To whom it may concern,

Re: Response letter to Reviewers, after submission of manuscript "Dansgaard-Oeschger events in climate models: Review and baseline MIS3 protocol" by I. Malmierca-Vallet et al. to Climate of the Past.

Thank you very much for informing me that the discussion period for the above manuscript is now over. Thank you also for the opportunity to resubmit a revised manuscript, according to the reviewer's comments.

I extend my sincere appreciation to the reviewers for their thorough examination of my manuscript, and their detailed and highly constructive comments. I propose to address all of their concerns, both minor and major, so please see attached for a revised manuscript, with the main changes marked in blue.

Here, I address the reviewers' suggestions, comment-by-comment. In the following, the reviewers' comments are in black, and my corresponding response follows in blue.

I very much hope that my responses will satisfy the reviewers and meet your expectations, and therefore request you to consider our revised manuscript for publication in your Journal.

Yours faithfully,

Irene Malmierca-Vallet, and co-authors

AMENDMENT

Before addressing the reviewers' suggestions, the authors would like to propose an amendment to the current manuscript.

Our discussion paper provides a common MIS3 protocol for modelling groups, which is aimed at maximising the chance of D-O type behaviour in the simulations. Subsequent to the submission of this Discussion article, we have found that the orbital forcing (time period) that we suggest for the MIS3 protocol can be improved. Using the model COSMOS, Zhang et al., (2021) already demonstrated that under intermediate glacial conditions, obliquity appears to play a significant role in the occurrence of D-O type behaviour. In particular, the orbital parameters at 40 ka do not produce D-O type behaviour, whilst at 34 ka lower obliquity leads to D-O type behaviour. See Figure 2 from Zhang et al. (below). In the last month, we have also found a similar (unpublished) result using the HadCM3 model. An obliquity-controlled reduction in high latitude annual insolation leads to D-O behaviour in HadCM3, with this model generating D-O type behaviour in a rather narrow window of orbital configurations ranging from 26 ka to 34 ka (Malmierca-Vallet in preparation). Additionally, the MIROC4m model produces D-O-type oscillations (under mid-glacial conditions) and low obliquity (22.9°) (Kuniyoshi et al., 2022). Similarly, a 70 ka (unpublished) simulation performed with the UofT CCSM4 model shows D-O behaviour with low obliquity (approximately 22.4°).

From these COSMOS, and new HadCM3, MIROC4m and CCSM4 results, we deduce that low obliquity (~22.2°-22.6°) seems conducive to D-O behaviour in models. As a result, our MIS3 protocol should be updated to reflect what we have recently learned; we thus propose here that the MIS3 baseline protocol should be modified to focus on conditions at 34 ka.

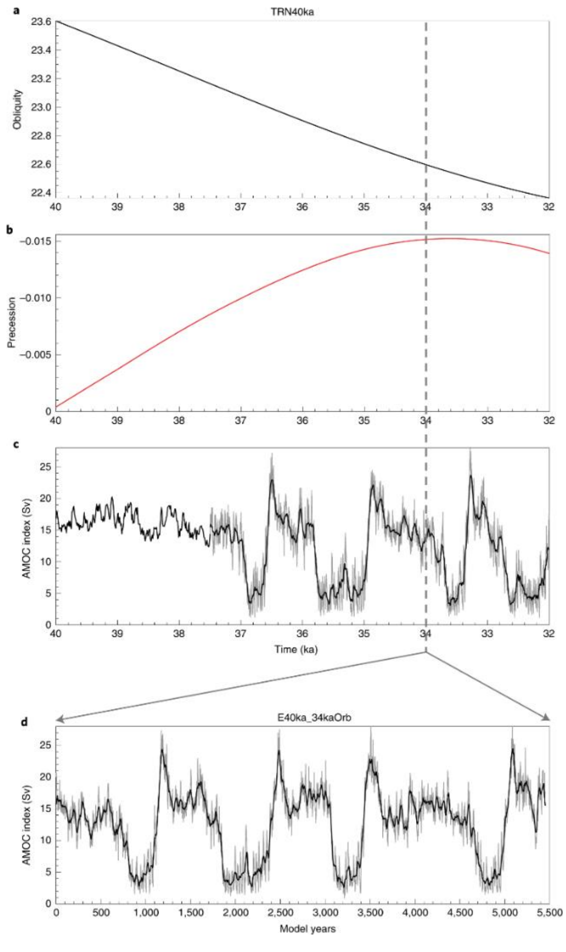

Figure 2 from Zhang et al., 2021. **a**–**c**, Imposed changes in obliquity (**a**) and precession (**b**) and simulated AMOC index (**c**) in experiment TRN40ka. **d**, AMOC index in experiment E40ka_34kaOrb, representing the AMOC response to constant orbital settings of 34 ka under intermediate glacial conditions.

