Response to comments by Erik Kjellström

Comments in *italics*, response in normal script, suggested changes to text in **bold.**

I'm not an expert in land-use or past changes in land-use but as a climate modeler with some limited experience in paleoclimate modelling I think that the paper would benefit from some more detailed discussion of potential limitations with the formulated strategy. In particular, the results illustrating the methods show: large spread, poor correlation and small differences between experiments with and without land-use (Figure 8). This could compromise the idea constraining land use change by climate model simulations.

These plots show the direct comparison between gridded values of simulated mean annual temperature, mean temperature of the coldest month and mean annual precipitation and reconstructions as reconstructed from pollen data at these same gridcells. The spread is therefore not indicative of uncertainty, as suggested by the reviewer, but the geographic spread in climate across the region. The motivation for including anthropogenic land use in these experiments was the fact that there is a poor correlation between simulated and observed climate in the original experiment without land use changes. LULC was implemented using KK10. The plot shows that the correlation becomes slightly better for MAP but does not improve significantly for MAT and becomes worse for MTCO. We already know from comparisons with pollen data that the KK10 scenario is not "perfect" and this is our motivation for improving the scenario -- so it would be hoped that the "improved" scenario leads to a better simulation of the climate. Certainly, if it does not lead to an improvement, then it will be meaningless to interpret the simulations as confirming the importance of LULC for correct simulation of climate during the Holocene. We have modified the caption to this figure in response to a specific comment (see below). We will modify the text describing this figure to clarify the expectations about the climate model tests, as follows:

A second test of the realism of the improved LULC scenarios is to examine whether incorporating LULC changes improves the realism of the simulated climate when compared to palaeoclimate reconstructions (Figure 8). The mid-Holocene (6000 years ago, 6ka BP) is an ideal candidate for such a test because benchmark data sets of quantitative climate reconstructions are available (e.g. Bartlein et al., 2011), the interval has been a focus through multiple phases of PMIP and control simulations with no LULC have already been run, and evaluation of these simulations has identified regions where there are major discrepancies between simulated and observed climates e.g. the observed expansion of northern hemisphere monsoons, climate changes over Europe, the magnitude of high-latitude warming, and wetter conditions in central Eurasia (Mauri et al., 2014; Harrison et al., 2015; Bartlein et al., 2017). There are discernible anthropogenic impacts on the landscape in many of these regions by 6 ka, although they are not as strong as during the later Holocene and they are not present everywhere. Nevertheless, the 6ka BP interval provides a good focus for testing improvements to the LULC scenarios. Such an evaluation would need to go beyond the global comparison made here (Figure 8) to regional comparisons to identify whether improvements in regions where there is a large anthropogenic impact on land cover do not result in a degradation in the simulated climate elsewhere.

In parallel to the climate model uncertainty, what is the uncertainty associated with the carbon cycle models proposed to be used for constraining the land use? Is it small enough to allow for a meaningful estimate of land use? I think the paper would benefit from a more in-depth discussion about these uncertainties.

It is important to separate out the two applications of the carbon-cycle model simulations: first as a test of whether the scenarios are plausible and second as part of the transient Holocene climate simulations. In the offline simulations, we will use a single climate forcing but the intention is to use multiple carbon-cycle models - and this will allow us to evaluate the uncertainty associated with different models. This perhaps should have been made clearer. The planned transient model intercomparison further serves to address model uncertainty by design in using an ensemble of model simulations. This allows us to quantify model spread and therefore account for uncertainty related to differences between the models. However, the fact that planned simulations cover a very large temporal (~12 kyr) and spatial (global) scale, restricts the possibilities to assess uncertainties in a more systematic way. In particular, with our activity, we do not aim at quantifying parametric model uncertainty because this would require a (very) large ensemble (on the order of thousands) of simulations with each individual model. This is not feasible. A single global model simulation covering 12 kyr takes on the order of weeks even for the fastest global models.
We will expand the text describing the initial testing of the scenarios using carbon-cycle models to make it clearer that this is envisaged as a multi-model test, as follows:
Transient simulations with a model that simulates CO_2 emissions in response to anthropogenic LULC can be used to test the reliability of the LULC changes through time, by comparing results obtained with prescribed LULC changes through time against a baseline simulation without imposed LULC. Here we envisage using several different offline carbon-cycle models for this purpose in order to take account of uncertainties associated with inter-model differences. The carbon-cycle simulations will be driven by climate outputs (temperature, precipitation and cloud cover) from an existing transient climate simulation made with the ECHAM model (Fischer and Jungclaus, 2011) and CO_2 prescribed from ice-core records. The CO_2 emission estimates from these two simulations will then be evaluated using C budget constraints. This evaluation will allow us to pinpoint potential discrepancies between known terrestrial C balance changes and estimated LULC CO_2 emission in given periods over the Holocene.

