

This document includes responses (in blue) to all the Reviewer comments.

Response to Interactive comments by Anonymous Referee #1 [RC1]

However, it seems to me that in the current setup there are perhaps too many sensitivity experiments proposed, with too many options, with the likely result that in the end a proper model inter-comparison becomes difficult since very few groups in fact performed the same sensitivity experiments. The authors have certainly carefully considered the issue and thus the proposed simulations are possibly the optimal solution, however, it seems to me that the topic is of such importance that a general reviewer comment is justified.

We agree. With so many sensitivity expts there is the possibility of too few modeling groups doing the same experiments. We have limited the lig116k glacial inception to the reconstructed value of CO2 and kept only the prescribed boreal forest and shrub savanna experiments to test the sensitivities to more idealized vegetation configurations. We also now suggest that the hol8.2k experiment can start from the hol9.5k simulation.

Therefore I strongly suggest to discuss the topic: What are the differences between early Holocene / Last Interglacial and future climate change? Which lessons from these past periods can be used to inform us about the future? Which systems and their sensitivity are influenced by the difference in the forcing and which ones are for instance solely driven by higher temperatures and can thus be seen as direct analogues? Do the authors have data available to provide some examples? Or have studies been done to investigate this? Possible examples that come to mind are Masson-Delmotte et al. (2006, DOI: 0.1007/s00382-005-0081-9c) and Blaschek et al. (2015, DOI: 10.1007/s00382-014-2279-1).

We have added a short discussion to the Conclusions. Neither time period can be suggested as a true paleo-analogue for the future because of the seasonal nature of the orbital forcing. That said, higher temperatures in the polar regions, particularly during the summer months and for the Last Interglacial, directly influence sea ice and the ice sheets. The data evidence provides a means of evaluating if we are capturing this sensitivity correctly in models being used for projections of future climate change.

The manuscript describes quite extensively the proxy-based paleoclimate datasets that are available for the early Holocene and the Last Interglacial (section 4). However, it does not seem to come to any conclusions. Is this part solely meant to provide an overview? If so, is this the correct journal and manuscript to do so? Or do the authors mean to provide some guidance for future model-data inter-comparisons? In which case the conclusions of this section should be made more clear.

The goal here was to summarise the data sets that are available for different types of model evaluation for the MH and LIG, and thus we explicitly draw attention to the state-of-the-art syntheses of data on e.g. hydrology, vegetation, climate reconstruction, climate variability etc in this section. We feel that it is important to do this in the current paper (a) to make it clear that model evaluation of the experiments is feasible, and (b) so that modelers are aware of the available data sets and their limitations. However, we agree that it would be useful to add something at the end of this section in conclusion. So we have added the following:

“The public-access data sets currently available for the MH and LIG serve different functions and address different aspects of the climate system. Modeling groups running MH and LIG simulations, or sensitivity experiments, are encouraged to work with data experts, using multiple data sets for a full diagnosis of the simulations. Many of these data sets provide measures of the uncertainty of the reconstructions and data-model comparisons should be designed to take these uncertainties into account.”

Minor and technical comments:

Lines 87-88: The second part of this sentence describes the possible results of these forcings, but since this paragraph solely describes the forcings themselves, it could be better to move this part to later in the manuscript.

Moved to next paragraph.

Lines 87-88: Provide references for the warmer NH summers and enhanced NH monsoons.

References have been added.

Lines 106-107: Different model biases and the issue of model complexity are discussed here. How do the presented experimental designs overcome these issues? Please shortly discuss, perhaps later in the manuscript.

Deleted sentence.

Lines 181-183: It would be easier for the reader if the order in which the two periods are described is always the same, either first early Holocene and then Last Interglacial or the other way around.

Order switched here to be consistent with discussion of other forcings

Line 197: What is meant with 'surface', really surface or something like 2m- temperature, reference temperature?

Line 198: Are these surface winds?

We have moved this paragraph and discussion to the Section 6 on Data Availability and made consistent with the text in the revised version of the CMIP6-PMIP4 Overview paper in GMDD

Line 284: Perhaps 130-126 ka for consistency.

Revised as suggested.

Lines 344-348: Consider mentioning again that daily output is needed to calculate output on angular calendar months vs a fixed calendar months. Is the second part of this paragraph clear to the reader? Do they know what 'output needed to force regional area-limited models' is?

We have made consistent with the text in the revised version of the CMIP6-PMIP4 Overview paper in GMDD. This text now appears in Section 6 on Data Availability of our paper. A supplementary information table has been added to the paper and also a link to the PMIP4 web site.

Lines 344-363: Are these paragraphs needed or can the manuscript simply refer to the website where all this information can be found?

