LULC response to editor's comments, January 2020

Although you have made a significant effort in addressing the comments/questions from the reviewers it is clear there is a significant amount of confusion regarding many aspects of the paper. These include whether models are used to generate datasets versus those that are used to validate, and the underlying algorithms used to generate the LULC-based focing datasets from the pollen proxy data. Given that this is a GMD and not a discipline-specific journal, I think it would make sense for you to pass through the article again with a mind to this, and also clarifying the nomenclature accordingly. Furthermore, uncertainties in the protocol will also be subject to significant uncertainties and more insight into these in the paper would be useful.

We agree that the use of the term model to apply both to climate models/ESMs, to biogeochemical/carbon-cycle models, and to models that are used to generate reconstructions (e.g. the pollen-source area model that underpins REVEALS) can be somewhat confusing. We also recognise that it was not always clear what kind of simulations we were talking about. We have been through the manuscript to standardise and enforce the nomenclature, ensuring that it is always clear when we are talking about a climate/ESM model and or model simulation, that LULC scenarios or reconstructions of LULC are consistently spelled out, that the statistical and/or process-based models to generate palaeo-reconstructions of climate and vegetation are also named consistently. In some places this makes the text a little repetitive, but we think it is better to be clear here than concise.

We have modified the final section of the paper to make it clear that there are uncertainties associated with the archaeological data and that there will therefore be uncertainties in the resulting LULC scenarios. However, the use of a wide range of archaeological information will certainly result in more realistic LULC scenarios than currently available and we have made this clear in the text. Most importantly, we are proposing to develop multiple scenarios taking the uncertainties into account, and testing these scenarios using independent data to determine which one(s) are most plausible.

A further comment on code and datasets. GMD endeavors to make paper submission code and datasets used to facilitate the research as available as possible. Would it be possible to make a limited subset of the data that was used to make the figures available on a code/data serving service (e.g., zenodo). In a perfect world it would be possible to recreate the figures with the provided code and data.

We conceived the figures in the paper as illustrative of our approach, and most of those that involve actual data draw on already public resources (cited in the text). However, we agree that it would be useful to make the specific data sets and the code used to generate these figures (Figure 1, Figure 3, Figure 5, Figure 7 and Figure 8) available.

We will reinstate the section on Data Availability and use this to document how the data and codes for these figures can be obtained (see below). Unfortunately, the person responsible for Figure 5 is currently moving universities and has been unable to upload the code/data for this figure, but this will be available shortly and we have therefor left a placeholder for the location details which we intend to fill in before publication of the article. I hope this is acceptable. The text reads as follows:

The data used for Figure 1 are publicly available. The HYDE3.2 data can be downloaded from https://doi.org/10.17026/dans-25g-gez3. The KK10 data can be downloaded from https://doi.org/10.1594/PANGAEA.871369. The data and code used to generate Figure 1 can be downloaded at: https://github.com/jedokaplan/ALCC_comparison_figure. The data and code used to generate Figure 3 are available from https://github.com/mavdlind/GMD.

The data and code used to generate Figure 5 are available from ?????. The European pollen-based reconstructions used in Figure 7 are available https://doi.org/10.1594/PANGAEA.897303. The pollen data used to generate Siberia reconstructions are available from https://doi.org/10.1594/PANGAEA.898616. An earlier version of this Figure was published in Dawson et al., 2018. The code and additional information used to generate Figure 7 is available from https://doi.org/10.5281/zenodo.3604328. The pollen-based reconstructions used in the generation of Figure 8 are available from 10.5281/zenodo.3601028. The climate model outputs used to generate Figure 8 are available from 10.5281/zenodo.3601040. The code used to generate Figure 8 is available from 10.5281/zenodo.3601011.

We unfortunately gave the wrong DOI for the HYDE3.2 paper, and we have now corrected this to: https://doi.org/10.5194/essd-9-927-2017

It was notable that there were a few grammatical and spelling errors in the revisions that need to be collected also.
We have now been through the revisions, and indeed the whole of the manuscript, to check for and correct typographic and grammatical errors. We hope that the manuscript is now free from such errors.

Response to editor's initial comments

We apologise if we have caused confusion through our inclusion of a data availability section in the manuscript. This manuscript describes the protocol for a set of envisaged simulations. As such, there are no data accompanying the paper. The climate model forcings that should be used for the planned LULC simulations are standard PMIP data sets. They have been referenced in other PMIP protocol papers (e.g. Otto-Bleisner et al., GMD 10 (2017); Kageyama et al. GMD 10 (2017) as follows: All the forcing data sets, their references, and their code can be found on the PMIP4 website (https: //pmip4.lsce.ipsl.fr/doku.php/exp_design:lgm, PMIP4 repository, 2017). The forcings will also be added to the ESGF Input4MIPS repository (https://esgf-node.llnl.gov/projects/input4mips/, with details provided in the "input4MIPs summary" link). We followed this format in our Code section. Model outputs that will be generated by groups following the protocol would normally be archived in the ESFG archive, following international standards and practice for CMIP6. The reference to PANGAEA was designed to indicate where validation data sets would be archived once they have been produced. Again, this paper indicates the type of validation that is envisaged as part of the protocol but does not present these data sets per se. We will remove the Data Availability section from the manuscript, since this paper only describes the protocol for running experiments. In this case, we would modify the final paragraph of the paper to indicate that groups wishing to run these simulations should follow standard CMIP practice for archiving as follows:

In addition to providing a protocol for the PMIP 6ka sensitivity experiments, we have devised a protocol for implementing the optimal LULC reconstructions for the Holocene in transient experiments. The goal here is to provide one of the necessary forc- ings that could be used for transient simulations in future phases of PMIP. This will allow an

assessment of LULC in these simulations, and therefore help address is- sues that are a focus for other MIPs e.g. LUMIP or LS3MIP. When these new forcings are created, they will be made available through the PMIP4 website (https: //pmip4.lsce.ipsl.fr/doku.php/exp_design:lgm, PMIP4 repository, 2017) and the ESGF Input4MIPS repository (https://esgf-node.llnl.gov/projects/input4mips/, with details provided in the "input4MIPs summary" link). Modelling groups who run either equilibrium or transient experiments following this protocol are encouraged to follow the standard CMIP protocol of archiving their simulations through the ESFG.

Response to comments by Almut Arneth

We thank Almut for her comments and suggestions. Comments in *italics*, response in normal script, suggested changes to text in **bold.**

..... one major aspect that seems missing from the approach. People need not only to eat, they also need to cook and heat, and to live. Has the group not discussed to -in addition to archaeological data- to also mine written historical records? This is probably most relevant for the last 1000+ years (rather than mid- Holocene), but surely there can be assumptions about wood requirements for building materials (analogue to a per-capita area needed to be fed: how many people would live in an 'average' house/farm and how much would this would need), shipping fleets (records from shipyards), charcoal making, furnaces for metal forging etc. I would imagine that at least in some regions this would have contributed perhaps already many centuries ago to deforestation. Could the authors comment on this aspect?

We agree that wood harvesting is an important issue. Historical wood demand estimates have been made at a regional scale (e.g. McGrath et al., 2015) and indeed estimates of wood harvest are included in LUH2 (https://luh.umd.edu/data.shtml). However, there are very few direct estimates of wood consumption on the longer Holocene timescale that is the focus of the LandCover6k work. While it would be possible to implement approaches based on e.g. population estimates and assuming constant wood use per capita, this is unlikely to be more than a first approximation but a rigorous site-by-site evaluation of wood use through time across the globe though worthwhile would be very time-consuming. Thus, we envisage that the first round of Holocene LULC experiments would focus on the impacts of agricultural expansion and that gathering data to refine population-based estimates of wood harvest could be a future focus on the work of LandCover6k. However, since we agree that this is an important issue and we should make this clear, we propose to add a paragraph to the section describing the archaeological data sources (line 255), as follows:

The harvesting of wood for domestic fires, building, and for industrial activities such as transportation, pottery-making and metallurgy is an important aspect of human exploitation of the landscape in the pre-industrial period (McGrath et al., 2015). It has been argued that even Mesolithic hunter-gatherer communities shaped their environment through wood harvesting (Bishop et al., 2015). Approaches have been developed to quantifying the wood harvest associated with archaeological settlements at specific times based on the evidence of types of wood use, household energy requirements, population size, and calorific value of the wood used (see e.g. Marston, 2009; Janssen et al., 2017). However, quantitative information on ancient technology and lifestyle is sparse and direct estimates of the amount of wood harvest through time are likely to remain highly uncertain (Marston et al., 2017; Veal, 2017). Nevertheless, by combining modelling approaches with improved estimates of population size should allow changes in wood harvesting to be taken into account in LULC scenarios.

Additional references

Bishop, R.R., Church, M.J., Peter A. Rowley-Conwy, P.A., 2015. Firewood, food and human niche construction: the potential role of Mesolithic hunter–gatherers in actively structuring Scotland's woodlands. Quaternary Science Reviews, 108: 51-75.
Janssen, E., Poblome, J., Claeys, J., Kint, V., Degryse, P., Marinova, E., Muys, B., 2017. Fuel for debating ancient economies. Calculating wood consumption at urban scale in Roman Imperial times. Journal of Archaeological Science: Reports 11: 592-599.