Consideration could also be given if there would be a place for more detailed regional and local studies to further constrain land use?
It is unclear what the reviewer is asking for here. The archaeological investigations are being carried out at a local scale and provide detailed regional records for some regions, which are then generalised for to continental scales. Both the detailed regional results and the continental maps will be used as inputs into the global LULC scenarios. The LULC scenarios necessarily have to be global for input into the climate model simulations. Similarly, the pollen-based constraints are site based and we have very detailed information on land use for some regions (e.g. Europe, North America) and less detailed information for others (e.g. tropics). Our evaluations will naturally make use of the detailed information where available.

General comments:
Some words and concepts are quite difficult for a climate modeler (definition of time periods like the Holocene and Mesolithic and Neolithic times, taphonomic (L190)). The manuscript needs to be checked for consistency in how time is referenced (sometimes 6 ka BP, sometimes 6 ka). Also please explain what this means at the first reference.
These points are raised below in the line-by-line specific comments, and our responses (and changes) are given there.

Line-by-line specific comments:

L1: Please don't use LULC in the title, better to spell out what it is about.
We will change this to read:
Development and testing of scenarios for implementing land use and land cover changes during the Holocene in Earth System Model experiments

L36: Unclear what is meant by "Current LULC scenarios". Is it current scenarios for the Holocene? Which part of the Holocene? Or, is it scenarios of LULC for the current climate (likely not, but it should be made more clear).
We are referring to scenarios of LULC during the Holocene. We will clarify this as follows:
Existing LULC scenarios of LULC changes during the Holocene are based on relatively simple assumptions and highly uncertain estimates of population changes through time.

L42-45: From this it is unclear if the paper is just on evaluation of scenarios or if it is also about further refinement of the scenarios.
Our goal here is to provide a protocol for refining existing scenarios iteratively so that these scenarios can be used for climate model experiments. We realise that the abstract does not make this clear and will modify it as follows:
In this paper, we document the types of archaeological data that are being collated and how they will be used to improve LULC reconstructions. Given the large methodological uncertainties involved, both in reconstructing LULC from the archaeological data and in implementing these reconstructions into global scenarios of LULC, we propose a protocol to evaluate the revised scenarios using independent pollen-based reconstructions of land cover and climate. Further evaluation of the revised scenarios involves carbon-cycle model simulations to determine whether the LULC reconstructions are consistent with constraints provided by ice-core records of CO_2 evolution and modern-day LULC. Finally, the protocol outlines how the improved LULC reconstructions will be used in palaeoclimate simulations in the Palaeoclimate Modelling Intercomparison Project to quantify the magnitude of anthropogenic impacts on climate through time and ultimately to improve the realism of Holocene climate simulations.

L44: What kind of "carbon-cycle simulations" are referred to here? Earth-system model simulations? Carbon cycle model simulations? Anything else?
We have modified the abstract (see above) to clarify this.

L53-54: The new IPCC special report on land states that 70% of land is being influenced by anthropogenic activities. Is there a discrepancy here?
It is obviously difficult to provide an overall estimate of how much of the land surface is affected by human activities because it depends on whether the focus is on direct appropriation for agriculture resulting in a fundamental change in land cover or whether any anthropogenic influence is being taken into account. The Land Report states (section 1.1.2.2) that between 60–85% of the total forested area and between 72-89% of non-forested land is used, but it also makes it clear that the level of usage is variable with only 10% being intensively managed, two-thirds being moderately managed and the remainder at low intensities. Only about one third of the used land is associated with changed land cover. The Report states that differences in definitions and lack of information about management practice means that the estimates of human usage are uncertain. So, in this sense our statement is compatible with the Land Report, in that the estimated 40% refers to the area being used for agriculture and we go on to say that large parts of the rest of the land area are being influenced in some way by human activities. However, our point here is not to quantify the extent of use but simply to point out that there