We have not shortened this section on the advice of GMD Chief Editor J. Hargreaves. We have added a list of variables to be saved in the SI.

Line 383: What feedbacks is this sentence referring to?

Albedo and freshwater

Line 404: 'can' or 'should' be prescribed from the last deglaciation experiment? It seems to me that these are the details that could in the end result in a model inter-comparison in which all simulations are slightly different from one another.

Should

Lines 409-412: Is the CO₂ concentration for 116ka so uncertain?

The Tier 2 experiment will use the EPICA Dome C data published by Schneider et al., 2013 CP (that is also publishing $\delta^{13}\text{-CO}_2$ for the same depth intervals) which is the currently best CO₂ data set for the Eemian. For the 116 ka BP, the nearest data point at 115909 ka BP (on AICC12) has 273 ppm in CO₂.

Lines 426-440: Make it more clear which of these sensitivity experiments are proposed for both the Holocene and the Last Interglacial and which ones only for the Holocene (in line with Table 2).

Revised as suggested.

Line 445: remove first 'is'.

Paragraph deleted.

Line 445: year of reference to Hoelzmann is missing.

Paragraph deleted.

Lines 446-449: This part is rather vague. Are sensitivity experiments in this direction foreseen in PMIP4 or not?

Paragraph deleted.

Line 470: This line seems to suggest that the 'coral records' do not provide evidence of sea level rise. Please rephrase.

Clarified

Line 471: Can this value also be given in Sv for easier comparison?

Done

Line 484: 0.13Sv doesn't seem small at all, is this typo?

Revised to be clearer

Line 489: I understand that the 8.2ka-event happened close to 8.5ka, but is it really necessary to introduce yet another simulation? Can't one simply use either 9.5 or 6ka in line with the other experiments?

We agree. This proposed experiment now starts from a 9.5 ka simulation.

Lines 489-492: This is somewhat unclear to me. Should the 2.5Sv pulse be introduced in year 500? Should both fluxes cease after this or should the background flux continue?

We agree that as first written the experimental protocol was unclear. Revised to be clearer.

Lines 492-493: Does 'evidence for the recovery' mean that some upward trend should be visible or that it should again be close to the initial state? Why not give a more concrete number like a minimum of 100yrs after the end of the pulse?

We agree. Sentence revised.

Lines 479-493: It seems that this experimental design is similar to experiments that have been performed previously (Wagner et al., 2013, DOI: 10.1007/s00382-013-1706-z and Morrill et al., 2013, DOI: 10.5194/cp-9-955-2013), please refer to these manuscripts and discuss how and why those simulations differed from the design that is proposed here.

We have revised the experimental design for the freshwater inputs to be the 'Lake + Ice_100 yrs' scenario of Wagner et al. (2013). As they suggest this design is more consistent with ice dynamics and the data of Carlson et al. (2009) than the shorter 1-yr flood scenarios described in Morrill et al., 2013.

Lines 509-512: Why not start from 127ka? Should groups perform a 128ka equilibrium simulation as initial condition for the transient 128-122ka simulation? According to table 2 one should use 127ka as initial condition, but will this not lead to some spurious jump in the climate?

The orbital and GHG forcings are similar for 128 ka and 127 ka but agree that a more consistent experimental protocol is to start from 127 ka. Ending date also revised to 121 ka to be of similar length as midHolocene transient simulation.

Line 540: Should this be 'small' rather than 'large'?

Yes, small is correct. We corrected.

Line 572: Bakker et al. 2013 does not include LIG proxy-based climate reconstruction data. Perhaps Bakker et al., 2014 (DOI: 10.1016/j.quascirev.2014.06.031) is meant?

We originally included this because it uses climate reconstructions to evaluate the simulations, including simulation of sea ice and high-latitude temperatures. We have now taken the reference out. Bakker et al (2014) is also a data-model comparison paper, rather than a primary source for reconstructions.

Lines 622-623: A number of datasets that are mentioned throughout the manuscript are not available on the website, when will they be?

Datasets are now available on the website

Table 1: For 'Other GHG gases' 6ka and 127ka say '0', is that different from 'CMIP DECK piControl'?

Yes, there are no CFCs included in the midHolocene and lig127k simulations.

What does 'SSI,ap if needed' mean?

Deleted. Refers to the global level of geomagnetic activity, and are used as inputs to parameterisations of magnetospheric particle precipitation.

Table 2: part of experiment 3.1 could also be considered as part of 3.3 (sensitivity to ice sheets).

Yes, but we have decided to keep separate.

The hol8.2 ka event simulation is somewhat confusing, should it be 8.2 or 8.5 ka orbital?

We now suggest that the hol8.2k experiment should start from the hol9.5k simulation. Except for the freshwater forcing, all other forcings and boundary conditions remain the same as the hol9.5k simulation.