The authors propose the following changes:

Page 1, line 11: We propose a MIS3 baseline protocol at 34 ka, which features low obliquity values, medium-to-low MIS3 greenhouse gas values and the intermediate ice-sheet configuration which our review suggests are most conducive to D-O like behaviour in models.

Page 5, line 107: D-O type quasi-oscillations are also found in MIROC4m under mid-glacial conditions (kuniyoshi et al., 2022). Some aspects of the D-O warming mechanism observed in the UofT-CCSM4, in particular the spatial location of the opening of a big polynya in the Irminger Sea, determining the stadial-interstadial transition, is also identified in MIROC4m (kuniyoshi et al., 2022) (Table A1).

Page 13, line 245: The impact of orbital parameters has been investigated in less detail than the role of GHGs and ice sheets. Using the model COSMOS, Zhang et al. (2021) demonstrated that under intermediate glacial conditions, obliquity appears to play a significant role in the occurrence of D-O type behaviour. In particular, the orbital parameters at 40 ka do not produce D-O type behaviour, whilst at 34 ka lower obliquity (~22.6°) leads to D-O type behaviour (See Figure 2 from Zhang et al. (2021)) Additionally, the MIROC4m model produces D-O-type oscillations (under mid-glacial conditions) and low obliquity (22.9°) (Kuniyoshi et al., 2022). From these COSMOS and MIROC4m results, we deduce that low obliquity seems conducive to D-O behaviour in models.

Page 13, Line 252:These considerations suggest that the interval starting at 38 ka to 32 ka is a good choice for the proposed baseline experiment: it is characterised by (1) a rather regular sequence of D-O events (Fig. 1), and (2) has the ideal intermediate MIS3 ice-sheet configuration conducive to generating D-O-type quasi-oscillations.

Page 15, Line 313: Given its strong evidence basis, we thus suggest the use of the maximum 35 ka Gowan et al. (2021) PaleoMIST ice sheet configuration. We note the LIS is considerably reduced in size, compared to the ICE-6G LGM reconstruction in the southeastern margin (Fig. 7 a,d); the EIS is also significantly smaller (Fig. 7 a,d).

Page 15, line 303: Gowan et al. (2021) provide a maximum and minimal MIS3 reconstruction, specifically for the Laurentide Ice Sheet. The maximum scenario is more consistent with recently discovered eastward oriented, pre-LGM ice flow direction indicators found in southeastern Manitoba (Gauthier et al., 2020), so we currently consider it to be more likely. However, at 35 ka, the difference between the two scenarios is minor. The difference is primarily with the thickness (and therefore also topography) of the ice sheet, rather than extent, but it amounts to less than 1 m of sea level equivalent. The 35 ka time slice represents conditions after Heinrich Event 4 (Andrews et al., 2018), and the ice margin in Hudson Strait is retreated about 350 km from the edge of the continental shelf. The ice margin elsewhere for the Laurentide Ice Sheet is based on chronological constraints, most that are documented in the compilation by (Dalton et al. 2019). The Cordillera Ice Sheet extent is based on evidence of relatively restricted ice cover during MIS 3 (Clague et al., 2011). The Greenland Ice Sheet margin is set to be intermediate of the LGM and present day extent. The Eurasian ice extent at 35 ka includes an advance of ice into the Baltic Sea, which happened after Heinrich Event 4 (Hughes et al, 2016). For East Antarctica, the margin is set to be the same as present. In West Antarctica, the margin at 35 ka is close to the self-edge, as the maximum extent may have been achieved by 30 ka (Larter et al., 2014)

Page 18, line 365: We have shown that reduced ice sheets relative to LGM, low obliquity values…

Page 18, line 367: Around 40% ….

Page 19, line 374: … low obliquity values ….

The following figures and table have been modified:

Figure 6 is now showing insolation anomalies at the top of the atmosphere between 34 ka and PI.

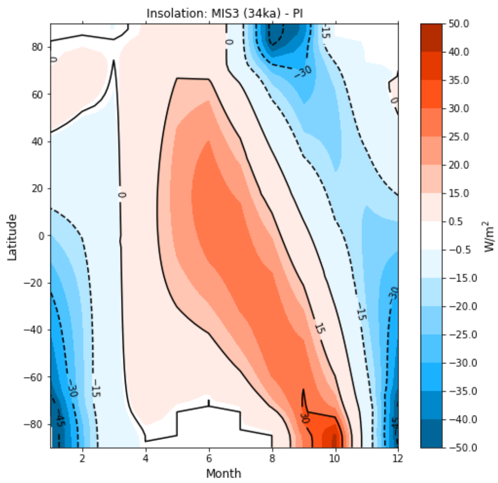

Figure 7 is now showing the 35 ka ice sheet configuration from Gowan et al. (2021).