Marston, J.M., 2009. Modeling wood aquisition strategies from archaeological charcoal remains. Journal of Archaeological Science 36: 2192-2200.
Marston, J.M., Holdaway, S.J., Wendrich, W., 2017. Early- and middle-Holocene wood exploitation in the Fayum basin, Egypt. The Holocene 27: 1812-1824.
McGrath, M. J., Luyssaert, S., Meyfroidt, P., Kaplan, J. O., Burgi, M., Chen, Y., Erb, K., Gimmi, U., McInerney, D., Naudts, K., Otto, J., Pasztor, F., Ryder, J., Schelhaas, M. J., & Valade, A. (2015). Reconstructing European forest management from 1600 to 2010. *Biogeosciences, 12*(14), 4291-4316. doi:10.5194/bg-12-4291-2015
Veal, R., 2017. Wood and charcoal for Rome: towards an understanding of ancient regional fuel economics, In de Haas, T. & Gijs, T. (eds), *Rural communities in a globalizing economy: new perspectives on the economic integration of Roman Italy,* Brill, (New York and Leiden): pp.388-406.

Specific comments

Lines 63-65: For correctness, I would avoid using the term "feedback" here in the sense of change in process A affects process B, feeding back to A. LUC impacts on the carbon cycle are nothing more than an additional emission (or uptake), similar to other anthropogenic emissions, and the biophysical processes are related to albedo or ET change – but these are not feedbacks.
We agree that this was not correctly phrased and will change this to:
Direct climate impacts occur through changes in the surface-energy budget resulting from modifications of surface albedo, evapotranspiration, and canopy structure (biophysical impacts, e.g. Pongratz et al., 2010; Myhre et al., 2013; Perugini et al., 2017). LULC affects the carbon cycle through modifications in vegetation and soil carbon storage (biogeochemical impacts, e.g. Pongratz et al., 2010; Mahowald et al., 2017) and turnover times, which changes the C sink/source capacity of the terrestrial biosphere.

Lines 89-99: might be worth pointing out that the large discrepancies between Hyde and KK10 arise mostly from the assumptions about per-capita land requirements; to my knowledge their estimates of historical population changes through time (at least global totals) are more or less the same.
We agree that we could be more explicit here and will change the sentence to read:
However, differences in the underlying assumptions about land-use per capita, which are generalized from limited and often site-specific data, have resulted in large differences in the final reconstructions (Gaillard et al., 2010; Kaplan et al., 2017).

Lines 125-132: Given that these MIPs are already well under way, could you pls comment how realistic it is that the communities will be able to take up these protocols in time? Is it not more likely that the work will be most useful for many other studies that may not follow the tight schedule of the current AR6 MIP-frenzy, including work that would be useful also in context of the IPBES; and/or might feed into the next IPCC cycle ?

Although the deadline for inclusion of material in the next IPCC Assessment report is looming, analyses of the CMIP6 simulations are not entirely tied to the current cycle and will continue after this year. Furthermore, the focus for most groups to date has been on Tier 1 type simulations and additional simulations will be made in the next years. This is certainly the case for the Palaeoclimate Modelling Intercomparison Project where, although the baseline mid-Holocene simulations are mostly completed, sensitivity simulations such as those we propose here will mostly not be started until 2020. However, we agree that we need to make it clear that the intention here is to provide a protocol for new model simulations. In order to do this, and in response to comments by RC1, we have modified this paragraph as follows:
The Past Global Change (PAGES, http://www.pastglobalchanges.org/**) LandCover6k Working Group (http://pastglobalchanges.org/ini/wg/landcover6k) is currently working to develop a rigorous and robust approach to provide data and data products that can be used to inform reconstructions of LULC (Gaillard et al., 2018). LULC changes are taken into account in simulations currently being made in the current phase of the Coupled Model Intercomparison Project (CMIP6) for the historic period and the future scenario runs (Eyring et al., 2016). They are also included in simulations of the past millennium (Jungclaus et al., 2017), in order to ensure that these runs mesh seamlessly with the historic simulations. However, the Land Use Harmonisation data set (LUH2: Hurtt et al., 2017) only extend back to 850 CE and thus LULC changes are currently not included in the CMIP6 palaeoclimate simulations, including mid-Holocene simulations, that are used as a test of how well state-of-the-art climate models reproduce large climate changes. In this paper, we discuss how archaeological data will be used to improve global LULC reconstructions for the Holocene. Given that there are large uncertainties associated with the primary data and further uncertainties may be introduced when this information is used to modify existing LULC scenarios, we outline a series of tests that will be used to evaluate whether the revised scenarios are consistent with the changes implied by independent pollen-based reconstructions of land cover and whether they produce more realistic estimates of both carbon cycle and climate change. Finally, we present a protocol for implementing LULC in Earth System Model simulations to be carried out in the current phase of the Palaeoclimate Modelling Intercomparison Project (PMIP: Otto-Bleisner et al., 2017; Kageyama et al., 2018). However, the data sets and protocol will also be useful in later phases of other CMIP projects, including the Land Use Model Intercomparison Project (LUMIP) and the Land Surface, Snow and Soil Moisture Model Intercomparison Project (LS3MIP) (Lawrence et al., 2016; van den Hurk et al., 2016).**

Lines 141/142: style; one 'required/requirements' might be sufficient. . .?
We will change this to read:
Generalising from site-specific data to landscape or regional scales involves making assumptions about human behavior and cultural practices.

Figure 4: 'Wetland cultivation' in Level 3 – would that mean wetland drainage for agriculture? I assume it does, please clarify.
The three categories under wet cultivations are: 1) creation of artificial wetlands for wetland crops, e.g., rice paddies, taro, 2) draining of wetlands in preparation for upland crops and pasture, e.g. polders or raised field systems, and 3) cultivation of existing wetlands (wetland cultivation). Thus, this last category does not mean drainage but rather preservation and use of existing wetlands. The terminology is explained in the cited Morrison et al. reference. However, we will modify the paragraph (lines 217-227) to make this clearer (and also to deal with comments made by Erik Kjellstrom), 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, using 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 assessments depending on the quality and quantity of the archaeological information available from different regions.

Lines 146-162 – bit of an unspecific list, can be more precise, give more concrete examples?
The aim of this text was to provide a general overview of the LandCover6k approach, to put subsequent sections into a broader context. each of the things listed are an explicit part of our strategy and thus further described. Since this obviously was not clear, we will revise the text and make explicit reference to the Figure describing the LandCover 6 scheme (Figure 2) and to the sections of the paper in which we develop each idea, as follows:
Because of the inherent uncertainties, we advocate an iterative approach to incorporate archaeological data into LULC scenarios in LandCover6k (Fig. 2). We propose to revise the LULC scenario by incorporation of diverse archaeological inputs (Fig. 2, phase 1; see Sections 3 and 4) and to test the revised LULC scenarios for their plausibility and consistency with other lines of evidence (Fig. 2, phase 2 with iterative testing; see Sections 5-7). As a first test, the revised LULC scenarios of the extent of cropland and grazing land through time will be compared with independent data on land-cover changes, specifically pollen-based reconstructions of the extent of open land (see e.g. Trondman et al., 2015; Kaplan et al., 2017) (Section 5). Further testing the LULC scenarios involve sensitivity tests using global climate models (Section 6) and global vegetation-carbon cycle models (Section 7). While the computational cost of the climate simulations can be minimized using equilibrium time-slice simulations, the carbon cycle constraint relies on transient simulations, but may be derived from uncoupled, land-only simulations. Simulated climates at key times can be evaluated against reconstructions of climate variables (e.g. Bartlein et al., 2011) (Section 6). The parallel evolution of CO2 and its isotopic composition (δ13C) can be used to derive the carbon balance of the terrestrial biosphere and the ocean separately (Elsig et al., 2009) and, in combination with estimates for other contributors to land carbon changes such as C sequestration by peat buildup, provides a strong constraint on the evolution of LULC through time. An under- or over-prediction of anthropogenic LULC-related CO2 emissions during a specific interval results in consequences for the dynamics of the atmospheric greenhouse gas burden in subsequent times (Stocker et al., 2017) (Section 7). Thus, these tests can be used to identify issues in the original archaeological datasets and/or the way these data were incorporated into the LULC scenarios that require further refinement. Phase 3 of the protocol (Fig. 2)

proposes specific implementation of the revised LULC in Earth System Model simulations (Section 8).

We will also modify the caption to Figure 2 as follows:
Figure 2: Proposed scheme for developing robust LULC scenarios through iterative testing and refinement, as input to Earth System Model (ESM) simulations. The archaeological inputs developed in Phase 1 can be used independently or together to improve the LULC reconstructions (Phase 2); iterative testing of the LULC scenario reconstruction (phase 2) will ensure that these inputs are reliable before they are used of ESM simulations (phase 3). The uppermost three LULC simulations capitalize on already planned baseline simulations without LULC; the lowermost two simulations are envisaged as new sensitivity experiments.

Section 3.1 – this section wasn't entirely clear to me. What samples are we talking about exactly, what is being dated, where do the samples come from? Could you provide an illustrative example?
We will modify this section to make it clearer that we are referring to dated archaeological material, as follows:
Radiocarbon is the most routinely used absolute dating technique in archaeology, especially for the Holocene. Many thousands of radiocarbon dates on archaeological material are available from the literature. A number of regional and pan-regional initiatives are compiling these records through exhaustive survey of the archaeological literature (e.g. the Canadian Archaeological Radiocarbon Database: https://www.canadianarchaeology.ca/).

We will also modify the text describing the sources of bias:
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.

Since there are several different ways this approach is being applied, we do not feel a single illustrative example would be adequate. We will therefore modify the final sentence of this paragraph to indicate that the references given refer to specific regional examples, as follows:
Radiocarbon dates have been successfully used in several regions to identify population fluctuations associated with the introduction of farming and subsequent changes in farming regimes (western Europe: Shennan et al., 2013; Wyoming: Zahid et al., 2016; South Korea: Oh et al., 2017; see also Freeman et al., 2018) as well as climatic oscillations (Ireland: Whitehouse et al., 2014; Japan: Crema et al., 2016).

