A, along with the paragraph quoted below:

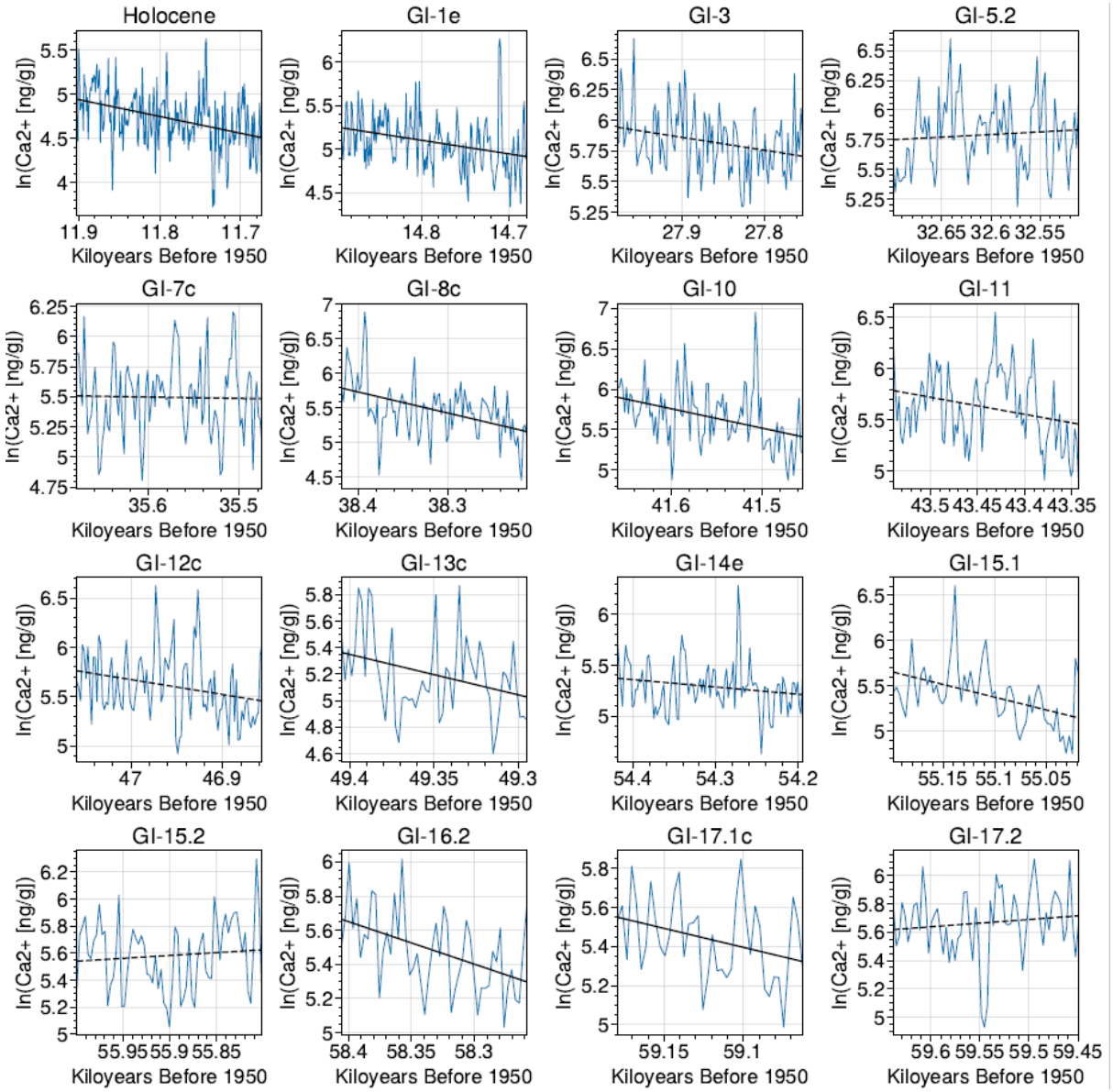

Figure A1: Slopes in the pre-transition stadial for NGRIP Ca2+. Seven of the sixteen events considered show significant downward slopes ($p<0.05$ assessed using phase-randomised Fourier surrogates).