is considerable anthropogenic modification on the landscape globally. We will acknowledge the work of the Land Report -- which came out after we submitted this paper -- and modify this sentence as follows:
Today, ca 10% the ice-free land surface is estimated to be intensively managed and much of the reminder is under less intense anthropogenic use or influenced by human activities (Arneth et al., 2019).

We will remove the following unnecessary references
Foley et al., 2005
Ellis and Ramankutty, 2008
Ellis et al., 2010
Ellis et al., 2013
and add the reference to Arneth et al. (2019)

Arneth et al., 2019. IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse gas fluxes in Terrestrial Ecosystems.

L56-57: Please define what is meant by "Mesolithic and Neolithic".
These archaeological periods are diachronous. The Mesolithic represents the final period of hunter-gather culture, and the Neolithic is associated with the emergence of agriculture, including domestication and more permanent settlements. We will modify the sentence to make this clearer for non-archaeologists as follows:
Substantial transformations of natural ecosystems by humans began with the geographically diachronous shift from hunting and gathering characteristic of the Mesolithic to cultivation and more permanent settlement during the Neolithic period ....

L79: "LULC change during the Holocene". It is unclear what is meant here. Is it over the full Holocene? Or, from any particular time in early or mid Holocene to any point during late Holocene (preindustrial?).
We agree that this is somewhat unclear. The experiments examine the impact of the change in 1850CE but this change represents the accumulated change in LULC through the Holocene. We will modify the text, as follows:
At the global scale, the biogeophysical effects of the accumulated LULC change during the Holocene which resulted in reconstructed land cover patterns in 1850CE have been estimated to cause a slight cooling (0.17 °C) that is offset by the biogeochemical warming (0.9 °C), giving a net global warming (0.73 °C) (He et al., 2014).

L189: "lack of uniform sampling through time" – does this include different national sampling strategies/resources for archeological excavations/sampling?
Most early archaeological sites represent occupation for only a limited period of time, although the same sites may be re-occupied at a later date. Differences in research traditions and foci in different regions means that particular periods may be intensively sampled and studied, while less interesting periods of time (from an archaeological perspective) are neglected. Lack of resources and preservation issues means that it is virtually impossible to obtain a uniform sampling of archaeological records in space and time and in any case such a sampling does not currently exist for most regions. In response to a slightly different comment by Almut Arneth, we propose to modify this sentence as follows:

There are biases that could affect the expected one-to-one relationship between number of people and number of radiocarbon dates on archaeological material, including lack of uniform sampling through time and space caused by different archaeological research interests and traditions in different regions) and increased preservation issues with increasing age.

L190: What is taphonomic?
Taphonomic processes are those which result in post-deposition modification of deposits, here including decomposition or erosion. Here we simply meant to say that there is a loss of information because preservation becomes less reliable with age. We have modified the sentence (see above) to remove the jargon.

L331-343: Here, it is unclear whether the "already produced reconstructions" are products of REVEALS or if there are any other methods that have been involved.
These reconstructions, which are illustrated in Figure 7, were made using REVEALs. We will clarify this and also include an additional references to the figure at this point, as follows:
LandCover6k has already produced reconstructions using REVEALS for the northern extratropics (see e.g. Figure 7). These reconstructions provide snapshots through the Holocene the Holocene with a time resolution of 500 years until 0.7ka BP, and three historical time windows (modern–0.1ka BP, 0.1–0.35ka BP, and 0.35–0.7ka BP).

