Response to reviews of "The PMIP4 contribution to CMIP6 - Part 3: the Last Millennium, Scientific Objective and Experimental Design for the PMIP4 *past1000* simulations" by Jungclaus et al.

For clarity, we reproduce the comments by the reviewers and the editor in blue/italic and provide answers in black. Changes to the manuscript are presented in bold face.

Reviewer 1:
This paper provides a useful overview of the objective and setup of the planned PMIP4 simulations covering the preindustrial millennium. It is generally well written and clear, and I have only a few minor suggestions for improvement. This paper will serve the community well and I would recommend accepting it for publication in GMD.

We thank the reviewer for his/her positive evaluation and suggestions that we found helpful to improve the manuscript. In the following, we address all the comments and suggestions.

Introduction. In my view, the introduction could be improved by stating more clearly the objective of this paper and the added value compared to PMIP3-CMIP5. The main improvements relative to PMIP3 are summarized in Section 6 (Conclusions), but after reading the last two paragraphs of the introduction, it was not clear to me what the exact innovation is in the PMIP4 past1000 simulations compared to PMIP3. So I suggest to revise the main paragraph on page 4 (starting at line 5) to clarify this point.

We have followed the reviewer's suggestion and included the following at the place suggested:

Further progress is expected for CMIP6 and PMIP4. Models with higher spatial resolution will be available for long-term paleo simulations, which has the potential to improve the representation of mechanisms controlling regional variability and to alleviate biases in the mean state (e.g. Milinski et al., 2016). Newly added model components, for example interactive chemistry and aerosol microphysics, will allow for more explicit representation of forcing-related processes in some models (LeGrande et al., 2016), and, as we outline below, improvements in forcing reconstructions regarding their accuracy and complexity will potentially lead to improved model data comparison. In addition, more stringent protocols for experimental set-ups and output data are implemented in the CMIP6 process, which also ensures a better interaction between related MIPs. For example, the PMIP4 past1000 experiment is closely related to the more process-oriented suite of simulations in the Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP, Zanchettin et al., 2016).

Scope of the paper. The title and the introduction suggest that this paper is about the tier-1 past1000 experiments, but in fact also the forcings for the historical simulations are discussed. Is there a separate paper planned to explain the setup of the historical simulations in detail? If not, I would suggest to slightly modify the title to broaden the scope.

The point is that it is very important to complement the past1000 simulations covering 850 to 1849 CE with historical simulations for the industrial period (1850 to 2014CE).

PMIP4 recommends strongly that these historical simulations will be carried out with the official CMIP6 historical forcing data sets that are documented in Eyring et al., GMD, 2016 and various other contributions to the GMD special issue. For some of the drivers, we are in the lucky situation that the forcings for the CMIP6 historicals have already been extended back in time to cover either the entire CE (as for the GHG forcing), or the last millennium (850 to 2014 CE for the land-use forcing. For other forcings, e.g. solar we have made an effort to provide a smooth transition between the pre-industrial and the historical forcings. This is documented in the respective sections of our manuscript. Since the 1850-2014 simulations following the past1000s fulfil all the requirements for CMIP6 "historicals" we do not intend to describe and document them in this manuscript. We have however, underlined more clearly that it is mandatory to add a "historical" simulation that is initialized with the 1849 conditions from past1000. Therefore, we have included the following statement at the end of the first paragraph of the introduction:

We emphasize, that the *past1000* simulations must be complemented by *historical* simulations for 1850 to 2014 CE following the CMIP6 protocol and applying the CMIP6 external forcing for the industrial period (Eyring et al., 2016 and references therein).

Section 3.3. I find the explanation of the tier-2 experiments rather vague. For instance, what is the time period to be covered in these simulations? Is it also the full 1000-year period of 850-1849 CE? Will the same initial conditions be used as in the tier-1 experiments? If the models are run in ensemble mode, what is the recommended number of ensemble members? I suggest making this section more specific.

We have added information on the simulation period etc. in the text

The "tier-2" past1000 experiment should be set-up in a similar way as the "tier-1" past1000 simulation, i.e. the simulation should cover 850 to 1849 CE and the same initial conditions should be used.