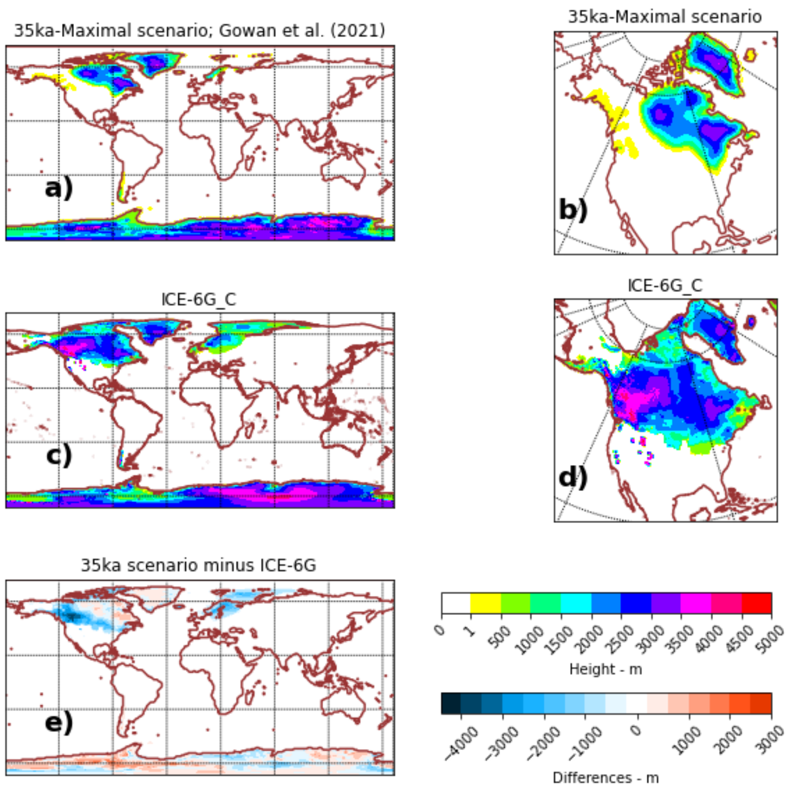

Table 2 is now showing the boundary conditions proposed for the 34 ka baseline experiment.

BC/Forcing	Suggested value MIS3-cnt
Atmospheric trace gases	CO_2: 208 ppm (Bauska et al., 2021) CH_4: 420 ppb (Loulergue et al., 2008) N_2O: 204 ppb (Schilt et al., 2010)
Insolation	Eccentricity: 0.01567 Berger et al. (1998) Obliquity: 22.6° Berger et al. (1998) Precession: -0.016 Berger et al. (1998)
Solar constant	Same as PI control
Ice sheets	35ka ice sheet reconstruction (Gowan et al., 2021); mean global salinity increased by 0.6PSU to account for ice volume
Global freshwater budget	Closed to avoid drifts; Snow should not accumulate over ice sheets and rivers should flow into the ocean. Models need to consider lakes when closing the global freshwater budget
Vegetation	Dynamic or fixed as in PI. If fixed vegetation: tundra in land new points
Dust	As in PI control
H-kicked variant	initial 0.04 - 1 Sv over 250-500 years followed by standard MIS3-cnt simulation

REVIEWER 1

Major comments:

Major comment 1: The variety of model setups and boundary conditions to simulate D-O-type oscillations that is presented in this study (e.g. Table A1) shows that each model has likely a very different threshold at which D-O events might occur. Hence, the models simulate D-O oscillations under very different ice-sheet states, utilized parameterizations and/or trace-gas concentrations. Proposing a common value for the trace-gases and two different ice sheet reconstructions will help to identify whether the models are capable of simulating D-O events under 'realistic' conditions. However, some models might have their threshold at conditions that are close but not within the proposed boundary conditions of the protocol. How to deal with experiments that might formally be outside of the bounds of the experimental protocol but show DO-type oscillations. We likely can still learn something about the mechanisms behind these changes. Hence, I would suggest one additional experiment alongside MIS3-cnt that allows for departures from the protocol (e.g. in terms of atmospheric trace gases – maybe a CO2_variant). Given the mentioned previous studies I doubt that any model will show oscillations at the exact same values that are proposed.

Response to major comment 1: The authors thank the reviewer for this very interesting comment. The authors agree that some models might not reproduce D-O type oscillations under the proposed boundary conditions of the protocol. However, as our amendment above suggests, it could be that the majority of models do oscillate under 34 ka conditions. In order to make clearer the aims/objectives and limitations of the protocol, the following text has been added:

Page 3, Line 68: "Given that D-O events did not occur under full glacial conditions in the last glacial period, the proposed modelling protocol is an important improvement on the use of an LGM PMIP protocol. It will undoubtedly help to shed light on the mechanism and processes involved in millennial-scale oscillations during MIS3. The common MIS3 climate modelling protocol is aimed at: 1) maximising the chance of the occurrence of D-O like events in the simulations; 2) improving model-data evaluation and; 3) providing an adequate central point for modellers to also explore model stability."

Page 14, line 272: We suggest performing a MIS3-cnt experiment centered at 34 ka, using GHG and orbital conditions for 34 ka (Fig. 6); and ice sheet configuration as outlined below (sections 3.1 and 3.2).