L361: Suggest changing "observed climate" to "reconstructed climate".
We will make this change (actually L357)

L386-390: Here it is discussed changes in land use over time. The text gives the impression that there is always increasing land use with time "more conversion in earlier periods implies less conversion in later periods". Seems logical, but does this argument hold in a situation when land use is fluctuating with time (e.g. no land use – some land use – forest regrowth – no land use – again more land use ...)?
We are not implying that land use always increases through time, because indeed the archaeological evidence shows that this is not the case and this is illustrated in Figure 5 for example. What we are trying to explain is that the cumulated amount of land converted to agriculture during the Holocene must sum to the amount of agricultural land today. So, if there is a lot of conversion early on, then there must either be less later or large parts of the converted land must have reverted to non-agricultural land. We will try to make this clearer by modifying the text. as follows:
First, reconstructions of the total land under agricultural use must converge on the present-day state, which is relatively well constrained by satellite land-cover observations and national statistics on the amount of land under use. Reconstructing the extent of past LULC thus reduces to allocating a fixed total amount of land conversion from natural to agricultural use over time. More conversion in earlier periods implies either abandonment of agricultural land or less conversion in later periods.

L395: "to" missing after "due".
We will correct this (actually line 390)

L440: How is land-use implemented in the models? Is it binary (i.e. 0 or 1) or fractional? In the latter case I guess that dynamical vegetation models could be used in combination with

the land use information to derive vegetation type for the part of a gridbox not associated with land use.

Land use is currently not implemented in the mid-Holocene simulations. The implementation in the CMIP6 past1000 and historic simulations varies with the model; most of the models use fractional coverage. Not all of the models include dynamic vegetation, or rather have dynamic vegetation "switched on" in their piControl experiment, but for those that do we are indeed proposing that the vegetation is simulated in that fraction of a gridcell that is not affected by LULC. We will revise the paragraph describing the mid-Holocene simulations to make this clearer, as follows:

The *mid-Holocene* (and its corresponding *piControl*) is one of the PMIP entry cards in the CMIP6-PMIP4 experiments (Kageyama et al., 2018; Otto-Bliesner et al., 2017) and it is therefore logical to propose this period for LULC simulations. The LULC sensitivity experiment (*midHoloceneLULC*) should therefore follow the CMIP6-PMIP4 protocol, that is it should be run with the same model components and following the same protocols for implementing external forcings as used in the two CMIP6-PMIP4 experiments (Table 1). Thus, if the *piControl* and *midHolocene* simulations are being run with interactive (dynamic) vegetation, then the *midHoloceneLULC* experiment should also be run with dynamic vegetation in regions where there is no LULC change. For most models, this means that the LULC forcing is imposed as a fraction of the grid cell and the remaining fraction of the grid cell has simulated natural vegetation.

L444-445: "free atmospheric CO2" needs a better explanation – for instance something like "…, allowing atmospheric CO2 concentrations to evolve in concert with fluxes to and from land and oceans".

We will change this to:

Thus, modelling groups who are running the *midHolocene* experiment with a fully interactive carbon cycle could also run the LULC experiment allowing atmospheric CO_2 to evolve interactively, subject to the simulated ocean and land C balance.

L466: Please elaborate a bit on how good the assumption on "equilibrium" is for the Mid-Holocene? Was the carbon cycle (and climate) at equilibrium at that time?

In the text, we are referring to starting the transient experiments from the mid-Holocene experiment because these equilibrium experiments are mandated to have a long enough spin-up to be in equilibrium before the experiment is run (see Otto-Bliesner et al., 2017). Whether the carbon cycle and climate was at equilibrium in the real world is not an issue. In the present context, where we address LULC CO_2 emissions that evolve over centuries to millennia, disequilibrium effects are relatively small. This is due to the much shorter time scale of emissions occurring after forest clearance (on the order of years to decades). The longer time scales of forest regrowth (centuries) might be relevant too, where agricultural land abandonment and forest regrowth are important. We will clarify the issue of the mid-Holocene experimental equilibrium in the protocol, as follows:

We suggest that this transient simulation (*holotrans*) should start from the pre-existing *midHolocene* simulation to capitalise on the fact that the *midHolocene* simulation have been spun up for sufficiently long (Otto-Bleisner et al., 2017) to ensure that the ocean and land carbon cycle is in equilibrium at the start of the transient experiment (Table 2).

L482-488: All references here are more than 10 years old. Are there no more recent studies of relevance?