Regarding the ensemble size, we can probably not demand too much, therefore we have included the following sentence:

While an ensemble size of ten has been shown to be desirable (Otto-Bliesner et al., 2016; Stevenson et al., 2016), we acknowledge that limits in computational resources or high computational demand of high-resolution models may prevent groups from conducting large ensembles.

Minor comments
Page 1, line 10. 'This is particularly acute for regional and sub-continental scales'. I suggest specifying what regions are of special interest here.

At this point in the introduction we would prefer not to go into details regarding individual regions. We have, however, changed the sentence to better specify the issue of spatial inhomogeneity and included a reference to a recent paper by Gagen et al., who discuss sub-continental spatial variations over Europe:

This is particularly acute for regional and sub-continental scales, where spatially heterogeneous variability modes potentially impact the climate signal (e.g., PAGES2k-PMIP3 Group, 2015; Luterbacher et al., 2016; Gagen et al., 2016).

Page 1, line 15. 'preindustrial millennium'. Please explain here that you mean the 850-1849 CE period.

Done

Page 6, line 11. 'a updated forcing datasets'. Remove 'a'.

Done

Page 6, Section 3. Please briefly explain already here what the difference is between tier-2 and tier-3 experiments.

We included the following statement here:

In contrast to the PMIP3 protocol, PMIP4-CMIP6 recommends a single collection of external forcing data sets (the default forcing) in the "tier-1" experiments while encouraging exploration of forcing uncertainty as part of dedicated "tier-2" experiments. Whereas these "tier-2" experiments only differ in the characteristics and combination of the external drivers from the "tier-1" *past1000* experiment, additional "tier-3" experiments are designed to allow clusters of modelling groups to perform dedicated research by exploring either specific episodes or advancing the scope of the *past1000* experiments by extending them in time.

Page 6, line 26. I suggest mentioning here that the historical simulations cover 1850-2014 CE, and not in line 31.

Done

Page 8, Section 3.4.2. Will the forcings for the past2K and past1000 experiments be identical for the period 850-1849 CE? I suggest clarifying this.

Yes, the forcing should be continuous. We have included the following statement after the first sentence of section 3.4.2. to clarify this:

In fact, except for the land-use change forcing, all forcing reconstructions described above for the "tier-1" *past1000* experiment are available for the entire CE and the groups need to make sure that the same forcing is used for *past1000* and *past2k* during the period 850 to 1849 CE.

Page 8, Section 3.5. I suggest briefly explaining here or in the Table caption the meaning of the capitals N, M and L.

We have expanded the description of the table in the text accordingly:

The experiments are defined by their short name (e.g., *past1000*) and an extension following the "ripf" classification, where "r" stands for "realization, "i"

for initialization, "p" for perturbed physics, and "f" for forcing (Table 1). The letters r, i, p, and f are followed by integers N, M, P, and L, respectively. For example, different realisations within an ensemble would have different values for "N". To classify a simulation with a model with modified physical parameterization, one would vary the integer "P". The experiments using the default forcing are defined by "f1", alternative or single forcing would be identified by a different integer value "L". It is the responsibility of the modelling groups to document the choices and settings.

Page 9, Section 4.2. What is the difference in the GHG radiative forcing compared to PMIP3-CMIP5? Please explain.

We have included the following addition:

Differences between the new CMIP6 data set and previous estimates for CMIP5 are rather small (e.g., for global mean surface mixing rations see figure 9 in Meinshausen et al., 2016). The CMIP6 reconstruction offers better representation of latitudinal and seasonal variations and we recommend using this data set for consistency throughout the CE.

Page 9, line, 8: 'Discrepancies in proxy-based temperature records'. Why are the temperature records mentioned here in the section on volcanic forcing? Please clarify.

We did not make it clear enough that we mean the discrepancies in timing of cooling events in temperature reconstructions and the occurrence of the sulphate signal in the forcing reconstruction. We have rephrased this sentence:

Discrepancies in the timing of volcanic events recorded in ice cores and short-term cooling events in proxy-based temperature records have been largely resolved by improvements in absolute dating of the ice core record (Sigl et al., 2015).