We acknowledge that some models might not oscillate under the proposed 34 ka baseline scenario. Indeed, this is expected for NorESM, which under 38 ka conditions, is in a stable regime and the model state seems to be far from a possible tipping point. In spite of that, standardised MIS3 simulations which do not show D-O like behaviour are still highly valuable, for exactly the same reasons that LGM simulations are relevant to the wider modelling community. These standardised MIS3 simulations could contribute to progress on the overarching CMIP6 questions 1 and 2 (Eyring et al, 2016): "How does the Earth System respond to forcing?", and "What are the origins and consequences of systematic model biases?" With a larger number of standardised MIS3 simulations, we would be able to answer questions such as:

- Are state-of-the-art climate models capable of representing D-O events under more realistic MIS3 conditions? Benchmarking these simulations will deliver a measure of how well models simulate abrupt changes, and tipping events.
- If models are too stable to simulate abrupt transitions, what are the processes that contribute to relative levels of model stability?
- Standardized MIS3 simulations can help explore the existence of a theoretical sweet spot for millennial activity in current climate models (Barker and Knorr 2021). As close to or within the sweet spot, the AMOC is characterized by high sensitivity to transient and/or noisy climatic

forcing (e.g. Zhang et al 2014; Lohmann & Ditlevsen 2018) or by self-oscillating behaviors (Zhang et al 2021).

- In addition, a larger number of standardised MIS3 simulations could encourage the creation of new data sets, improving model-data evaluation.

Page 6, line 118: The NorESM - with a reasonably simulated AMOC and Arctic sea ice distribution in the PI and historical simulations (as documented by Guo et al., 2019b) - simulated a MIS3 climate that is in a stable regime with relatively strong convections in the Norwegian and Labrador seas; indeed, NorESM sensitivity experiments including large reductions in atmospheric CO2 levels and Laurentide Ice Sheet heights, aimed at perturbing the system into a cold stadial-like climate, indicate that the model state appears to be far from a possible threshold (Guo et al., 2019).

Page 2, line 47: An important question is if model stability is caused by the model parameters and MIS3 conditions are such that the models are in a mono-stable state, in an oscillatory state or if the models exhibits bi-modality where noise-induced transitions are not induced due to too low model variability (Ditlevsen et al., 2010). Previous studies have questioned the significance of the periodic occurrence of DO events in MIS3 (~ 1470) (Ditlevsen et al., 2007). If the full glacial period is included, the distribution of waiting time between DO-events is consistent with a random process (Ditlevsen et al., 2005). Durations of stadials vs. interstadials indicate correlations with global ice volume and orbital parameters (Lohmann et al., 2018; Mitsui et al., 2017), thus underpinning the decision to focus on MIS3 boundary configurations.

And on the reviewer's suggestion of performing one additional sensitivity experiment, alongside MIS3-cnt that allows for departures from the protocol (e.g. a CO2_variant):

The authors thank the reviewer for this most interesting suggestion. Answering it is outside the scope of this current work. Our suggested approach of performing a single MIS3 baseline experiment, is more or less a direct equivalent to the tier 1 PMIP4-LGM experiment. We agree with the reviewer that additional sensitivity experiments could help disentangle the impact of individual changes in boundary conditions (CO2, ice sheet configuration, diapycnal mixing, orbital parameters) and identify the parameter space where models reproduce D-O type oscillations. We believe that this sub-aim deserves a separate manuscript, particularly the question of how we might more efficiently and effectively explore MIS3 model stability. We are currently working on a second manuscript where we explore in more detail strategies for tuning or exploring model parameter or state-space to find where models may exhibit D-O type oscillatory sweet-spots. This major piece of work is underway, but it takes time mostly because 1) we are analyzing/processing more than seventy sensitivity simulations with HadCM3, MPI-ESM and CCSM4 (run with different CO2 values, orbital configurations, ice sheet morphologies and ocean diffusion coefficients), and 2) we are in the process of adding some COSMOS output to the analysis as well.

In addition, a larger number of standardised MIS3 simulations can help explore the existence of a theoretical sweet spot for millennial activity (Barker and Knorr 2021). We add two sentences to stress on this point in the main text (see above).

Major comment 2: Following the previous point, I am wondering what the authors can learn from the simulations in case the majority of the models (if not all) do not show D-O oscillations under the given parameter ranges. Are there still questions that might be answered. This might add not only scientific credibility to this study but also a motivation for the modelling groups to conduct such experiments.

Response to major comment 2: The authors agree that some models might not reproduce D-O type oscillations under the proposed boundary conditions of the protocol. However, as our amendment above suggests, it could be that the majority of models do oscillate under 34 ka conditions (an intermediate glacial condition). Standardized (34 ka) MIS3 simulations which do not show D-O like behaviour are still highly valuable because 1) they will contribute to progress on the overarching

CMIP6 questions, 2) provide a good starting point for modellers to test the sensitivity of the AMOC to transient or noisy climatic forcing and, 3) they could encourage the creation of new MIS3 data sets for model-data comparison. See also response to major comment 1 from the Referee 1.