Unfortunately, there are no more recent continental scale reconstructions of climate through the Holocene -- although there are ongoing projects that are planning to revisit these

reconstructions for Europe and North America taking advantage of more extensive pollen data sets and newer reconstruction techniques. There are newer reconstructions for Europe and the USA for individual sites, but site-based model evaluation is difficult and here we only give references to individual sites in regions where there are no continental-scale reconstructions. We could add more references to reconstructions at individual sites but perhaps it would be better to clarify why such data is not particularly helpful for model evaluation, as follows:
Quantitative climate reconstructions through the Holocene at a regional scale are currently only available for Europe (Davis et al., 2003) and North America (Viau et al, 2006; Viau and Gajewski, 2009). There are time series reconstructions for individual sites outside these two regions (e.g. Nakagawa et al., 2002; Wilmshurst et al., 2007; Ortega-Rosas et al., 2008), but it is difficult to rely on such reconstructions for model evaluation because of the differences in resolution between the models and the geographic scale sampled by individual sites. However, the simulated time-course of CO_2 emissions can be compared to the ice core records.

Figure 1: The color scale with the relatively dark green makes it difficult to see any of the rather small areas with land-use. It is difficult to understand why these two years have been chosen from the datasets (why not use the same reference year?). The font size at the color bar is too small.
The two data sets (KK10, HYDE3.2) do not have outputs for every year and so we have chosen the two available intervals that correspond most closely to the mid-Holocene time interval from each. They are 50 years apart, which given the uncertainties on radiocarbon dating of this time interval can be considered indistinguishable from one another. We will redraw this figure (and the other figures) to ensure that the font size is readable throughout.

Figure 2: The figure is difficult to read and it is not easy to see what is the final outcome of the scheme based on the figure. If it is something like "LULC scenario" I guess this should be something popping out on the right-hand side after going through the three steps in Phases 1-3. Also, it is not clear from the figure if there is any iterative part in the process where info is added to the scenarios based on constraints from phases 2-3? This could be better explained here and would also help to make the paper a bit more clear on a general level.
We have expanded our description of this Figure and the general protocol in response to comments by Almut Arneth. We will redraw this figure (and the other figures) to improve readability.

Figure 3. Here, font sizes are too small everywhere. What is SDPs? Please explain what the shading is for the maps (areas under human use?) and give a color bar. What are the circles in the lowermost panels?
In addition to revising the figure to improve readability, we will change the to explain the abbreviation SDP and the shading, as follows:
Reconstruction of changes in population size in the Iberian Peninsula during the Holocene (9000 to 2000 BP, 9ka to 2ka BP) using summed probability distributions (SPDs) of radiocarbon dates (data after Balsera et al., 2015). The red line indicates the onset of agriculture in the region. The lower panels show areas under human use at 6ka (left) and 4ka (right) using kernel density estimates, where the white dots are actual archaeological sites and the shading shows the implied density of occupation.

Figure 4. Here is a box (Extensive/Minimal land use) that lacks some Level 2/3 information. Or it is redundant and can be removed? The labels on the land-use classes are quite specialized and several of the words are not everyday terms from my perspective

(pastoralism, chinampas, taro pondfields, Peri-urban, Swidden). It would be good if these were a bit better explained, alternatively use different words). Also, why are there only Level 3 boxes for some of the Level 2 boxes?

The Figure is included for illustrative purposes and shows the scheme of land-use categories developed by LandCover6k to be used by the archaeological community to map land-use in different regions of the world. The terminology is that used to describe different kinds of agriculture by archaeologists, and there is a handbook (which we can refer to) that defines these terms. As we explain in the text, these land-use types will have to be translated to the anthropogenic land-use types used in ALCC scenario models and then trasnslated again in land-use harmonization schemes to produce quantitative estimates before being used for climate model simulations. The level of categorisation that is possible or necessary varies depending on the type of land use: it is clearly not useful to subdivide categories such as "no human land use" or "extensive/minimal land use". In the same way, there is no basis for subdividing some of the level 2 categories. For example, if there is "specialised fish production" it doesn't much matter what kind of fish are being farmed whereas if there is wet cultivation it does matter what type of crop is being grown and whether the wetland was natural or created for the purpose. We have already expanded this paragraph somewhat in response to comments by Almut Arneth, but we will further refine it to clarify the scheme as follows:

Maps of the distribution of archaeological sites or of areas linked to a given food production system have been produced for individual site catchments or small regions (e.g. Zimmermann et al., 2009; Barton et al., 2010; Kay et al., in press). LandCover6k is developing global land-use maps for specific time windows, based on a global hierarchical classification of land-use categories (Morrison et al., 2018) based on land-use types that are widely recognised from the archaeological record. At the highest level, the maps distinguish between areas where there is no (or only limited) evidence of land use, and areas characterized by hunting/foraging/fishing activities, pastoralism, agriculture, and urban/extractive land use (Fig. 4). Except in the cases where land use is minimal (no human land use, extensive/minimal land use), further distinctions are subsequently made to encompass the diversity of land-use activities in each land-use type (Fig. 4). A third level of distinction is made in the case of two categories (agroforestry, wet cultivation) where there are very different levels of intervention in different regions. Explanations of this terminology are given in Morrison et al. (2018). The LandCover6k land-use maps (see e.g. Fig. 5) will be based on different methods ranging from kernel-density estimates to expert knowledge depending on the quality and quantity of the archaeological information available from different regions.

Figure 5. This figure is not easily readable. The font size in the legends is way too small, the red dots in the upper panels are hardly distinguishable and the land-cover classes in the lowermost figure are not readable. Is the order left/right OK here? The figure indicates more people and land use at the earlier period (right panels) if I'm interpreting the figures correctly. In the figure caption "cal BC and BP" are used without definition anywhere. Also in the figure caption intervals defining the Middle and Early Neolithic time periods are given. Are these related to the more general statement on l56/57?

We will redraw all the figures to make them more readable. Indeed the figure does show that there were more people during the earlier period than the later period, and this is one of the reasons we chose this as an illustration to make the point that the impact of human activities is not unidirectional! The more general statement does not imply that the changes are unidirectional, as we have now clarified (see above). We realise that there are inconsistencies in the way time is expressed in the figures and figures captions (we do not refer to specific times in the text). We would like to keep both BP and BCE dates because the former

terminology is used by climate modellers and the Quaternary geology community, and the latter by archaeologists. However, we will define the terms consistently in each of the captions, as follows:

Figure 1: **Land use at ca 6000 years ago (6ka BP, 4000 years BCE) from the two widely used global historical land-use scenarios HYDE 3.2 (top panel, Klein Goldewijk et al. 2017a) and KK10 (bottom panel, Kaplan et al. 2011), illustrating the large disagreement between LULC scenarios at a regional scale. In both scenarios, the land-sea mask and lake areas are for the present day.**

Figure 3: **Reconstruction of changes in population size in the Iberian Peninsula during the Holocene (9000 years to 2000 years ago, 9ka BP to 2ka BP) using summed probability distributions (SPDs) of radiocarbon dates (data after Balsera et al., 2015). The red line indicates the onset of agriculture in the region. The lower panels show areas under human use at 6ka BP (left) and 4ka BP (right) using kernel density estimates, where the white dots are actual archaeological sites and the shading shows the implied density of occupat***ion.*

Figure 5: An example of regional land-use mapping. The plots show the distribution of archaeological sites superimposed on kernel density estimates of the extent of land-use based on the density of sites (top panels), and superimposed on the LandCover6ka land-use classes (bottom panels) for the Middle Neolithic (3600-3400 years BCE, 5600-5400 years BP, 5.6-5.4 ka BP) (left panels) and the Early Neolithic (3750-3600 years BCE, 5750-5600 years BP, 5.7-5.6 ka BP) (right panels) of Ireland. Data points derive from ^{14}C dated archaeological sites and distributions of settlements and monuments that have been assigned to each archaeological period following the dataset published in McLaughlin et al. (2016). In areas characterized by low-level food production, agricultural land (crop growing and grazing, combined) probably occupies between 10-15% of the total grid cell area in eastern and western coastal areas, whilst inland agricultural land likely represent 5% or less of the total grid cell area.

Figure 7: Northern extratropical (>40°N) mean fractional cover of open land at 6000 years ago (6ka BP: left panel) and 200 years ago (0.2ka BP: centre panel) estimated using REVEALS, and the difference in fractional cover between the two periods (right panel), where red indicates an increase in open land and blue a decrease (after Dawson et al., 2018).