Pages 10-11, Section 4.4. What is the difference in solar forcing compared to PMIP3? Please elaborate.

We have included a summary paragraph at the end of the section on solar forcing that describes what is new and more robust in PMIP4:

In summary, PMIP4 provides three reconstructions of TSI and SSI from the most-up-to-date records of cosmogenic radioisotopes ^{14}C and ^{10}Be using a chain of models, all of which have been improved and updated since PMIP3. In contrast to CMIP3, for all provided reconstructions, total and spectral irradiance are computed in a self-consistent manner. In particular, the same model has been used to reconstruct irradiance from each radioisotope to allow an estimate of the uncertainty due to the effect of local conditions on their formation and deposition. Two irradiance reconstructions were obtained from ^{14}C data using different irradiance models to allow for sensitivity experiments testing the response to the amplitude of the solar forcing. The default forcing for CMIP6-PMIP4 *past1000* is the ^{14}C SATIRE-based reconstruction. The PMOD-based reconstruction provides

an upper limit on the magnitude of the long-term changes in irradiance. Since the historical CMIP6 recommendation is an arithmetic average of two conceptually different models with significant differences in the SSI variability, special care has been taken to combine the PMIP4 data sets with the historical forcing. The approach we have chosen here allows for a smooth transition but might nevertheless produce some artefacts.

Page 11, line 26. Will ozone variations be provided by PMIP4 for the period 850-1849 CE? Please discuss.

We have updated the section on solar-related ozone changes. This part was provided by Amanda Maycock, who we suggest to include as co-author. At the end of section 4.4, we now specify:

Hence we have re-performed the regression of the same ozone fields but with respect to solar UV irradiance averaged over the spectral range from 200 to 320 nm. We recommend calculating time varying ozone input for PMIP4 by scaling these coefficients with the anomaly of the respective UV flux during the simulation period and add it to the CMIP6 preindustrial ozone climatology. The UV flux anomaly should accordingly be calculated with respect to the CMIP6 preindustrial irradiance data (Matthes et al., 2016).

Page 11, last line. Klein Goldewijk et al. 2016. The reference list only mentions Klein Goldewijk 2016, so without co-authors. Is the reference in the list incomplete?

We have corrected Klein Goldewijk et al., 2016 to Klein Goldewijk, 2016.

Page 12. Section 4.5 discusses quite extensively the wood consumption. I wonder if this paper is the right place for this discussion, as it seems incompatible compared to level of detail in the rest of the manuscript. Wouldn't it fit better in a manuscript on LUMIP?

We agree that the level of details on the reconstruction of wood consumption is a bit out of balance compared with the other ingredients to the forcing. We have shifted a larger part of this paragraph to the appendix section.

Reviewer 2:

In this manuscript the authors describe the major goals of the last millennium experiments within the forth phase of PMIP, and the experimental protocol that have been proposed to address them. This is an important well-organised initiative that will shed new light on both the internally driven and externally forced contributions to the climate of the last millennium, and will complement other additional efforts by the paleoclimate community (e.g. PAGES2K). Therefore, I find the article timely and worthy of publication in Geoscientific Model Development. The paper is well written and the experimental protocol is well justified and thoroughly explained. There are, however, some key choices of the experimental setup that could be better highlighted (see points below). I thus recommend

acceptance pending a few minor clarifications and comments that would need to be addressed.

We thank the reviewer for his/her positive evaluation of our manuscript and the suggestions that we found helpful to improve the manuscript. In the following we will address all comments and suggestions.

#1 I think that the article would benefit if the default forcings for the Tier1 experiments were more clearly synthetized, e.g. summarized in a Table and/or highlighted in the legends of the different figures. Otherwise, that key information is scattered throughout the text, and not always easy to find.

Thanks for pointing this out. We have followed the reviewer's suggestion and included such a table as table 2. The table includes also direct links to the respective forcing data repositories.

#2 This article describes the third part of the PMIP4 contribution to CMIP6, but there is no mention to the other parts (are there more than three?), and how they complement with each other. A brief explanation in the introduction would be helpful.