Why is a freshwater forcing coming from the Antarctic Ice Sheet not taken into account?

The early last interglacial freshwater forcing associated with the H11 event is likely to have come from the melting of the MIS6 NH ice sheets. Reconstructions from marine data support this.

Figure 1: Is the horizontal placement of the global sea-level peak in panel k suggesting the timing of the Last Interglacial high-stand?

No. We will clarify this in the legend.

Figures 3 and 4: Color bars are missing and quality is rather low.

Redrafted to improve quality. Color bars added.

Response to Interactive comments by Julie Brigham-Grette [RC2]

My major complaint is that the modeling design does not really get at sea ice. Its mentioned a few times. It would seem absolutely necessary that different sea ice configurations are included the same way that different (or prescribed) ice sheet geographies are included. For 6k (with nearly ice free summers from 9 to 6ka? (Funder et al) and for 127k (possible ice free summers at peak interglacial?; no sea ice south of Bering Strait, several papers) sea ice variability or 2-3 different modeled geographies might be considered.

All the sensitivity simulations proposed are to use the same coupled atmosphere-ocean-sea ice models as the midHolocene and lig127k CMIP6 or PMIP4 simulations to assess missing forcings/boundary conditions that affect the coupled climate state. Although AMIP simulations with different scenarios of seasonal sea ice extent could be of interest, we do not include here.

The paper is well written and easy to follow however I had to read parts of the Eyring et al paper 2016 in this journal to find some of the terminology. I am not a modeler, yet I am among those in the community who would like to read about modeling project plans, but might not immediately

understand what the “CMIP DECK” is. The paper does an impressive job listing summaries and paleodata compilations that might be used for input however it is not exhaustive.

This paper should move forward to publication with only a few picky edits:

Line 110: I suggest for non-modelers that you add a footnote about what an “entry card” is? I understand this refers to a specific list of requirements.

As indicated in subsequent text: “all modeling groups contributing to PMIP4-CMIP6 must perform either the midHolocene experiment or a simulation of the Last Glacial Maximum (Kageyama et al., 2016).

Line 129: define ISMIP6 – Ice Sheet Model Intercomparison 6 contribution to CMIP.

Defined.

Line 153: typo, Mollusc shells, not mullusc shells.

Corrected

Line 164: write out the meaning of DECK – Diagnostic, Evaluation and Characterization of Klima. One should not have to read the Eyring et al. 2016 paper to get all of the acronyms

We have revised as suggested.

.Line 429: move the Lozhkin and Anderson reference. So it reads: . . . vegetation and climate in which vegetation cover in the high-latitudes is changed from tundra to boreal forest (experiment a) (Lozhkin and Anderson, 1995) and the Sahara desert is replaced by evergreen shrub to 25°N and savanna/. . .

Paragraph deleted.

Line 445: Add year to the Hoelzmann et al reference.

Paragraph deleted.

Line 465 and section 3.4: Should/Could H11 be added to figure 1? Line 490: The location of the freshwater flux is extremely important and there might be reasons for the freshwater to hug the coast rather than be flooded over the entire Labrador sea. So this might also be part of the experimental design?

We have removed all gray lines from Figure 1. We propose starting the H11 experiment from the 127ka simulation. True, this is an idealization but more feasible than adding an additional time slice simulation for some time between 132-130ka. This should give the first-order approximation of the climate state at 130ka. Again to make this simulation feasible for many modeling groups, we have suggested for this Tier 2 simulation to just add the freshwater across the North Atlantic. Although the European ice sheet is reconstructed to be larger in MIS6 than MIS2, the partitioning of the H11 freshwater between the European and Laurentide ice sheets is not clear.

Line 1158: Fig. 1 caption. Add Veres et al, 2013 to the sentence containing AICC2012. Or perhaps better yet, cite the editors of the AICC2012 volume.

Veres et al, 2013 moved as suggested.

Lines 1175 and 1180: Add color bars to figure 3 and 4 because the print on the lines in the figure are very tiny.

Done.

Response to Interactive comments by Anonymous Referee #3 [RC3]

I tend to disagree with Referee 1 that the paper should include some speculations on lessons learnt from the interglacial simulations for possible future climate change. The story of future climate change is about the response of our climate system to unprecedented strong variation in external forcing, while the mid-Holocene and the early Eemian pose the challenge of explaining climate change in the presence of weak variation in external forcing. In any case, simulating the subtleties of past climate variability is a prerequisite for gaining confidence in understanding the dynamics of our climate system – which the authors clearly state.