Figure 6. Realizing that these figures are conceptual, but they still need some better illustration. What are the different "squares" in the left panel second from the top? Grid squares on a spatial map? Same question for the plots on the third row (and what is the bar with shading representing?)? Units lowermost left panel? Why is there a label "HYDE 3.x" on the top?

We have already modified the caption to this figure in response to comments from Almut Arneth (see below) to explain more clearly what this illustrative figure is about.

Schematic illustration of the proposed implementation of ^{14}C-based population estimates, date of first agriculture, land-use maps, and land-use per capita information in the HYDE model (here indicated as HYDE3.x). The archaeological data are represented as values for a grid cell in geographic space at a given time for date of first agriculture and land use, but as a time series for a specific grid cell for population and land-use per capita. In the case of population estimates, date of first agriculture and land-use per capita data, we

show the initial estimate and the revised estimate after taking the archaeological information into account in the HYDE3.x plot. It should be assumed in the case of the land-use mapping that the original estimate was that there was no land use in this region.

Figure 7. A suggestion here could be to remove the panel with the differences and make the other two a bit bigger and more easy to read (including larger font size on the color bar).
We will replot this figure to make it clearer.

Figure 8. What are all the dots in the panels? Are the sites covering large areas? Biased to some regions? Evenly spread? Are all three panels for areas north of 30N? What are the associated uncertainty bars with the proxy-based data? With the models?
The dots represent the individual grid cells where comparisons are possible. The Bartlein et al data set is a gridded data set derived from site-based pollen-based reconstructions. The original sites are certainly not evenly spread and there are more grids in some regions than others. All this information is given in the Bartlein et al. paper from which these data are sourced. As it says in the caption, all of the plots are for the region north of 30° N, and this region was chosen because it has the most even coverage. We do not show uncertainty bars here, either for the model or for the data. What we show is the strength of the relationship between the observations and the simulations in the two experiments. Nevertheless, we will expand the caption to make it clearer what this comparison involves, as follows:
Figure 8: *Quantitative comparison of the change in climate between the mid-Holocene (6ka) and the pre-industrial period as shown by pollen-based reconstructions gridded to* 2 x 2° resolution to be compatible with the model resolution *(from Bartlein et al., 2011) and in simulations with and without the incorporation of land-use change (from Smith et al., 2016). The imposed land-use changes at 6000 years ago (6ka BP) were derived from the KK10 scenario (Kaplan et al., 2011). The plots show comparisons of mean annual temperature (MAT), mean temperature of the coldest month (MTCO) and mean annual precipitation (MAP) for the northern extratropics (north of 30° N), where each dot represents a model grid cell where comparisons with the pollen-based reconstructions is possible. Although the incorporation of land use produces somewhat warmer and wetter climates in these simulations, overall the incorporation of land-use produces no improvement of the simulated climates at sites with pollen-based reconstructions.*

Comments on Table 1: Why is "Modern" paleogeography and ice sheets used instead of "piControl"? And, how (if at all?) are these two differing? In the table "LC6k" is used supposedly for "LandCover6k", please spell out. What does it mean that pasture and crop distributions are "imposed"? I guess "imposed on top of the default vegetation in the 6ka experiment".
These simulations follow the standard PMIP protocol for the mid-Holocene simulation as described by Otto-Bleisner et al. (2017). We say this in the text. These mid-Holocene simulations make no change in geography (land-sea distribution and topography) or ice sheet extent, i.e. they prescribe modern values for these. In point of fact, the real-world difference in these two things between the modern day and the pre-industrial (1850 CE) is negligible and not distinguishable at the model resolution. We will change the description of the imposition of crop and pasture in the table to read:
pasture and crop distribution prescribed from the revised scenario
We will also change the caption to clarify the relationship with the PMIP simulations, as follows:

Boundary conditions for CMIP6-PMIP4 and the mid-Holocene LULC experiments. The boundary conditions for the CMIP6-PMIP4 *piControl* and *midHolocene* are described in Otto-Bleisner et al. (2017) and are given here for completeness.