Figure 5: I liked the Figure, is nice to see a concrete, illustrative example of the planned approach. However, it was not entirely obvious to me what the top and bottom panels in Fig. 5 are meant to convey: is it to show the improvements that can be made by adding the new information to the existing LandCover 6a? Or what is exactly the added value of the two combined? And what's the reasoning behind the 10-15% and the 5% mentioned in lines 269/270?
This figure illustrates alternative approaches to mapping land use, with the upper panels showing the distribution of archaeological sites and how these data are generalised to an provide an estimate of the extent of land use. The lower panels show the same data but

superimposed on the land use classification scheme used by LandCover6k. It is unrealistic for these periods - or even today - to consider that the entire 64km2 is continuously covered with fields, and the percentages given are estimates of how much of each grid cell was being used in cells assigned to low-level agriculture in different parts of Ireland. We will modify the caption to make this clearer, as follows:

An example of regional land-use mapping. The upper panels show the distribution of known archaeological sites superimposed on kernel density estimates of the extent of land-use based on the density of observations, and the lower panels show these data superimposed on the LandCover6k land-use classes for the Middle Neolithic (3600-3400 cal BC, 5600-5400 BP) (left panels) and the Early Neolithic (3750-3600 cal BC, 5750-5600 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). The assigned land-use classes are inferred from archaeological material from one (or more) sites within the grid box. It should not be assumed that the whole gridcell was being used for agriculture during the Middle and Early Neolithic. Informed assessment suggests that agricultural land (crop growing and grazing, combined) probably occupied between 10-15% of the total grid area in the low-level food production regions of the eastern and western coastal areas, whilst agricultural land likely represents 5% or less of the total grid cell area in inland areas.

Lines 288/289: how do you obtain information about past irrigation? From archaeological data (irrigation structures?) I assume? Likewise, per-capita land needs surely change over time, agreed. But how can these estimates be obtained, could you provide more explanation and/or references to methods as to how to do this?

We will expand the text to clarify these points, as follows:

Information on the extent of rain-fed versus irrigated agriculture, as indicated by the presence of irrigation structures associated with archaeological sites, can also be used to refine the distribution of these classes in the LULC scenarios. Per-capita land-use estimates and their changes through time (see e.g. Hughes et al., 2018; Weiberg et al., 2019) provide a further refinement of the LULC scenarios, allowing a better characterization of the distinction between e.g. areas given over to extensive versus intensive animal production (rangeland versus pasture in the HYDE 3.2 terminology).

Additional references

Weiberg, E., Hughes, R. E., Finné, M., Bonnier, A., & Kaplan, J. O. (2019). Mediterranean land use systems from prehistory to antiquity: a case study from Peloponnese (Greece). *Journal of Land Use Science*, 1-20. doi:10.1080/1747423x.2019.1639836

Hughes, R., Weiberg, E., Bonnier, A., Finné, M., & Kaplan, J. (2018). Quantifying land use in past societies from cultural practice and archaeological data. *Land, 7*(1), 9. doi:10.3390/land7010009

Figure 6, just for illustrative purpose only: the panels 'land use classification input' and 'revised land use allocation' look identical, might be illustrative to not only change the legend but also the drawing.

These two panels necessarily look identical because the archaeological data shown in the lefthand panel are explicitly incorporated into the scenario. Unlike in the other examples, it is difficult to show the before/after situation here. However, we can expand the caption to make this clearer, as follows:

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.

Line 327-329: what's the basis for the optimism that 'eventually' these pollen-based reconstructions will also be available elsewhere (presumably: the tropics), is there initial work that points in that direction? And what's the pros/cons of the "other" pollen-based reconstructions that are mentioned?

There is indeed work going on the collect RPP data in other parts of the world, and we will expand the text to explain this and to explain the pros/cons of the other techniques, as follows:

The REVEALS approach has been used to reconstruct changes in the amount of open land through time across the northern extratropics (Figure 7; Dawson et al., 2018) through the Holocene with a time resolution of 500 years from 11.5ka to 0.7ka BP, and three historical time windows (modern–0.1ka BP, 0.1–0.35ka BP, and 0.35–0.7ka BP). A major limitation in applying REVEALS globally is requirement for information about the relative pollen productivity (RPP) of individual pollen taxa, which is currently largely lacking for the tropics. However, LandCover6k has been collecting RPPs for China, South-East India, Cameroon, Brazil and Argentina and pollen-based land-cover reconstructions will be available for at sufficient parts of the tropics to allow testing of the scenarios. Another limitation of REVEALS reconstructions is that RPP estimates are available for cultivated cereals but not for other cultivars or cropland weeds, so the LandCover6k reconstructions will generally underestimate cropland cover (Trondman et al., 2015). It may also be possible to use alternative pollen-based reconstructions of land cover changes, such as the Modern Analogue Approach (MAT: e.g. Tarasov et al., 2007; Zanon et al. 2018); pseudo-biomization (e.g. Fyfe et al., 2014) or STEPPS (Dawson et al., 2016). While none of these methods require RPPs, MAT and STEPPS can only be applied in regions where the pollen datasets have dense coverage (such as Europe and North America) and pseudo-biomization is affected by the non-linearity of the pollen-vegetation relationship that the REVEALS approach is designed to remove.

Lines 385/386: "known" today is not quite true unfortunately. There are still sizeable discrepancies in today's land cover estimates in terms of major classes such as crop- land, pasture, forest, 'other' (let alone in the degree to which these are being used). Partially this arises from disagreements in terms of how a pasture or forest is defined. There is no need to add a long discussion but pls. revise the sentence slightly to express that there is also uncertainty for today.

We agree that this statement was a little too optimistic and will change the text to read:

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.

Lines 383-399: The scaling aspect is important. However, cumulative LUC C emissions differ substantially depending on whether "net" or "gross" area changes are being calculated. The

total agricultural area might be the same in both approaches, but the 'gross' approach considers expansion and reduction that might occur within a gridcell. The most prominent example is shifting cultivation, and today is mostly restricted to tropical regions. However, others have pointed out that such gross transition of course also are relevant on other parts of the world (see e.g., Fuchs et al., GCB, 2015), and were possibly even more so further back in time. The challenges that arise from this aspect are mentioned later in the Outcomes section but I wonder if it's not better to introduce these already here.

We agree that it would be important to account for the difference, and this is one reason that we discuss this issue in the Outcomes section (lines 532 et seq.). Unfortunately, the only way to do this globally at the present time is by making assumptions about farming practices (e.g. how much land is abandoned or fallowed in a given year). The archaeological record does not provide a very strong basis for quantifying this. We will modify the text describe the carbon-cycle simulations to clarify that these simulations will necessitate making assumptions about the nature of land-use turnover, 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. This will necessitate making informed decisions about the fraction of land under cultivation that is abandoned or left fallow each year, and the maximum extent of land affected by such episodic cultivation. The simulations will be driven by climate outputs (temperature, precipitation and cloud cover) from an existing 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.

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. This will necessitate making informed decisions about the fraction of land under cultivation that is abandoned or left fallow each year, and the maximum extent of land affected by such episodic cultivation. 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 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 1850 CE 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 reference to the figure at this point, as follows:
The REVEALS approach has been used to reconstruct changes in the amount of open land through time across the northern extratropics (Figure 7; Dawson et al., 2018)

through the Holocene with a time resolution of 500 years from 11.5ka to 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). The assigned land-use classes are inferred from archaeological material from one (or more) sites within the grid box. It should not be assumed that the whole gridcell was being used for agriculture during the Middle and Early Neolithic. Informed assessment suggests that agricultural land (crop growing and grazing, combined) probably occupied between 10-15% of the total grid area in the low-level food production regions of the eastern and western coastal areas, whilst agricultural land likely represents 5% or less of the total grid cell area in inland areas.

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.

Response to Reviewer RC1

The referee comments that it is difficult to judge the manuscript because it is a proposal for work to be done, although they recognise that the approach outlined is novel in several regards and would make a substantial contribution towards providing higher fidelity estimates of past LULC. We recognise that the paper is a somewhat unusual protocol in that it combines the development and testing of input data sets for model simulations as well as the description of the proposed simulations themselves. We have chosen to do this because we feel it is important that the palaeoclimate modellers who will be running these simulations understand the strengths and weaknesses of the input data sets that are being developed. However, the ultimate

goal here is to provide the protocol for simulations to be run by the PMIP group, building on the Holocene simulations that are already underway as part of CMIP6. The creation of the archaeological data sets and their use to improve LULC scenarios is currently being carried out by the PAGES LandCover6k working group, and it is anticipated that these data sets will be available for the PMIP community to use in 2020 -- hence the need for a protocol to describe the planned experiments.

We can perhaps make the situation clearer by revising the introductory text to make it clear that work on the production of the input data sets is ongoing. (We will also be clarifying the status of individual components of the work in response to comments by Arneth and Kjellstrom). Specifically, we propose revising the final paragraph of the introduction to read:

The Past Global Changes (PAGES, http://www.pastglobalchanges.org/) LandCover6k Working Group (http://pastglobalchanges.org/landcover6k) is currently working to develop a rigorous and robust approach to provide data and data products that can be used to inform reconstructions of LULC (Gaillard et al., 2018). LULC changes are taken into account in simulations currently being made in the current phase of the Coupled Model Intercomparison Project (CMIP6) for the historic period and the future scenario runs (Eyring et al., 2016). They are also included in simulations of the past millennium (Jungclaus et al., 2017), in order to ensure that these runs mesh seamlessly with the historic simulations. However, the Land Use Harmonisation data set (LUH2: Hurtt et al., 2017) only extend back to 850 CE and thus LULC changes are currently not included in the CMIP6 palaeoclimate simulations, including mid-Holocene simulations, that are used as a test of how well state-of-the-art climate models reproduce large climate changes. In this paper, we discuss how archaeological data will be used to improve global LULC reconstructions for the Holocene. Given that there are large uncertainties associated with the primary data and further uncertainties may be introduced when this information is used to modify existing LULC scenarios, we outline a series of tests that will be used to evaluate whether the revised scenarios are consistent with the changes implied by independent pollen-based reconstructions of land cover and whether they produce more realistic estimates of both carbon cycle and climate change. Finally, we present a protocol for implementing LULC in Earth System Model simulations to be carried out in the current phase of the Palaeoclimate Modelling Intercomparison Project (PMIP: Otto-Bleisner et al., 2017; Kageyama et al., 2018). However, the data sets and protocol will also be useful in later phases of other CMIP projects, including the Land Use Model Intercomparison Project (LUMIP) and the Land Surface, Snow and Soil Moisture Model Intercomparison Project (LS3MIP) (Lawrence et al., 2016; van den Hurk et al., 2016).