“To test whether ice-core data show significant slopes in the stadial periods preceding DO events, we consider the data segments used by Erhardt et al. (2019). We only consider data up to the fifth percentile of the posterior distribution for the transition onset time, as given by Erhardt et al. (2019), to ensure that the abrupt transition itself does not influence our results. We also only consider the sixteen DO events where Erhardt et al. (2019) provide onset times for all four proxies. We calculate linear slopes and test the significance of these using phase-randomised Fourier surrogates, which preserve the autocorrelation structure of the data. We find that seven of the sixteen events show a significant slope ($p<0.05$) in Ca2+ that is in the same direction as the following abrupt transition (Figure A1), as well as four for

the Annual Layer Thickness and one for Na+. This demonstrates the need to consider the impact of pre-transition slopes on the accuracy of the ramp fitting method, and to test whether this can be improved by directly incorporating said slopes into the transition model. This is in addition to the more obvious need to consider post-transition slopes due to the classic "saw-tooth" shape of DO events."

Capron et al. test 20 realisations of a synthetic ramp, and find no significant evidence of a bias in either the transition midpoint or duration. Neither of these are directly comparable to the transition onset time, which we focus on as we feel it is more physically meaningful when trying to understand the progression of DO events. Nonetheless, if there were a bias in the onset time then Capron et al. would surely have seen this reflected in either the midpoint or duration, and they do not in fact see this. As the referee notes, the uncertainties shown in Supplementary Figure 2 of Capron et al. are large, and so could include a bias of a few years. Even so, the test conducted by Capron et al. would seem to rule out decadal-scale bias of the kind that we find in our study, at least for this particular combination of transition shape and noise. We would therefore suggest that the transitions tested by Capron et al. happen to lie in a region of parameter space for which the bias is small, even at high levels of noise. When using the original Erhardt et al. implementation we find very little bias for synthetic transitions with no pre- or post-ramp slopes (see Additional Figure 2 on page 10 of this response). …

I think this is a likely explanation. I guess it also means that the bias problem is expected to be relatively small for most of the data sets of transitions derived from ice cores. …

We are glad that the referee is satisfied by this explanation. As shown in Table 3, for the four different NGRIP proxies we estimate early biases between five and ten years when using the original Erhardt et al. Method. Whether or not these are "relatively small" is subjective, as it entirely depends on what they are being compared to. Nonetheless, we feel that these biases are sufficient that they could potentially lead to false conclusions regarding leads and lags between different proxies, and so we feel that it is important to document the presence of these biases and warn the community of this possibility.

Although the shape of the deterministic ramps used by Capron et al. is not made clear, it is likely that they are flat before and after the ramp. …

Indeed, this is the case, and this should be clear from Capron et al.: "i.e., a linear change in the raw or logarithmically-transformed data between two stable states"

…Instead, we demonstrate that their finding depends strongly on the unstated assumption that there are no pre- or post-ramp slopes. …

Citing Capron et al.: "The assumption that the transitions are adequately described by a linear change from one stable state to another is not trivial and has been challenged previously, but neither our observations nor the current understanding of the nature of the transitions justifies employing a model with more degrees of freedom." The authors are very welcome to challenge this assumption, but it is hardly unstated.

We apologise for failing to notice that Capron et al. do in fact spell this out clearly, and we thank the referee for drawing it to our attention.

Line-by-line comments:
- Line 1 and throughout: Hyphens should be used in "sea-ice extent", "ice-core records" and other similar compound adjectives, but not when "ice core", "sea ice" etc. appear as nouns.

We thank the referee for this grammatical correction, which we have implemented throughout.

- Line 68. Revise grammar.

We have rewritten this sentence as follows:
"Unlike the previous study, Capron et al. (2021) suggested that any leads and lags between climate elements might be impossible to detect due to both the tight coupling of the different climate elements and the substantial variability between different DO events."

- Line 74-76 seems inconsistent with line 12.

We apologise for this inconsistency. In line 12 we are describing the implications of our results, whereas in lines 74-76 we are discussing the state of affairs which motivated this research. Prior to this study, it was in our view possible that the application of the Erhardt et al. method to new data could still resolve the question of leads and lags between climate elements. As a result of our finding of bias, we feel that this now seems highly unlikely. We have written the last sentence of the paragraph including lines 74-76 in an attempt to reflect this, which we hope satisfies the referee:
"Nonetheless, the previous research in this area has left open the possibility that such an understanding could be achieved in future through the application of this method to either improved ice-core records or data from model simulations."

