Dear Laurie,

first, we would like to thank you for editing our paper and coordinating this detailed and fruitful review process. We appreciated your last set of comments and implemented the proposed changes. There are just two points we would like to mention:

- l.4: You rightfully noted, that the sea ice retreat can be evidenced for Nordic Seas. Apparently, for the North Atlantic, it is harder to find evidence from marine cores for sea ice extent during stadials and interstadial. Erhardt et al. introduce sodium as a proxy for North Atlantic sea ice:
  - ‘In turn, this allows us to interpret the stadial–interstadial changes in sodium concentrations in the ice cores as qualitative indicators of the extent of the sea-ice cover in the North Atlantic.’ (Erhardt et al., 2019)

  In the manuscript we follow the proxy interpretation as provided by Erhardt. Also, our description of DO events is by far not complete in term of changes that accompany Greenland warming. Thus, we think it is reasonable to not mention the Nordic Seas sea ice dynamics explicitly, even though these are better established than the North Atlantic sea ice dynamics.

- l.55: you proposed to replace ‘According to this proxy variable interpretation’ with ‘According to these proxy records’. Here, we intentionally chose the word ‘interpretation’. With this, we want to express, that the statement on DO events that follows, is conditioned on the interpretation of the proxy variables. If this makes sense to you, we would like to keep the word ‘interpretation’, but replace ‘variable’ with ‘records’.
  - According to this proxy record interpretation, DO events are found to comprise not only sudden warming, but also sudden increase in local precipitation amounts, retreat of the Nordic Seas’ and North Atlantic sea ice cover, and changes of hemispheric circulation patterns.

- l309: You questioned, if the example given to motivate uncertainty propagation was a good one. We agree, that the provided example is in fact not very close to the study and thus we replaced it. While the former example aimed at showing how averaging uncertainties before testing leads to meaningless results, the new example makes a slightly different point but is more related to the study: It highlights the fundamental fact, that large uncertainties in the measurement prevent detailed inference on the population. That is, even in the presence of arbitrarily large uncertainties, the scheme of averaging first and testing second, suggests that one could still take an informed decision about rejecting or accepting the null hypothesis. However, this cannot be true and is thus a fundamental inconsistency of this scheme. We hope that the revised manuscript now conveniently motivates the need for uncertainty propagation by pointing toward this inconsistency that arises when uncertainties are averaged out.

- In the previous version of the manuscript, in the introduction we formulated the null hypothesis ‘that pairwise no transition sequence is physically favoured’ (l.99 in the version edited by you) which corresponds to a population mean equal to zero. Later in the manuscript we slightly modified the null hypothesis and made it one-sided: ‘testing if the sample favours no or the opposite lag’ (l.260 in the edited version).
  - Now, we reconciled this, and changed the abstract and introduction accordingly:
    - ‘with respect to the null hypothesis that the proposed transition order is in fact not systematically favoured’ (l.9)
    - ‘Accordingly, we formulate the null hypothesis that the proposed transition sequence is in fact not physically favoured. In mathematical terms this corresponds to an underlying

population of lags with mean equal to zero or with reversed sign with respect to the observed lags.’ (l.89)

- ‘Here, we formalize the investigation of systematic lead-lag relationships between the proxy transitions. The random experiment framework allows to relate a suspected transition sequence to a mean of the generating population P ΔT which differs from zero in the according direction. Evidence for the suspected sequence can then be achieved by testing the null hypothesis of a population mean equal to zero or opposed to the suspected lag direction. If this null hypothesis can be rejected based on the observations, this would constitute a strong hint for a systematic, physical lag, and would hence potentially yield valuable information on the search for the mechanism(s) and trigger(s) of the DO transitions.’ (l.237)

- We slightly modified the table B1 for sake of clarity and added the table in csv format as a supplement

- Other then that, we adopted all proposed changes.

all the best,
Keno