Response to reviewer RC2

Comments in *italics*, response in normal script, suggested changes to text in **bold.** We note that several of these comments are similar to those posted by Erik Kjellström, and in these cases we have already responded and note this here.

How are these LULC reconstructions better/different than HYDE and KK10? Are the methods different? Do we know that it is better? This may be obvious for everyone in the LULC business, but it is not explicitly explained in the text, at least not as far as I can see.

The LULC reconstructions we are proposing will be refinements of HYDE and KK10 that take account of a wider range of archaeological data. We describe these data in Section 3 and how they will improve the current HYDE and KK10 scenarios in section 4. In response to comments by the other reviewers, we have expanded the text in both of these sections to be more explicit about the data and how these data will be incorporated into the existing scenarios. The main improvements hinge on having better estimates of population changes based on the density of archaeological settlement evidence, better information for the initiation of agriculture in a region, more regionally specific information about land use, and more nuanced information about land-use per capita than the somewhat generic estimates used in KK10 or the single global assumption about land-use per capita that underpins HYDE. Until these data are used to revise the scenarios, and tested against independent data (as described in Sections 5, 6 and 7), we cannot be sure what impact they will have. Our contention is that it is surely better to incorporate information about human exploitation of the landscape than to rely on estimates that we know are based on relatively simple assumptions and which, in any case, differ markedly from one another as a consequence of these assumptions. We will take the opportunity to make a clearer statement about this in our final outcomes and perspective section, as follows:

LandCover6k has developed a protocol for using archaeological information to improve existing scenarios of LULC changes during the Holocene, specifically by using archaeological data to provide better estimates of regional population changes through time, better information on the date of initiation of agriculture in a region, more regionally specific information about the type of land use, and more nuanced information about land-use per capita than currently implemented in the LULC scenarios generated by HYDE and KK10. While the final global archaeological data sets are still in production, fast-track priority products have been created and their impact on current LULC scenarios is being tested.

Is it possible to do uncertainty ranges? Some regions will inevitably be more uncertain than others. When you do a global map you tend to think that the uncertainties are the same everywhere. How do you deal with that? Also, the paper kind of assumes that data availability is as good as for the northern hemisphere in all of the world. I guess a lot of your methods won't work that well in parts of the world. How do you deal with that?

We are fully aware that the amount and quality of the archaeological data inputs is not the same everywhere, and indeed we state this in our outcomes and perspective section (line 512 et seq.). Nevertheless, incorporating information from regions where the data is good and identifying regiona where there is less certainty will certainly go some way to improving the scenarios. It should be remembered that the archaeological itself is only input to the scenarios and that both HYDE and KK10 interpolate these data to generate global scenarios of land use. It is certainly possible and our intention to provide uncertainty ranges on the estimates (see for e.g. the caption to Figure 5). These can be used to generate for example high-end and low-end scenarios of LULC change, a practice that parallels the implementation of LULC changes in future simulations. We did not spell this out clearly in the paper, and so we will take the opportunity to do so, as follows:

Although the work of LandCover6k will provide more solid knowledge about anthropogenic modification of the landscape, some information will inevitably be missing and some key regions will be poorly covered. There will still be large uncertainties associated with LULC scenarios. Documenting these uncertainties is an important goal of the LandCover6k project, and will allow the generation of multiple scenarios comparable to the "low-end", "high-end" scenarios used for e.g. in future projections. Furthermore, we have proposed a series of tests that will help to evaluate the realism of

the final scenarios, based on independent evidence from pollen-based reconstructions of land cover, reconstructions of climate, and carbon-cycle constraints. These tests should help in identifying which of the potential LULC reconstructions are most realistic and constraining the sources of uncertainty.

We will also be using the REVEALS land cover information to evaluate the archaeology-based maps, and we can also address uncertainties in these reconstructions. The uncertainty of a pollen-based REVEALS estimate of cover for a plant taxon or a group of plant taxa is partly expressed by its standard error (SE). This SE takes into account the SE on the relative pollen productivity (RPP) of each plant taxon included in the REVEALS reconstruction, and the variability between the site-specific REVEALS estimates (e.g. Trondman et al., 2015). This allows us to take account of the uncertainty on the pollen-based land cover when compared with the LULC scenarios. We will modify the text (line 320) as follows:

The more pollen records per grid cell and pollen counts per time window, the smaller the estimated error on the land-cover reconstruction. The uncertainties on the pollen-based REVEALS estimates are partly expressed by their standard errors (SEs). These SEs take into account the SE on the relative pollen productivity (RPP) of each plant taxon included in the REVEALS reconstruction and the variability between the site-specific REVEALS estimates (e.g. Trondman et al., 2015). These uncertainties on the pollen-based land cover are considered when these reconstructions are compared with LULC scenarios (Kaplan et al., 2017).

I think Section 2 is a bit confusing to follow. What is it that you want to show? Is it only to give a hint of the outline of the paper? That could be done much simpler. Section 1 introduces about the same concepts in a nice way, and the rest of the paper gives the details. It's hard to know if this is a description of the paper or something more general about the LandCover6k methodology (if these two are the same, please say so). I think that the rest of the paper will be easier to read if Section 2 clearly lists the three main points: 1) ways to improve data 2) ways to test data 3) the protocol. If this structure is kept and clear for the rest of the paper it will be easier to follow. Because it's mixture of methods and results that is not always so easy to follow.

This Section was designed to explain the methodology we are using and in particular the different phases of work. within the protocol. In response to comments by Almut Arneth we propose to revise this section to make it clearer about the three different phases of work outlined in this protocol, i.e. (a). using archaeological data to refine LULC scenarios, (b) testing the revised scenarios and (c) running climate model simulations to examine the impact of LULC changes on climate, as follows:

Because of the inherent uncertainties, we advocate an iterative approach to incorporate archaeological data into LULC scenarios in LandCover6k (Fig. 2). We propose to revise the LULC scenario by incorporation of diverse archaeological inputs (Fig. 2, phase 1; see Sections 3 and 4) and to test the revised LULC scenarios for their plausibility and consistency with other lines of evidence (Fig. 2, phase 2 with iterative testing; see Sections 5-7). As a first test, the revised LULC scenarios of the extent of cropland and grazing land through time will be compared with independent data on land-cover changes, specifically pollen-based reconstructions of the extent of open land (see e.g. Trondman et al., 2015; Kaplan et al., 2017) (Section 5). Further testing the LULC scenarios involve sensitivity tests using global climate models (Section 6) and global vegetation-carbon cycle models (Section 7). While the computational cost of the climate simulations can be minimized using equilibrium time-slice simulations, the carbon cycle constraint relies on transient simulations, but may be derived from uncoupled, land-only simulations.

Simulated climates at key times can be evaluated against reconstructions of climate variables (e.g. Bartlein et al., 2011) (Section 6). The parallel evolution of CO2 and its isotopic composition (δ13C) can be used to derive the carbon balance of the terrestrial biosphere and the ocean separately (Elsig et al., 2009) and, in combination with estimates for other contributors to land carbon changes such as C sequestration by peat buildup, provides a strong constraint on the evolution of LULC through time. An under- or over-prediction of anthropogenic LULC-related CO2 emissions during a specific interval results in consequences for the dynamics of the atmospheric greenhouse gas burden in subsequent times (Stocker et al., 2017) (Section 7). Thus, these tests can be used to identify issues in the original archaeological datasets and/or the way these data were incorporated into the LULC scenarios that require further refinement. Phase 3 of the protocol (Fig. 2) proposes specific implementation of the revised LULC in Earth System Model simulations (Section 8).

In Section 5 I don't get if REVEALS is used as an input to the LULC reconstructions or if it is used to evaluate the reconstruction. Is it only the fraction of open land that is evaluated? How is land cover reconstructed without REVEALS as the archaeological data (as I understand it) only give fraction of open land/land use.