Major comment 3: Abstract: The mechanisms behind D-O events varies between models, as presented in Section 2 and Table A1. In the abstract you only state the question: are our climate models too stable to simulate D-O events? I think the model intercomparison will likely be able to answer more than just this question, hence, I would try to include a little more information on how such an intercomparison study may help to improve our process understanding and/or past and future abrupt changes.

Response to major comment 3: In answer, the following text has been added to the abstract to make this point clearer. See also response to major comment 1 from the Referee 1.

Page 1, line 16: This review provides modelling groups investigating MIS3 D-O oscillations with a common framework, which is aimed at 1) maximising the chance of the occurrence of D-O like events in the simulations; 2) allowing more precise model-data evaluation and; 3) providing an adequate central point for modellers to explore model stability.

Major comment 4: Line 60: In fact, so far you have only posed one question (see comment above). It might make sense to shortly summarize the key questions here.

Response to major comment 4: In answer, the authors have added a new list of questions that could be answered through this model intercomparison study. See also response to major comment 1 from the Referee 1.

Major comment 5: Table 1: I liked this short summary of the key terms used within the paper. I was just wondering what you mean by 'A further issue arises…' under the term Heinrich stadial. Could you please elaborate in what sense it is an issue whether the H-event stems from the Laurentide or Fennoscandian ice sheet.

Response to major comment 5: We have elaborated, and rephrased, whilst following INTIMATE naming system. In addition, the part regarding Laurentide vs. Fennoscandian has been removed

Page 4, table 1: Heinrich stadial (HS): this term refers to a stadial containing a specific Heinrich Event. Rasmussen et al. (2014) indicates that the term of HS can refer to the complete stadial period, or to part of a stadial only, characterized by changes shown in proxies of IRD, AMOC or SST (Barker et al., 2009).

Major comment 6: Section 2.2: A sensitivity to the ice sheet height is evident in all model simulations presented in the present study. In the protocol two different ice-sheet reconstructions are proposed as boundary condition and Fig. 7 shows that the reconstructions are quite different, specifically in terms of the extent and height of the Laurentide ice sheet. I assume that the modelers can choose between either one of the reconstructions. It would be good to add a small discussion on the uncertainties and differences that are expected to arise in the model ensemble in terms of these uncertainties in the ice sheet reconstructions. Previous studies for glacial-interglacial conditions have shown, that a different height and extent of the Laurentide ice sheet has a significant effect on the AMOC (e.g. Löfverström et al., 2014 - https://doi.org/10.5194/cp-10-1453-2014 or Kapsch et al., 2022 - https://doi.org/10.1029/2021GL096767).

Response to major comment 6: Subsequent to the submission of this discussion article, we have found that the orbital forcing (time period) that we suggest for the MIS3 protocol can be improved. From the COSMOS, and new HadCM3, MIROC4m and CCSM4 results, we deduce that low obliquity (~22.2°-22.6°) seems conducive to D-O behaviour in models. As a result, we propose to update our MIS3 protocol to reflect what we have recently learned. We suggest performing a MIS3-cnt experiment centered at 34 ka, using GHG and orbital conditions for 34 ka; and the 35 ka ice sheet configuration published in Gowan et al., (2021). Although, we no longer proposed different ice-sheet reconstructions for the baseline experiment, we agree with the reviewer that consideration of uncertainties in boundary conditions is particularly important when comparing the model results to paleoclimatic reconstructions and drawing conclusions about the capabilities of the state-of-the-art models that are used for future climate projections. In answer, the following text has been modified in section 3.2.

Page 15, Line 316: The single ice sheet reconstruction MIS3 set-up summarized above contrasts with the PMIP4 LGM protocol, which provides three different possible ice sheet configurations (PMIP3, ICE6G-C and GLAC-1D) for the tier 1 LGM experiment. This partially reflects the more limited knowledge of ice sheet in pre-LGM periods. Exploration of the effect of MIS3 ice sheet reconstructions uncertainties on climate models, particularly on model stability, would be valuable. For this purpose, further additional 34 ky / MIS3 ice sheet reconstructions would be very valuable.

Technical corrections:

Technical correction 1: Line 75: 'which do not show these' – 'n' removed, also might be good to refer to 'oscillations' instead of 'these'. Done

Technical correction 2: Line 77: 'A number …' – this sentence is a little confusing and becomes only clear after reading Table A1. Please revise. Done

Line 77: A number of PI/present-day model simulations exhibit spontaneous centennial-length cold events (Table. A1), however, they do not appear to be D-O like events. We deal with these first.