We have clarified this point and provide a paragraph in the introduction that puts our paper into the context of the suite of PMIP4/CMIP6 papers.

This paper is part of a suite of four manuscripts documenting the PMIP4 contributions to CMIP6. Kageyama et al. (2016a) provide an overview on the tier-1 experiments dedicated to CMIP. More specific information on the other paleoclimate experiments as well as the design of additional tier-2 and tier-3 experiments are given in the contributions for the mid-Holocene and the previous interglacial by Otto-Bliesner et al. (2016), for the last glacial maximum by Kageyama et al. (2016b), for the mid-Pliocene warm period by Haywood et al. (2016), and the present manuscript on the last millennium. Our *past1000* manuscript is organized as follows….

#3 I presume that the notation past1000 comes from the previous PMIP3 experimental protocol, and have been kept for coherence. I, however, think that the term is misleading, as it seems to suggest that the experiments cover the past millennium. But instead they target the "preindustrial" last millennium. I don't think that it's worth to change it now, but a more appropriate term could be considered in the future (e.g. preind1000).

We agree that the name past1000 could be misleading, but we are not in a position to change it now. Wes hall keep this issue for further discussion on transient simulations within PMIP.

#4 [Page 3, lines 7-10] I would recommend rephrasing this sentence for clarity. For example, to something of the sort of ": : :the relative contribution of internal variability and external forcing factors to natural fluctuations in the Earth's climate system: : :".
#5 [Page 4, line 15]

done

Two other relevant articles that could be cited here are Lehner et al (2012) and Ortega et al (2015).

Included

#6 [Page 4, line 37] As it is written, it seems to imply that the MCA-LIA transition is only explained by these clusters of eruptions. But changes in solar irradiance most probably played some (minor) role. I suggest rephrasing to "Clusters of eruptions have been identified as the major contribution to the transition: : :"

done

#7 [Page 6, lines 10-12] Remove "a" from "a updated". The final part could also be slightly rephrased to "a new generation of climate models in which the different forcings will be better represented". Also, it is not clear to me if this sentence refers exclusively to the changes in land-use, or to all the forcings previously described. If it's to all forcings, it might work better at the end of the paragraph (as the next sentence refers only to land-use changes).

We agree that this did not fit very well. We have deleted the sentence on further progress in the land-use forcing (as it is irrelevant for the existing protocol) and moved the more general statement on the new generation models to the first paragraph of section 2:

The PMIP4 experiments will revisit the questions regarding the relative role of external drivers using updated forcing datasets and a new generation of climate models, in which the different forcing will be better represented.

#8 [Page 5, line 13] Correct to "initiative".

This sentence was deleted

#9 [Page 7, lines 4-8] It is not totally clear to me from this paragraph whether there are two different sets of historical CMIP6 simulations according to their initial conditions (are they taken from picontrol experiments, past1000 experiments, or both?). Is that why you say that it will be possible to assess the impact of initial conditions on the climate of the 19th and 20th centuries?

Indeed, there will be two sets of *historical* simulations. We designed the experimental set-up so that the 1850 to 2014 CE *historical* simulations should be identical in terms of forcing (and, as such serve as contributions to the CMIP6 multi-model ensemble of historical simulations), but differ in the initial conditions. On the other hand, we want to make sure that we have continuous simulations from the pre-industrial past to the historical period. We have slightly modified the text to make this point more clear:

The standard PMIP4-CMIP6 *past1000* experiment applies the default forcing data set (see below) and is complemented by an *historical* (1850 – 2014 CE) simulation that uses the end state of the *past1000* simulation in 1850 CE for initialization and

that follows the CMIP6 protocol (Eyring et al., 2016). This procedure provides a consistent data set for past and present climate variations. Comparing historical simulations initialized from a *piControl* run (the CMIP6 default) with those starting from 1849 CE conditions from *past1000* serves to assess the impact of initial conditions on the evolution of the 19th and 20th century climate.

#10 [Page 13, line 21] Change to "impacted-related".

Done

#11 [Page 14, line 17] I suggest specifying "new climate reconstructions", to distinguish from forcing (reconstructions) just mentioned before.