The REVEALS reconstructions are being used here as a way of evaluating the LULC reconstructions derived from archaeological information. REVEALS reconstructions could be used as input to the LULC scenarios, especially in regions where the archaeological information is sparse, but as we explain in the text (lines 333-339) there are problems in doing this because (a) pollen-based reconstructions cannot distinguish between anthropogenic and climatically determined natural open land (e.g. natural grasslands, steppes, wetlands) and (b) REVEALS underestimates cropland cover because there are no RPP estimates for cultivars other than cereals. In contrast, the archaeological data provides information on different types of agriculture (crops versus grazing versus mixed) and the types of crops being grown, direct information on the area affected and indirect estimates of the land-use per capita associated with different types of agriculture at different times that can be used to infer the area being used. However, since there is some confusion about the different information obtained from the two different sources and how we will use the REVEALS data for evaluation we will expand the text to explain this procedure more explicitly, as follows:

Pollen-based vegetation reconstructions can be used to corroborate archaeological information on the date of first agriculture from the appearance of cereals and agricultural weeds. These reconstructions can also be used to test the LULC reconstructions, either using relative changes in forest cover or reconstructions of the area occupied by different land cover types. LandCover6k uses the REVEALS model (Sugita, 2007) to estimate vegetation cover from fossil pollen assemblages. The REVEALS model predicts the relationship between pollen deposition in large lakes and the abundance of individual plant taxa in the surrounding vegetation at a large spatial scale (ca. 100 km x 100 km; Hellman et al., 2008a, b) using models of pollen dispersal and deposition. REVEALS can also be used with pollen records from multiple small lakes or peat bogs (Trondman et al., 2016) although this results in larger uncertainties in the estimated area occupied by individual taxa. The estimates obtained for individual taxa are summed to produce estimates of the area occupied by either plant functional (e.g. summer-green trees, evergreen trees) or land cover (e.g. open land, grazing land, cropland) types.

We will also add a final sentence to this section as follows:

However, overestimation of the area of open land in the LULC scenarios might suggest problems either in the archaeological inputs or their implementation, especially for times

or regions when other evidence indicates cereals were the major crop. In this sense, despite potential problems, the LandCover6k pollen-based reconstructions of land cover will provide an important independent test of the revised LULC scenarios.

For Section 6 I have a few concerns. First, should results be a part of a protocol paper? If it should, why are the results buried in the caption of Fig. 8? Are they old or new results? Make a proper paragraph explaining the results.

Section 6 is describing our approach for evaluating the new LULC scenarios by seeing whether they have an impact on simulated climate, and whether this impact is to produce a better a better simulation of climate or not. We illustrate this approach by showing two existing simulations, one with and one without LULC changes. The simulations are published and we cite this publication (Smith et al., 2016). It is not our intention here to comment on the simulations themselves, simply to illustrate how we would evaluate new simulations. We can clarify this by modifying the caption, as follows:

Quantitative comparison of the change in climate between the mid-Holocene (6ka) and the pre-industrial period as shown by pollen-based reconstructions (from Bartlein et al., 2011) and in simulations with and without the incorporation of land-use change (from Smith et al., 2016). This figure illustrates the approach that will be taken to evaluate the impact of new LULC scenarios on climate. The imposed land-use changes at 6ka 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). 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.

Second, the studies of LULC effects on simulated paleo climate that I'm familiar with tell clearly that despite radical changes in land cover the, although significant, differences in simulated climate are small compared to the uncertainty range in the proxies. It is not possible to assess which land-cover description is the most reasonable on the basis of a comparison of modelled climate with paleo climate reconstructions. (e.g. Strandberg et al., 2011; Strandberg et al., 2014). Your own results show this also. How do you plan to overcome this?

The Smith et al. simulations show regional changes in summer temperature (JJA) due to LULC of 2-3 degrees C in e.g. North America, Europe and China in the late Holocene, and changes of the same magnitude for more limited regions in the early Holocene. This is certainly within the detection range of the pollen-based reconstructions of summer temperature for these regions. Thus, we are sure that such comparisons will be a useful additional assessment of the new LULC simulations. In fact, in the Smith et al. simulations shown in Figure 8 to illustrate our approach, show an improvement in simulated climate in the high latitudes (increased warming) that is offset in this comparison by a degradation in simulated climate elsewhere. Thus, in our evaluations of the impact LULC on simulated climate we will necessarily have to make more detailed regional comparisons -- and this will be useful information for the diagnosis of the improved LULC simulations because it might pinpoint regions where the imposed LULC is wrong. We have already modified this paragraph in response to comments by Kjellström to clarify this point, 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.

Minor comments

L53: IPCC SRLUCC says 70% did you do a different kind of estimate? If you did, please explain why it's different.

To clarify, the estimate we provide is taken from the cited references. 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. In fact, 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).

L61: I don't think it's good to have the abbreviation LULC after the sentence "...as a result of land use". I guess LULC means land use and land cover. Spell out LULC before "affects the carbon cycle" on line 64 instead.

The sentence currently reads "changes in land cover as a result of land use (LULC)". We can expand this as follows:

.... changes in land cover as a result of land use (land use land cover: LULC)

L95: "differences in the underlying assumptions" It would be interesting to know about what these assumptions are.
We agree that we could be more explicit here and will change the sentence to read:
However, differences in the underlying assumptions about land-use per capita, which are generalized from limited and often site-specific data, have resulted in large differences in the final reconstructions (Gaillard et al., 2010; Kaplan et al., 2017).

L175. "LULC scenarios" Is "scenarios" the right word here? I would go for "reconstruction" as "scenario" for me means an assumption about the future, with emphasis on the word assumption. These "LULC scenarios" are not based on assumptions but "a number of products", i.e. they are in some way based on facts.
The term scenario is indeed used to describes a trajectory of change in the future based on making assumptions about e.g. behaviour patterns. It can equally well be used to apply to the past LULC changes which may be informed to some extent by data but are also underpinned by assumptions. Indeed, as we point out (see response above) it is these assumptions that give rise to the very large differences between the different "products" currently available. We do not claim that incorporating archaeological information will change the basis for scenario-creation; merely that incorporating more data that will help refine these assumptions, the resulting scenarios will become more realistic.

L229. "expert knowledge". How is "expert knowledge" done, is it even a method? Please explain and/or rephrase.
There are some regions where there are very few archaeological sites and where statistical methods are therefore difficult to apply. In such regions, we will be forced to use the insights of the archaeologists who worked on the sites about what kind of land use the archaeological records imply. We feel that this is more informative than leaving grid cells blank. We will change the sentence to read:
The LandCover6k land-use maps (see e.g. Fig. 5) will be based on different methods ranging from kernel-density estimates to expert assessments depending on the quality and quantity of the archaeological information available from different regions.

L281-295. Here, references to the different panels in Fig. 6 would be helpful.
We will modify the figure to add labels so that we refer to the separate panels in the text.

L328-329. How is this done globally, is it possible to do on a global scale?
It is not necessary to have global reconstructions to evaluate LULC scenarios, although this is of course desirable. The ultimate goal of PAGES LandCover6k is to provide such reconstructions globally, and we explain that lack of tropical RPPs is the current limitation on providing a global reconstruction using REVEALS. As we point out in our response to a comment by Almut Arneth about the likelihood of having global reconstructions, LandCover6k has been collecting tropical RPPs which will thus facilitate global reconstructions. Furthermore, as we point out in the paper, there are alternative methods that have been used in regions where there are no RPPs and these reconstructions can also be used to evaluate the LULC scenarios. We have expanded the text describing the pollen-based reconstructions (in response to Almut's comments), as follows:
The REVEALS approach has been used to reconstruct changes in the amount of open land through time across the northern extratropics (Figure 7; Dawson et al., 2018)

through the Holocene with a time resolution of 500 years from 11.5ka to 0.7ka BP, and three historical time windows (modern–0.1ka BP, 0.1–0.35ka BP, and 0.35–0.7ka BP). A major limitation in applying REVEALS globally is requirement for information about the relative pollen productivity (RPP) of individual pollen taxa, which is currently largely lacking for the tropics. However, LandCover6k has been collecting RPPs for China, South-East India, Cameroon, Brazil and Argentina and pollen-based land-cover reconstructions will be available for at sufficient parts of the tropics to allow testing of the scenarios. Another limitation of REVEALS estimates is that RPP estimates are available for cultivated cereals but not for other cultivars or cropland weeds, so the LandCover6k reconstructions will generally underestimate cropland cover (Trondman et al., 2015). It may also be possible to use alternative pollen-based reconstructions of land cover changes, such as the Modern Analogue Approach (MAT: e.g. Tarasov et al., 2007; Zanon et al. 2018); pseudo-biomization (e.g. Fyfe et al., 2014) or STEPPS (Dawson et al., 2016). While none of these methods require RPPs, MAT and STEPPS can only be applied in regions where the pollen datasets have dense coverage (such as Europe and North America) and pseudo-biomization is affected by the non-linearity of the pollen-vegetation relationship that the REVEALS approach is designed to remove.

L332. "transient" and "500 years". Is it correct to call something with 500 year resolution transient? Or should it rather be time slices. Compare the use of "transient" in Section 8.
It is true that in a modelling context we use the term transient to mean "every year" whereas the pollen-based reconstructions are currently the average plant cover over a 500-year time window, except in the last millennium (when we use shorter intervals). The time slices include one to several pollen counts, i.e. the REVEALS estimates of plant cover will represent the number of years one to several pollen samples represent within the 500 year-time interval. The more pollen samples per time intervals and the more pollen records per grid cells, the more years within the 500 yrs time slice will be represented in the reconstruction. This implies that the number of years represented in a reconstruction will very a lot through space and time. In palaeoecology we would consider such a reconstruction to be "transient". However, to avoid confusion, we will not use this term in the context of this paper. It would be possible to provide reconstructions at higher time resolution, for example to provide average cover over a 50-year window subject to the sampling resolution and the uncertainty of the age model of the individual pollen cores. Increasing the time resolution will imply that some grid cells will have no reconstructions for some time intervals. By using 500 yrs time slices we maximise the number of grid cells with reconstruction. We will modify the wording here to differentiate between the model simulations and the pollen-based reconstructions, as follows:
The REVEALS approach has already been used to produce gridded reconstructions of changes in the amount of open land through time across the northern extratropics. These reconstructions provide mean plant cover for time slices of 500 years through the Holocene until 0.7ka BP, and three historical time windows (modern–0.1ka BP, 0.1–0.35ka BP, and 0.35–0.7ka BP). The more pollen samples per time intervals and pollen records per grid cells, the more years within the 500 yrs time slice will be represented in the reconstruction. This implies that the number of years represented in a time-slice reconstruction varies in space and time.