Technical correction 3: Line 161: 'the sub polar gyre contracts, an inflow of'. Done

Technical correction 4: Table 1: Bond cycle: 'following H-events H5 and H6' – hyphen missing. Done

Technical correction 5: Table A1-A3: There are several small typos throughout the three tables, please revise. E.g. Drijfhout et al.: 'spontaneous cold event that last around 100 years' – missing s, but there are several more. Also the format is not optimal in some places, specifically when there are several simulations per study. Done

REVIEWER 2

Major comments:

Major comment 1: Unclear aim or focus of proposed experiments. The abstract states: "The climate modelling community up to now has not been able to answer the question: Are our climate models too stable to simulate D-O events?" and this appears to be the central question driving this study. Surely we are interested in much more than this binary question? Some might argue we already know the answer – yes, some models are too stable – the important questions are WHY this is, what are the processes that contribute to relative levels of model stability? I suggest that the scientific motivation behind this proposed model intercomparison protocol is explicitly laid-out and justified by the data available. For example at Line 60 "The design of a common MIS3 experimental protocol would allow the modelling community to address the questions posed above" – I have a hard time working out what the questions above are. I note that some questions are introduced later at Line 78.

Response to major comment 1: The authors thank the reviewer for this most interesting comment. In order to make clearer the scientific motivation of the protocol, the authors have added a new list of questions that could be answered through this model intercomparison study. The proposed changes are given in the response to comment 1 from Reviewer 1.

Major comment 2: Conflation of model and data. Related to the above, the logic of this manuscript is quite difficult to follow because from the abstract onwards knowledge of DO events from primary data (e.g., ice cores, marine sediments) is conflated with model simulations. It is not clear if the project's motivation is grounded in the data or in model nuances (that may of course be relevant). For example, Line 104-5 states that D-O type behavior shouldn't be seen under full glacial conditions because it isn't present in the data – surely that information should be upfront, included as a primary pillar of your protocol design, not mentioned because some models might agree with it? In Section 2.1, please make clear with D-O type oscillations are in models…in which parameters? What timescale and amplitude satisfies your definition? [n.b., Table 1 states "It is not clear that series of D-O events are oscillations in the strict sense."]

Response to major comment 2: Out of the six models that reproduce D-O type oscillations (HadCM3, COSMOS, CM2Mc, MIROC4m, CCSM4 and MPI-ESM), CCSM4 is the only one that oscillates under full LGM conditions. Ideally, models should not oscillate under full glacial conditions because DO events did not occur during the LGM. Following the reviewer's recommendation, we have added the following text, making clear that one of the primary pillars of the protocol is to find out if models are able to oscillate under more "realistic" MIS3 conditions.

Page 3, Line 68: "Given that in reality D-O events do not occur under full glacial conditions, the proposed modelling protocol is an important improvement on the use of an LGM PMIP protocol. It will undoubtedly help to shed light on the mechanism and processes involved in millennial-scale oscillations during MIS3. The common MIS3 climate modelling protocol is aimed at: 1) maximising the chance of the occurrence of D-O like events in the simulations; 2) improving model-data evaluation and; 3) providing an adequate central point for modellers to also explore model stability."

And on the reviewer's suggestion of making clear the D-O type oscillation term used in the manuscript, we have added a new definition in table 1. We also indicate more precisely (in table A1 and B1) the models that show D-O type oscillations.

Table 1 : D-O type oscillations: For the purpose of this MIS3 DO protocol, the term of D-O type oscillation refers to D-O scale climate variability reproduced by climate models, comparable to the D-O events observed in the Greenland ice core record.

Major comment 3: Nomenclature.

(1) Table 1 is not a useful addition to the literature as it stands. It is littered with errors (too many for me to highlight every one) and introduces further ambiguity and confusion. The ambition to reset nomenclature usage across the entire palaeoclimate community is a large one. It would better to state that you wish to define nomenclature for use within this proposed PMIP protocol and associated papers only.

(2) In table 1: "stadial-interstadial" – these occur in many other time periods in addition to MIS3. Might be worth referencing the INTIMATE naming system here.

(3) "Heinrich events" – entirely inaccurate definition. There is little, if any, evidence to support your claim that Heinrich events "have a role in DO oscillations". There are small methane peaks within some (but not all) Heinrich stadials that have been linked to Heinrich events.

(4) They typically occur mid-stadial not "before a stadial has begun". "Heinrich Stadial" This term denotes a Greenland Stadial in which a Heinrich event is thought to occur. A stadial can therefore be both a HS and a GS. Sentence about H event provenance is out of place here!

(5) "Bond cycle" is the term/concept useful or relevant here? Heinrich "

Response to major comment 3: We now state that we wish to define nomenclature for use within the proposed MIS3 DO protocol. We have redefined all terms in Table 1 following the INTIMATE naming system. In addition, we explain more clearly why the term Bond cycle is relevant for the protocol - since D-O event Bond cycles, which begin with a Heinrich Event, are essentially the whole rational for the Heinrich event preconditioned proposed simulation variant.

Page3, Line 63: Firstly, Table 1 and Figure 1 provide a framework for a more consistent terminology for use within this proposed MIS3 DO protocol.