- Line 84: Models provide a lot more direct insights into the dynamics than ice-core proxies, but "they provide complete information" seems like an overstatement.

We agree with the referee and have removed this statement.

- Line 109: "appears to depends" .. no s or just "depends" … the paper details the event duration's dependence on CO2 quite explicitly.

We thank the referee for this suggestion and have changed "appears to depends" to "depends"

- Line 109: The "chosen range" range does not make sense. The model oscillates with a range of CO2 values (which is not chosen), but the range of CO2 investigated is broader than this,
The referee is entirely correct. We have removed this sentence as we do not feel that it is necessary.

- Line 174: Missing "the" - Caption fig. 1: The last sentence does not apply to a) and d)

We agree with the referee and have clarified in fig. 1 that the improvement is only visually clear for the CCSM4 model data in (e) and (f).

- Line 191: in terms in terms

We thank the referee for pointing out this mistake and have corrected it.

- Line 243: Rather "synthetic data series"?

We agree with the referee and have implemented this suggestion.

- Line 279-280: yes, but especially for the Greenland stadial slope, the range of values used in the tests is much larger than the values observed in data, so the effect may look more dramatic than what is realistic.

Whilst we agree that the majority of events in both the ice core and model data show Greenland stadial slopes that lie within a smaller range than that which we use, there are several individual events in NGRIP with slopes that are outside of our considered range. Table D2 shows that all four proxies in NGRIP have a posterior-mean Greenland stadial slope that is above the maximum of our considered range (1.8 kiloyears^{-1}) for at least one DO event. However, we agree that it is necessary to note that effect may be overdramatised here, and have extended the sentence on lines 281-282 so that it now reads as follows:

“We find that this leads to a too-early bias that can exceed 10 years, although the extremities of the range considered here (and so the largest biases) occur only rarely in the NGRIP data.”

We hope that this caveat satisfies the referee.

- Line 307: Challenging

We thank the referee for this correction.

- Line 424: Suggestion: One important caveat is that the bias we have identified is generally fairly small relative to the timing uncertainty of individual proxies across single DO events, …

We have adopted this suggestion.

- Line 427: Suggestion: may involve

We have adopted this suggestion.

- Line 433: It seems unlikely that 20 synthetic data series with different noise realizations would not show any bias if it indeed was a problem. A more likely explanation is given in the authors’ comments: “We would therefore suggest that the transitions tested by Capron et al. happen to lie in a region of parameter space for which the bias is small, even at high levels of noise.”

We agree with the referee, and we thank them for this comment. We have condensed the paragraphs beginning on lines 430 and 435 into a single paragraph that includes this explanation, as follows:

“Capron et al. (2021) also test for possible bias using synthetic transitions with autocorrelated noise and find no significant bias in either the transition midpoint or duration, implying that the transition onset time must also be unbiased. This stands in contrast to our findings in this study, and so merits further consideration. We suggest that the transitions tested by Capron et al. (2021) lie in a region of the parameter space for which the bias is small, even at high levels of noise. The most obvious reason for this is that the absence of any slopes before or after the ramp favours small biases. We observe in Figure 3 that, when using the original Erhardt et al. (2019) implementation, synthetic transitions without any slopes before or after the ramp show very little bias, even for relatively high levels of noise. However, the bias grows rapidly when even slight slopes are present. We therefore suggest that a key cause of the discrepancy is the assumption made by Capron et al. (2021) that there are no slopes in either the stadial that precedes the transition or the interstadial that

follows. Similarly, Figure 3 shows that for certain “sweet spots” of the transition duration, the bias remains small even as the level of noise increases. It could also be the case that the transition duration chosen by Capron et al. happens to lie in such a sweet spot, and that this is partly why they do not find significant bias.

- Line 473-477: It would be fair to mention around here – or elsewhere in the conclusion - that (significant) slopes are not always present in the data and that the original model outperforms the extended model in the absence of slopes in the data.

We agree that this would be fair. We have therefore changed the sentence starting on line 476 from:
“Furthermore, this bias cannot be alleviated by incorporating said slopes directly into the ramp fitting method.”

to:
“Directly incorporating said slopes into the ramp fitting method reduces the extent of the bias when significant slopes are in fact present in the data, but doing so also worsens the problem when they are not, as is often the case.”