Before publication, I would appreciate, if the authors could consider the following issues.

a) The authors correctly highlight uncertainties arising from prescribing or simulating Holocene and Eemian vegetation patterns. The authors recommend using the reconstruction by Hoelzmann et al. (1998) for Holocene North Africa. Is this still the best reconstruction? What about the reconstructions mentioned in the papers cited by the authors or by Lézine et al. (2011), Larrasoana et al. (2013), : : :? Perhaps there are good reasons to still use Hoelzmann's et al data. But this should be critically reassessed.

We agree with Reviewer #1 that there are perhaps too many sensitivity experiments proposed, with too many options. With so many sensitivity expts there is the possibility of too few modeling groups doing the same experiments. We have kept only the prescribed boreal forest and shrub savanna experiments to test the sensitivities to more idealized vegetation, consistent with reconstructions, among models and for comparing these two time periods.

b) In the same line: What about lakes? Lakes potentially matter in the mid-Holocene Sahara (e.g., Krinner et al., GRL, 2012). Perhaps also for lakes, the reconstruction by Hoelzmann et al. would be useful.

We agree that lakes could potentially impact the monsoon climate, and there were a number of papers showing this prior to the Krinner et al (2012) paper. Indeed, the Hoelzmann et al (1998) data set was explicitly constructed to allow such sensitivity experiments to be made, see Broström, A., Coe, M., Harrison, S.P., Gallimore, R., Kutzbach, J.E., Foley, J., Prentice, I.C. and Behling, P., 1998. Land surface feedbacks and palaeomonsoons in northern Africa. Geophysical Research Letters 25: 3615-3618. However, previous work suggests that the vegetation changes appear to be the most important element of the land-surface feedback – and given limited resources, our focus in PMIP and in this paper is therefore on designing common experiments to examine the sensitivity to vegetation.

c) SST biases in the coupled atmosphere – ocean models presumably contribute to an

underestimate of Interglacial Monsoon strengths. Hence some SST sensitivity experiments (e.g., strong vs weak gradient between tropical and extratropical SST differences between pre-industrial and mid-Holocene / LIG climate) might be worth considering.

All the sensitivity simulations proposed are to use the same coupled atmosphere-ocean-sea ice models as the midHolocene and lig127k CMIP6 or PMIP4 simulations to assess missing forcings/boundary conditions that affect the coupled climate state. Although AMIP simulations with different gradients of SST or scenarios of seasonal sea ice extent could be of interest, we do not include here.

d) Likewise, sensitivity experiments with respect to changes in Arctic sea-ice could be instructive to explore the role of high latitude climate system feedbacks (cf. the papers cited by the authors). *See comment above.*

Minor issues:

i) The term Tier 1 (explained in line 132) should be defined earlier.

Defined.

ii) Fig. 6: I do not understand what Figure 6a refers to. Is it just the geographic location of dust sources, or the difference in locations – irrespective of their strength? Please use a superscript in the dimension $\text{g/m}^2/\text{a}$.

Revised.

iii) Fig. 5: What is the meaning of the different shades of grey on the continents? In the upper figure, the color grey also appears on the temperature scale (temperature difference between 1.5 and 2 K)?!

We agree that the color scales are confusing in this figure. This figure has been redrafted.

iv) Line 754 ff: Something went wrong with the citation of the Dahl Jensen and : : : and : : : and .

Corrected.

References to papers cited (and not included in the manuscript):

Krinner, G., Lézine, A. M., Braconnot, P., Sepulchre, P., Ramstein, G., Grenier, C., & Gouttevin, I. (2012). A reassessment of lake and wetland feedbacks on the North African Holocene climate. *Geophysical Research Letters*, 39. doi:10.1029/2012gl050992.

Larrasoana JC, Roberts AP, Rohling EJ (2013) Dynamics of Green Sahara Periods and Their Role in Hominin Evolution. *PLoS ONE* 8(10): e76514. doi:10.1371/journal.pone.0076514

Lézine, A. M., Zheng, W., Braconnot, P., & Krinner, G. (2011). Late Holocene plant and climate evolution at Lake Yoa, northern Chad: pollen data and climate simulations. *Climate of the Past*, 7 (4), 1351-1362.

Now included.

Additional note from referee #1 [RC4]

As a follow-up on the interactive comment from referee #3, I would like to add some clarifying words.

Past interglacials, especially the early Last Interglacial, are often regarded as analogues for future climate change. Whether this is justified or not is an open question and I'm not asking the authors to provide any proof in favour or against this idea. However, it seems to me that the authors have made a well-considered choice not to mention the Last Interglacial (or the Early Holocene for that matter) in this context and it

Would be highly interesting for the larger community, especially considering the wealth of knowledge on the topic held within the long list of authors, if these considerations would be part of the manuscript

See response to Referee 1 point 2. We have added some discussion to the Conclusions.