L405. "contributions to the land C inventory can be specified..." Is this possible to achieve? Your assumption builds on that.
The main independent contribution to the land C inventory is the build up of peat through the Holocene and this is, at least to first order, known from syntheses of peat records. We can expand this text to be more specific, as follows:

Providing that all of the natural contributions to the land C inventory (e.g. the build up of natural peatlands: Loisel et al., 2014) can be specified from independent evidence, the anthropogenic sources can be estimated as the difference between the total terrestrial C budget and natural contributions (Figure 9) at any specific time.

Additional reference

Loisel J, Yu Z, Beilman DW, Camill P, Alm J, Amesbury MJ, Anderson D, Andersson S, Bochicchio C, Barber K, Belyea LR, Bunbury J, Chambers FM, Charman DJ, Vleeschouwer FD, Fiałkiewicz-Kozieł B, Finkelstein SA, Gałka M, Garneau M, Hammarlund D, Hinchcliffe W, Holmquist J, Hughes P, Jones MC, Klein ES, Kokfelt U, Korhola A, Kuhry P, Lamarre A, Lamentowicz M, Large D, Lavoie M, MacDonald G, Magnan G, Mäkilä M, Mallon G, Mathijssen P, Mauquoy D, McCarroll J, Moore TR, Nichols J, O'Reilly B, Oksanen P, Packalen M, Peteet D, Richard PJ, Robinson S, Ronkainen T, Rundgren M, Sannel ABK, Tarnocai C, Thom T, Tuittila E-S, Turetsky M, Väliranta M, Linden Mvd, Geel Bv, Bellen Sv, Vitt D, Zhao Y & Zhou W, 2014. A database and synthesis of northern peatland soil properties and Holocene carbon and nitrogen accumulation. *The Holocene* 24: 1028-1042

L542-545. This is not possible without first improving proxy data.

We do not understand this comment. The point of this protocol paper is to explain how we will improve the land use scenarios so that they can be used to drive model simulations. The point here is that these experiments could be used to explore whether the land-use changes are implicated in e.g. abrupt events or whether specific land-use changes associated with population changes used in the scenarios produce significant effects on climate.

Fig. 3 The text is far too small. No explanation for the grey shading or the white dots is given.

A similar point was raised by Kjellström and we have expanded the text and modified the caption to explain this figure better

Fig. 4 Two boxes in Level 1 don't connect to Level 2. I can see that "No human land use" doesn't have to connect to Level 2, but is it then necessary to include it in the figure? I don't see how "Extensive/Minimal land use" fits in the picture.

As we have said in our response to Kjellström, 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.

Fig. 5 Too small legends.

We will provide new figures to ensure that they are readable. Please see detailed explanations in the response to Kjellström.

Fig. 6 I don't understand the coupling between "LandCover 6k working group" and "HYDE 3.x". What does "→" mean? I don't understand many of the panels. What are the axes? What are the squares? What is the grey shading?

A similar point was raised by Kjellström and we have expanded the text and modified the caption to explain this figure better

Fig. 7 Far too small legends.

We will provide new figures to ensure that they are readable. Please see detailed explanations in the response to Kjellström.

Fig. 9 I don't understand this, but it seems to be more complicated than it sounds, but the surrounding text doesn't give much help.

The text here describes the basis for using carbon cycle constraints on LULC. We will modify the caption to the Figure to clarify what this illustrative figure shows and so that it can be better understood in relation to the surrounding text, as follows:

Illustration of the terrestrial C budget approach to evaluate LULC. The total terrestrial C balance (green circle 'total') is constrained by ice core records of CO_2 and its isotopic signature ($\delta^{13}C$). Estimates for C balance changes of different natural land carbon cycle components (e.g., peatlands, permafrost, forest expansion/retreat, desert greening) can are estimated independently (blue slices 'Natural components') either from empirical upscaling of site-scale observations or from model-based analyses (BGC models forced with varying climate).The remainder (yellow slice 'remainder') is then calculated as the total terrestrial C balance (green circle 'total') minus the sum of separate estimates of natural components (blue slices 'Natural components') The remainder is effectively the emissions resulting from LULC changes, and can therefore be compared to LULC CO_2 emission estimates by carbon-cycle models.

Table 1 What does "Modern" mean here? If it is pre-industrial say so. If it is modern (= 20th century) explain why you don't use pre-industrial.
This point has been raised by Kjellström and we have explained in that response that the PMIP protocol mandates modern geography and ice sheets for the pre-industrial simulation. We have expanded the text to explain this and modified the caption also.

Development and testing of scenarios for implementing land use and land cover changes during the Holocene in Earth System Model Experiments

Deleted: Holocene LULC

Sandy P. Harrison[1], Marie-José Gaillard[2], Benjamin D. Stocker[3,4], Marc Vander Linden[5], Kees Klein Goldewijk[6,7], Oliver Boles[8], Pascale Braconnot[9], Andria Dawson[10], Etienne Fluet-Chouinard[11], Jed O. Kaplan[12,13], Thomas Kastner[14], Francesco S.R. Pausata[15], Erick Robinson[16], Nicki J. Whitehouse[17], Marco Madella[18,19, 20], Kathleen D. Morrison[8]

Ms for Geoscientific Model Development (PMIP special issue)

1: Department of Geography and Environmental Science, University of Reading, Reading, UK
2: Department of Biology and Environmental Science, Linnaeus University, Kalmar, Sweden
3: Ecological and Forestry Applications Research Centre, Cerdanyola del Vallès, Spain
4: Department of Earth System Science, Stanford University, Stanford, CA 94305, USA
5: Department of Archaeology, University of Cambridge, UK
6: PBL Netherlands Environmental Assessment Agency, The Hague, The Netherlands
7: Copernicus Institute of Sustainable Development, Utrecht University, The Netherlands
8: University Museum of Archaeology & Anthropology, University of Pennsylvania, Philadelphia, USA
9: Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France
10: Department of General Education, Mount Royal University, Calgary, Canada
11: Department of Earth System Science, Stanford University, California, USA
12: Department of Earth Sciences, University of Hong Kong, Hong Kong
13: Institute of Geography, University of Augsburg, Augsburg, Germany
14: Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany
15: Centre ESCER, Department of Earth and Atmospheric Sciences, University of Quebec in Montreal, Montreal, Canada
16: Department of Anthropology, University of Wyoming, Laramie, Wyoming, USA
17: School of Geography, Earth and Environmental Science, University of Plymouth, Plymouth, UK
18: Department of Humanities (CaSEs), University Pompeu Fabra, Barcelona, Spain
19: ICREA Passeig Lluís Companys 23 08010 Barcelona, Spain
20: School of Geography, Archaeology and Environmental Studies, University of Witwatersrand, Johannesburg, South Africa

Abstract: Anthropogenic changes in land use and land cover (LULC) during the pre-industrial Holocene could have affected regional and global climate. Existing scenarios of LULC changes during the Holocene are based on relatively simple assumptions and highly uncertain estimates of population changes through time. Archaeological and palaeoenvironmental reconstructions have the potential to refine these assumptions and estimates. The Past Global Changes (PAGES) LandCover6k initiative is working towards improved reconstructions of LULC globally. 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.

Deleted: Current LULC scenarios

Deleted: 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, we propose methods to evaluate the revised scenarios by using independent pollen-based reconstructions of land cover and of climate. A further test involves carbon-cycle 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, we outline a protocol for using the improved LULC reconstructions in palaeoclimate simulations within the framework of the Palaeoclimate Modelling Intercomparison Project in order to quantify the magnitude of anthropogenic impacts on climate through time and ultimately to improve the realism of Holocene climate simulations.…

1 Introduction and Motivation

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). 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 (Mazoyer and Roudart, 2006; Zohary et al., 2012; Tauger, 2013; Maezumi et al. 2018), although there is controversy about the relative importance of climate changes and human impact on landscape development both during and since that time. Resolving the uncertainty about the extent and timing of land use is important because changes in land cover as a result of land use (land use land cover: LULC) have the potential to impact climate and the carbon cycle. Direct climate impacts occur through changes in the surface-energy budget resulting from modifications of surface albedo, evapotranspiration, and canopy structure (biophysical impacts, e.g. Pongratz et al., 2010; Myhre et al., 2013; Perugini et al., 2017). LULC affects the carbon cycle through modifications in vegetation and soil carbon storage (biogeochemical impacts, e.g. Pongratz et al., 2010; Mahowald et al., 2017) and turnover times, which changes the C sink/source capacity of the terrestrial biosphere. LULC changes have contributed substantially to the increase in atmospheric greenhouse gases during the industrial period (Le Quéré et al., 2018). It has been suggested that greenhouse gas emissions associated with Neolithic LULC changes were sufficiently large to offset climate cooling after the Mid-Holocene (the overdue-glaciation hypothesis: Ruddiman 2003). Although this has been challenged for several reasons, including inconsistency with the land carbon balance derived from ice-core and peat records (e.g. Joos et al., 2004; Kaplan et al., 2011; Singarayer et al., 2011; Mitchell et al., 2013; Stocker et al. 2017), a LULC impact on climate in more recent millennia appears more plausible.

Climate model simulations have shown that LULC changes have discernible impacts on climate, both in regions with large prescribed changes in LULC and in teleconnected regions with no major local human activity (Vavrus et al., 2008; Pongratz et al., 2010; He et al., 2014; Smith et al., 2016). At the global scale, the biogeophysical effects of the accumulated LULC change during the Holocene which resulted in reconstructed land cover patterns in 1850 CE 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). However, in these simulations, biophysical and biogeochemical effects were of comparable magnitude in the most intensively altered landscapes of Europe, Asia, and North America (He et al., 2014). Using parallel simulations, with and without LULC changes, Smith et al. (2016) showed that detectable temperature changes due to LULC could have occurred as early as 7000 years ago (7ka BP) in summer and throughout the year by 3ka BP. All of these conclusions, however, are obviously contingent on the imposed LULC forcing, which is highly uncertain.