We changed this to:

…new reconstructions of climate variables…

to emphasize that it is not only about temperature.

#12 [Page 14, first and second paragraph] These two collaborations with PAGES2K to investigate the past changes in the ocean circulation and hydroclimate are really important to bridge the existing gaps between models and paleo records. Will key variables for these model-data intercomparison studies, such as the AMOC and barotropic streamfunction and some drought severity indices, be consistently stored by the different modelling groups?

Yes, this is a major effort that has been done in the CMIP6 community. We have added the following statement to the end of section 5.

PMIP has provided to CMIP6 a comprehensive list of output variables that includes all necessary variables for analyses of atmospheric, oceanic and land-surface processes (see Appendix 6). CMIP6 will make sure that all groups store the output variables in a consistent way (see https://earthsystemcog.org/projects/wip/CMIP6DataRequest).

Please see also our response to the editor's comments below.

Comments by the Editor (J.C. Hargreaves):

Please can you include details of where the model output data will be stored. Is it all to be uploaded to ESGF? If other databases are to be used, please give details and explain the terms of use. This basic information could be added to the data availability section. Secondly, please provide a list or table detailing the required output variables for the experiments. This could be added as an appendix or supplement.

We have expanded the Appendix 6 to include more information on data distribution and the list of variables requested by PMIP4. We include a link to the CMIP6 data request

web page for PMIP and we will provide a list derived from the CMIP6 excel files that can be attached as supplementary material to our manuscript.

The "tier-1" *past1000* simulation is part of the CMIP6 experiment family and data will be distributed through the official CMIP6 channels via the Earth System Grid Federation (ESGF, https://earthsystemcog.org/projects/wip/CMIP6DataRequest). Data from PMIP4-only "tier-2" and "tier-3" simulations must be processed following the same standards as 'tier-1" for data processing (e.g. CMOR standards) and should be distributed via ESGF. Modelling groups producing these simulations are responsible to secure suitable space on ESGF nodes.

The list of variables requested for the PMIP4-CMIP6 palaeoclimate experiments can be found here: http://clipc-services.ceda.ac.uk/dreq/u/PMIP.html, and as a supplement to this manuscript.

The PMIP4 variable selection reflects plans for multiple analyses and for interactions with other CMIP6 MIPs (see Kageyama et al., 2016). The only variables defined specifically in PMIP are those describing oxygen isotopes for model systems that calculate these data interactively (Kageyama et al., 2016).

Groups contributing *past1000* simulations to CMIP6-PMIP4 should ideally deliver the entire set defined in the data request. However, an important issue for long-term simulations such as *past1000* is storage demand for high-frequency output. As a minimum, we ask for a subset of daily variables that allow investigations on extreme events and particular dynamical features:
Near surface air temperature (tas), daily maximum near surface air temperature (tasmax), daily minimum near surface air temperature (tasmin), daily maximum near-surface wind speed (sfcWindmax), precipitation (pr), daily 500 hPa geopotential (zg500), daily maximum hourly precipitation rate (prhmax).

Groups participating in PMIP and VolMIP should pay attention to the new diagnostics of volcanic instantaneous radiative forcing defined by VolMIP, whose calculation is recommended for some major volcanic events simulated in the past1000 experiment (for details, see Zanchettin et al., 2016). Groups that run the PMIP4-CMIP6 experiments with the carbon cycle enabled should pay attention to the output variables requested by OCMIP and C4MIP.

Further unsolicited changes

We have refined the description of the solar forcing at several places and have collaborated with Prof. Raimund Muscheler (Lund University). We received important input from R. Muscheler regarding the methodology presented in section 4.4. In particular, the comparison between his recently published neutron-monitor based estimate for the solar modulation has increased confidence in our solar forcing reconstruction.
We would like to include R. Muscheler as co-author in the revised version of our manuscript

We have further elaborated on the parameterization of solar-related ozone changes. We received important input from Dr. Amanda Maycock (University of Leeds), who has also been instrumental in defining the respective forcing for the CMIP6 historical experiments. We therefore wish to include A. Maycock as co-author in the revised version of our manuscript.