The definitions included in Table 1:

Abrupt change: We follow the IPCC Assessment Report 4 (IPCC AR4) definition of abrupt event/change (Meehl et al., 2007; Brokin et al., 2021). This term refers to a large-scale change, which is much faster than the change in the pertinent forcing (e.g. rising atmospheric CO2 concentrations).

Tipping point: This term refers to a critical threshold at which a small perturbation can qualitatively modify the development or state of a system (Lenton et al., 2018).

Tipping element: This term describes large-scale components of the Earth system that could pass a tipping point (Lenton et al., 2018). Earth system components are the ocean, atmosphere, cryosphere, anthroposphere and biosphere, which have further important sub-components e.g. the meridional ocean circulation, the monsoon systems, sea ice, and various ecosystems (Brokin et al., 2021).

D-O event: During the Last Glacial period, a series of dramatic climatic fluctuations occurred in the North Atlantic known as D–O events, during which atmospheric and oceanic conditions alternated between relatively mild (interstadial) and full glacial (stadial) conditions (Dansgaard et al., 1982, Johnsen et al., 1992). Around 25 abrupt transitions (each completed within a decade) from stadial to interstadial conditions occurred during the Last Glacial period and their amplitude vary from 5 to 16°C (Rasmussen et al., 2014). The duration of interstadials varies from approximately a century to many millennia (Rasmussen et al., 2014).

D-O type oscillations: For the purpose of this MIS3 DO protocol, the term of D-O type oscillation refers to D-O scale climate variability reproduced by climate models, comparable to the D-O events observed in the Greenland ice core record.

Greenland stadial and interstadial: We follow the INTIMATE (Integration of Ice core, Marine and Terrestrial records of the North Atlantic) definition of stadial/interstadial terms (Rasmussen et al., 2014). The Greenland Interstadials (GI) and Greenland Stadials (GS) periods terms are the Greenland expressions of the D-O events and represent warm and cold phases of the NA area, respectively.

Heinrich events: these are defined by the presence of layers of ice-rafted debris (IRD) of primarily (not exclusively) Laurentide origin in North Atlantic sediment cores (e.g., Heinrich, 1988, Hemming, 2004, and references therein). Heinrich events have been observed during some of the longer stadials, but likely do not cover the entire period of these longer stadials (Roche et al., 2004, Marcott et al., 2011).

Heinrich stadial (HS): this term refers to a stadial containing a specific Heinrich Event. Rasmussen et al. (2014) indicates that the term of HS can refer to the complete stadial period, or to part of a stadial only, characterized by changes shown in proxies of IRD, AMOC or SST (Barker et al., 2009).

Bond cycle: D-O events tend to follow a pattern of diminishing amplitude (or a general cooling trend of the GSs) following each HE (Bond et al., 1992; Alley et al., 1998,1999; Clark et al., 2007; Rousseau et al., 2022). These cycles of HE grouped D-O events were named Bond cycles by Broecker et al. (1994) and Alley et al. (1998). The average gap between HEs is around 7 ka, so this is the average length of a Bond Cycle (Clark et al., 2007).

Major comment 4: Heinrich “kick” Line 68: “In addition to the protocol for a baseline simulation, we also outline a protocol for a Heinrich event (Bond cycle event one type; Table 1) preconditioned variant.” This statement makes no sense to me and Table 1 does not help. Scanning down to section 3.3…the term “kicked Heinrich event” is still not explained. This section needs dedicated attention from co-authors. Stocker and Johnsen 2003 do not discuss “The freshwater delivered during Heinrich event iceberg discharge extends GS duration and suppresses the AMOC”. Why is it “logical to presume that these freshwater events are important in preconditioning the climate system with respect to D-O behaviour”. There seems to be an equally strong argument which states DO events would continue to occur without Heinrich events.

Response to major comment 4: We agree with the reviewer that the term “kicked Heinrich meltwater” variant is confusing and we now use the terms unforced (D-O type) and Heinrich-Event preconditioned (H-E preconditioned D-O type) abrupt climate change. The following text has been modified/added:

Page 1, Line 14: We also provide a protocol for a second freshwater (Heinrich-Event preconditioned) experiment, since previous work suggests that this variant may be helpful in preconditioning a state in models which is conducive to D-O events.

Section 3.3: This section has a new title now: “Heinrich-Event preconditioned option”. In addition, the following text has been added/modified in section 3.3.

Page 17, line 331: The term “kick” Heinrich event-like was initially introduced by Peltier and Vettoretti. (2014) to invoke a ’kicked’ salt oscillator hypothesis (pseudo-Heinrich type behaviour), to induce D-O type oscillations in an LGM simulation performed with the UofT CCSM4. During the first thousand years of the LGM simulation, the UofT CCSM4 is spun up and the ocean cools to reach a state consistent with glacial boundary conditions, there are two thermal thresholds during which the strength of the AMOC rapidly reduces (see Figure 2 in Peltier et al., 2020). These abrupt transitions in the AMOC coincide with abrupt reductions in surface temperatures in the North Atlantic and abrupt expansions of sea ice coverage. During the second of these events, the AMOC is reduced to approximately 12 Sv, about half its strength in the pre-industrial control (Peltier et al., 2020). This event may resemble the impact of a Heinrich event-like "kick" to the AMOC though no freshwater perturbation was imposed (Peltier et al., 2020).