There have been several attempts to map LULC changes through time (e.g. Ramankutty and Foley, 1999; Pongratz et al., 2008; Kaplan et al., 2011; Klein Goldewijk et al. 2011; Klein Goldewijk et al. 2017a, b). All of these reconstructions assume that anthropogenic land use is a function of population density and the suitability of land for crops and/or pasture. They then use estimates of regional population trends through time in combination with assumptions about per-capita land use and spatial land use allocation schemes to estimate anthropogenic changes in LULC across time and space. However, differences in the underlying assumptions about land-use per capita, which are generalized from limited and often site-specific data, have resulted in large differences in the final reconstructions (Gaillard et al., 2010; Kaplan et al., 2017). Hence, there are still very large uncertainties about the timing and magnitude of LULC changes, both at a global and at a regional scale (Figure 1).

Deleted: Today, nearly 40% of the ice-free land surface is under anthropogenic use, and large parts of the remaining land area is influenced by human activities (Foley et al., 2005; Ellis and Ramankutty, 2008; Ellis et al., 2010; Ellis et al., 2013).…

Deleted: with the shift from hunting and gathering to cultivation during the Mesolithic and Neolithic periods …

Deleted: climate feedbacks

Deleted: climate feedbacks

Deleted: At the global scale, the biogeophysical effects of LULC change during the Holocene 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).…

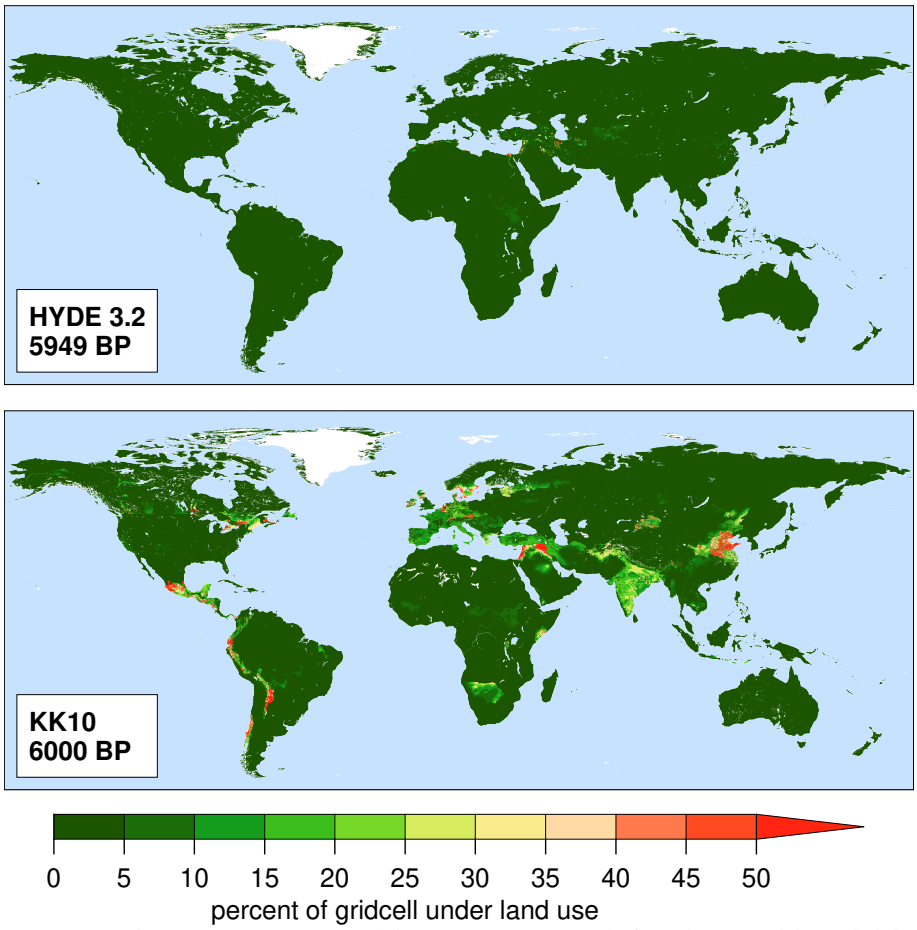

Figure 1: *Land use at ca 6000 years ago* (6ka BP, 4000 years BCE) *from the two widely used global historical land-use scenarios HYDE 3.2 (top panel, Klein Goldewijk et al. 2017a) and KK10 (bottom panel, Kaplan et al. 2011), illustrating the large disagreement between LULC scenarios at a regional scale.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

Deleted: *(6ka BP)*

There is a wealth of archaeological, historical and palaeo-vegetation data that could be used to improve the relatively simple rules used to generate global LULC reconstructions. For example, settlement density and numbers of radiocarbon-dated artifacts can be used to infer population sizes and their temporal dynamics (Rick, 1987; Williams, 2012; Silva and Vander Linden, 2017). Carbonised and waterlogged plant remains and animal bones can be used to infer the occurrence and nature of agriculture at a site, although their presence provides no quantitative information about the area under cultivation (Wright, 2003; Lyman 2008; Orton et al., 2016). Although the record of LULC is likely to be patchy and incomplete, because of preservation and sampling issues, systematic use of archaeological data is one important way to improve current LULC scenarios.

The Past Global Change (PAGES, http://www.pastglobalchanges.org/) LandCover6k Working Group (http://pastglobalchanges.org/ini/wg/landcover6k) is currently working to develop a rigorous and robust approach to provide data and data products that can be used to inform the development of LULC scenarios (Gaillard et al., 2018). LULC changes are taken into account in climate-model simulations currently being made in the current phase of the Coupled Model Intercomparison Project (CMIP6) for the historic period and the future scenario runs (Eyring et al., 2016). They are also included in climate-model simulations of the past millennium (Jungclaus et al., 2017), in order to ensure that these runs mesh seamlessly with the historic simulations. However, the Land Use Harmonisation data set (LUH2: Hurtt et al., 2017) only extends back to 850 CE and thus scenarios of LULC changes are currently not included in the CMIP6 palaeoclimate simulations, including mid-Holocene simulations, that are used as a test of how well state-of-the-art climate models reproduce large climate changes. In this paper, we discuss how archaeological data will be used to improve global LULC scenarios for the Holocene. Given that there are large uncertainties associated with the primary data and further uncertainties may be introduced when this information is used to modify existing LULC scenarios, we outline a series of tests that will be used to evaluate whether the revised LULC scenarios are consistent with the changes implied by independent pollen-based reconstructions of land cover and whether they produce more realistic estimates of both carbon cycle and climate changes. Finally, we present a protocol for implementing LULC in Earth System Model simulations to be carried out in the current phase of the Palaeoclimate Modelling Intercomparison Project (PMIP: Otto-Bleisner et al., 2017; Kageyama et al., 2018). However, the data sets and protocol will also be useful in later phases of other CMIP projects, including the Land Use Model Intercomparison Project (LUMIP) and the Land Surface, Snow and Soil Moisture Model Intercomparison Project (LS3MIP) (Lawrence et al., 2016; van den Hurk et al., 2016).

2 LandCover6k Methodology

The primary source of information about human exploitation of the landscape comes from archaeological data. In general, these data are site specific and spatiotemporal coverage is often patchy, and the types and quality of evidence available vary between sites and regions. Generalising from site-specific data to landscape or regional scales involves making assumptions about human behaviour and cultural practices. Because of the inherent uncertainties, we advocate an iterative approach to incorporate archaeological data into LULC scenarios in LandCover6k (Fig. 2). We propose to revise the existing LULC scenarios by incorporation of diverse archaeological inputs (Fig. 2, phase 1; see Sections 3 and 4) and to test the revised LULC scenarios for their plausibility and consistency with other lines of evidence (Fig. 2, phase 2 with iterative testing; see Sections 5-7). As a first test, the revised LULC scenarios of the extent of cropland and grazing land through time will be compared with independent data on land-cover changes, specifically pollen-based reconstructions of the extent of open land (see e.g. Trondman et al., 2015; Kaplan et al., 2017) (Section 5). Further testing the LULC scenarios involves sensitivity tests using global climate models (Section 6) and global carbon cycle models (Section 7). While the computational cost of the climate-model simulations can be minimized using equilibrium time-slice simulations, the carbon cycle constraint relies on transient simulations, but may be derived from uncoupled, land-only simulations. Simulated climates at key times can be evaluated against reconstructions of climate variables (e.g. Bartlein et al., 2011) (Section 6). The parallel evolution of CO_2 and its isotopic composition ($\delta^{13}C$) can be used to derive the carbon balance of the terrestrial biosphere and the ocean separately (Elsig et al., 2009) and, in combination with estimates for other contributors to land carbon changes such as C sequestration by peat buildup, provides a strong constraint on the evolution of LULC through time. An under- or over-prediction of anthropogenic LULC-related CO_2 emissions during a specific interval results in consequences for the dynamics of the atmospheric greenhouse gas burden in