In a more recent study, Pedro et al. (2022) examine the CCSM4 simulations that shows unforced D-O type oscillations (Peltier and Vettoretti, 2014; Vettoretti and Peltier., 2018), but with the addition of a (freshwater) H-like event. The freshwater flux (0.05 Sv injected into the NA for 500 years) leads to (1) a 5 Sv weaker AMOC compared to the one seen in the unforced model stadial and, (2) a stronger D-O warming transition into the interstadial phase (Pedro et al., 2022). Thus this type of H-E preconditioning can trigger abrupt reductions in the AMOC strength and in NA surface temperatures and sea ice coverage – and it may also help induce a stadial state in other models which is more conducive to unforced (D-O type) oscillations.

Page 17, line 345: Given the importance of HEs to starting Bond Cycles of D-O events, an additional experiment to investigate how HE meltwater preconditioning impacts the simulation of D-O like oscillations under MIS3 boundary conditions would be valuable. HE freshwater preconditioning may, as in reality, be more conducive to a (Bon Cycle-like) sequence of spontaneous D-O type oscillations (see Table 1).

Page 17, line 360: After 250-500 years this freshwater forcing should be switched off.

Major comment 5: Could the authors clarify what they are trying to achieve or test with this "kicked Hevent"? How will they distinguish freshwater flux related to a Heinrich event (iceberg discharge from Laurentide) with other ice sheet instability delivering IRD and [potentially] contributing to AMOC variations that occurred within every DO cycle, H event or not.

Response to major comment 5: We thank the reviewer for this most interesting comment. We suggest that some models may require a H-E preconditioned (freshwater) as a precursor to the unforced (D-O type) oscillatory behavior. Indeed, Peltier and Vettoretti (2014), Peltier et al. (2020) and Pedro et al (2022) have suggested that some models require a freshwater Heinrich-like event as a precursor to the D-O type quasi-oscillatory behaviour. See also response to major comment 4 from reviewer 2.

In addition, we agree with the reviewer that it would be most interesting to distinguish freshwater flux related to H event (iceberg discharge from Laurentide) with other ice sheet instability delivering IRD. Our suggested approach of performing a single freshwater (H-like preconditioned) event experiment, is to keep the MIS3 DO protocol as simple and accessible as possible to the different modelling groups. The single suggested freshwater (H-like preconditioned) experiment is aimed to represent a Greenland stadial that contains a HE.

Major comment 6: Sections 2.1 and 2.2 contain interesting discussion of the likely processes and feedback involved in generating D-O like events, and the associated figures are well-presented. Throughout though, it needs to be clear how the new protocol will address key questions related to the role of sea ice etc. that the discussion identifies.

Response to major comment 7: The following sentences have been added to make this point clearer.

Page 19, Line 331: More model simulations run under the here proposed MIS3 DO protocol together with analyses across models, could provide better insights, along the lines of atmospheric-ice-ocean feedbacks behind DO events. These simulations will allow us to answer questions such as: are current climate models able to reproduce DO-type behaviour under more realistic MIS3 conditions? How well models simulate tipping events, abrupt changes? What are the mechanisms that lead to relative levels of model stability? Moreover, a large number of standardised MIS3 simulations could encourage the creation of new data sets, improving model-data evaluation.

Major comment 7: Atmospheric gases. Could you comment further on choice to keep atmospheric CO_2 constant? (section 3.1). What does that assume about the role of CO_2 in D-O behavior?

Response to major comment 7: The authors thank the reviewer for this most interesting question. Answering it is outside the scope of this current work. We agree with the reviewer that additional sensitivity experiments could help disentangle the impact of individual changes in boundary conditions (e.g. CO2) and identify the parameter space where models reproduce D-O type oscillations. We believe that this sub-aim deserves a separate manuscript, particularly the question of how we might more efficiently and effectively explore MIS3 model stability. We are currently working on a second manuscript where we explore in more detail strategies for tuning or exploring model parameter or state-space to find where models may exhibit D-O type oscillatory sweet-spot. This major piece of work is underway, but it takes time mostly because 1) we are analyzing/processing more than seventy HadCM3, MPI-ESM and CCSM4 sensitivity simulations (run with different CO2 values, orbital configurations, ice sheet morphologies and ocean diffusion coefficients), and 2) we are in the process of adding some COSMOS output to the analysis as well. See also response to major 1 comment from reviewer 1.

Technical corrections

Minor comment 1: Line 62: "Given…" This is a phrase not a sentence. Done