Deleted: The goal of the Past Global Change (PAGES, http://www.pages-igbp.org/) LandCover6K working group (http://pastglobalchanges.org/ini/wg/landcover6k/intro) is to develop a rigorous and robust approach to provide data and data products that can be used to inform reconstructions of LULC (Gaillard et al., 2018). In this paper, we present a protocol to use archaeological data to improve global LULC reconstructions for the Holocene. Given the large uncertainties associated both with the primary data and in how these data are incorporated into LULC scenarios, we propose a series of tests to evaluate whether the revised scenarios are consistent with the changes implied by independent pollen-based reconstructions of land cover and whether they produce more realistic estimates of both carbon cycle and climate change. Finally, we present a protocol for implementing LULC in Earth System Model simulations to be carried out in the Coupled Model Intercomparison Project (CMIP: Eyring et al., 2016), specifically within the Palaeoclimate Modelling Intercomparison Project (PMIP: (Jungclaus et al., 2017; Otto-Bleisner et al., 2017; Kageyama et al., 2018). However, the data sets and protocol will also be useful for other CMIP projects, including the Land Use Model Intercomparison Project (LUMIP) and the Land Surface, Snow and Soil Moisture Model Intercomparison Project (LS3MIP) (Lawrence et al., 2016; van den Hurk et al., 2016).¶

Deleted: therefore requires

Deleted: Specifically (Figure 2), we propose testing the revised LULC scenarios to see whether the changes resulting from incorporating diverse archaeological inputs are plausible and consistent with other lines of evidence. As a first test, the revised LULC scenarios of the extent of cropland and grazing land through time will be compared for consistency with independent data on land-cover changes, for example pollen-based reconstructions of the extent of open land (see e.g. Trondman et al., 2015; Kaplan et al., 2017). Further testing the LULC scenarios would involve sensitivity tests using global climate models and global vegetation-carbon cycle models.…

subsequent times (Stocker et al., 2017) (Section 7). Thus, these tests can be used to identify issues in the original archaeological datasets and/or the way these data were incorporated into the LULC scenarios that require further refinement. Phase 3 of the project (Fig. 2) provides a protocol for the implementation of the revised LULC scenarios in Earth System Model simulations (Section 8).

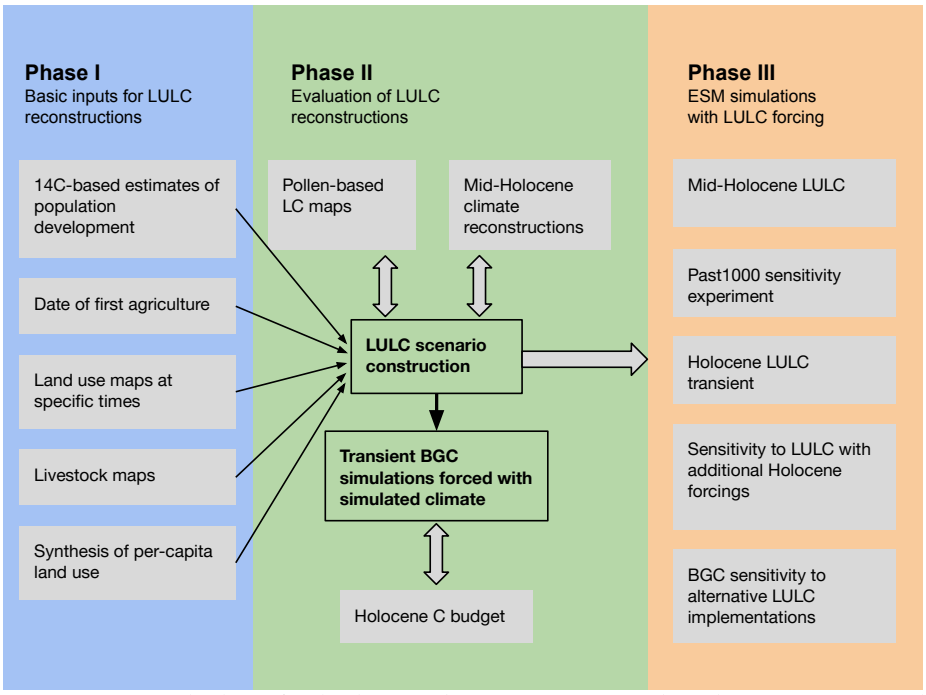

Figure 2: *Proposed scheme for developing robust LULC scenarios through iterative testing and refinement, as input to Earth System Model (ESM) simulations. The archaeological inputs developed in Phase 1 can be used independently or together to improve the LULC reconstructions; iterative testing of the LULC scenario reconstruction (Phase 2) will ensure that these inputs are reliable before they are used of ESM simulations (Phase 3). The uppermost three LULC simulations capitalize on already planned baseline simulations without LULC; the lowermost two simulations are envisaged as new sensitivity experiments.*

3 Archaeological data inputs

LandCover6k is creating a number of products that will be used to improve the LULC scenarios (Figure 2). Here, we summarise the important features of these data products before showing how they will be incorporated within a scenario-development framework.

3.1 Population dynamics from ^{14}C data

Radiocarbon is the most routinely used absolute dating technique in archaeology, especially for the Holocene. Many thousands of radiocarbon dates are available from the archaeological literature. A number of regional and pan-regional initiatives are compiling these records through exhaustive survey of the archaeological literature (e.g. the Canadian Archaeological Radiocarbon Database:

https://www.canadianarchaeology.ca/). Statistical approaches, such as summed probability distributions (SPDs), can then be used to infer past demographic fluctuations from these compilations (Figure 3). This method assumes that the more people there were, the more remains of their various activities they left behind, and that this is directly reflected in the number of samples excavated and dated (Rick, 1987: Robinson et al., 2019). 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, but these can be minimised through auditing the datasets. Assessment of the robustness of population reconstructions through time can be made statistically, by comparing a null hypothesis of demographic growth constructed from an exponential fit to the data with the actual record of number of dates through time (Shennan et al., 2013; Timpson et al., 2014). Mathematical simulations show that the method is relatively robust for large sample sizes (Williams, 2012). Radiocarbon dates have been successfully used in several regions to identify population fluctuations associated with hunter-gatherers (Japan: Crema et al., 2016) and the introduction of farming and subsequent changes in farming regimes (western Europe: Shennan et al., 2013; Wyoming: Zahid et al., 2016; South Korea: Oh et al., 2017; see also Freeman et al., 2018) as well as climatic oscillations (Ireland: Whitehouse et al., 2014).

Deleted: taphonomic

Deleted: losses with

Deleted: Radiocarbon dates have been successfully used in several regions to identify population fluctuations associated with the introduction of farming and subsequent changes in farming regimes (Shennan et al., 2013; Zahid et al., 2016; Oh et al., 2017; Freeman et al., 2018) as well as climatic oscillations (Whitehouse et al., 2014; Crema et al., 2016).…

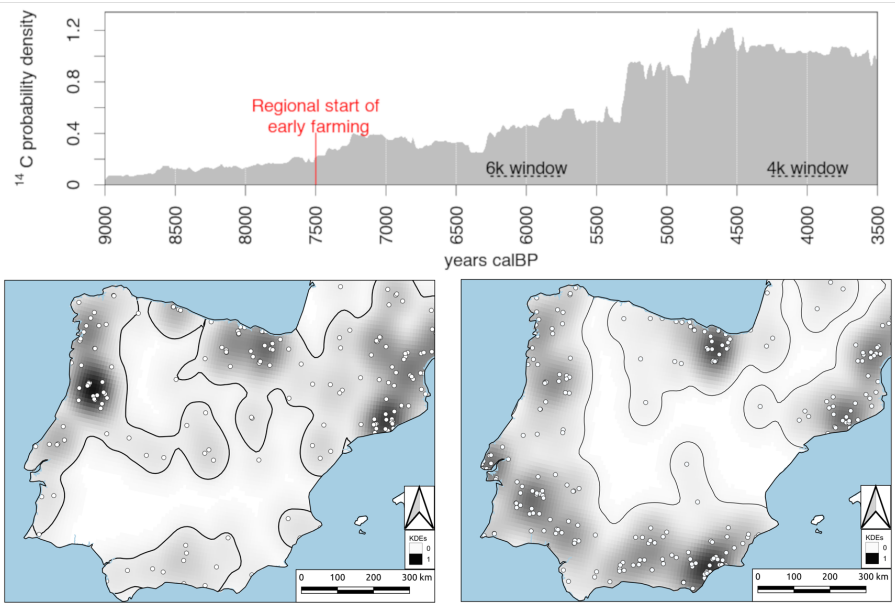

Figure 3: *Reconstruction of changes in population size in the Iberian Peninsula during the Holocene (9000 to 2000 BP, 9ka to 2ka BP) using summed probability distributions (SPDs) of radiocarbon dates (data after Balsera et al., 2015). The red line indicates the onset of agriculture in the region. The 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.*

3.2 Date of first agriculture

Radiocarbon dates can also be used to track the timing and process of dispersal events, such as the diffusion of plant and animal domesticates from their initial centres of domestication. Since the distribution of samples is often patchy, geostatistical techniques such as kriging and splines are used to spatially interpolate the information in order to provide quantitative estimates of the timing of spread. Work carried out in Europe (Bocquet-Appel et al., 2009), Asia (Silva et al., 2015), and Africa (Russell et al., 2014) demonstrates that there are different rates of diffusion even within a region, reflecting the possible impact of natural features (e.g. waterways, elevation, ecology) on diffusion rates (Davison et al., 2006; Silva and Steele, 2014). Numerous studies provide robust local estimates for the earliest regional occurrence of agriculture and these are being synthesized to provide a global product within LandCover6k (Figure 2).

3.3 Global land-use and livestock maps

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., 2019). LandCover6k is developing global land-use maps for specific time windows, using 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 assessments depending on the quality and quantity of the archaeological information available from different regions.

Deleted: *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 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. …*

Deleted: in press

Deleted: based on

Deleted: Further distinctions are subsequently made within each of these categories to encompass the diversity of land-use activities in each land-use type (Fig. 4). …

Deleted: knowledge

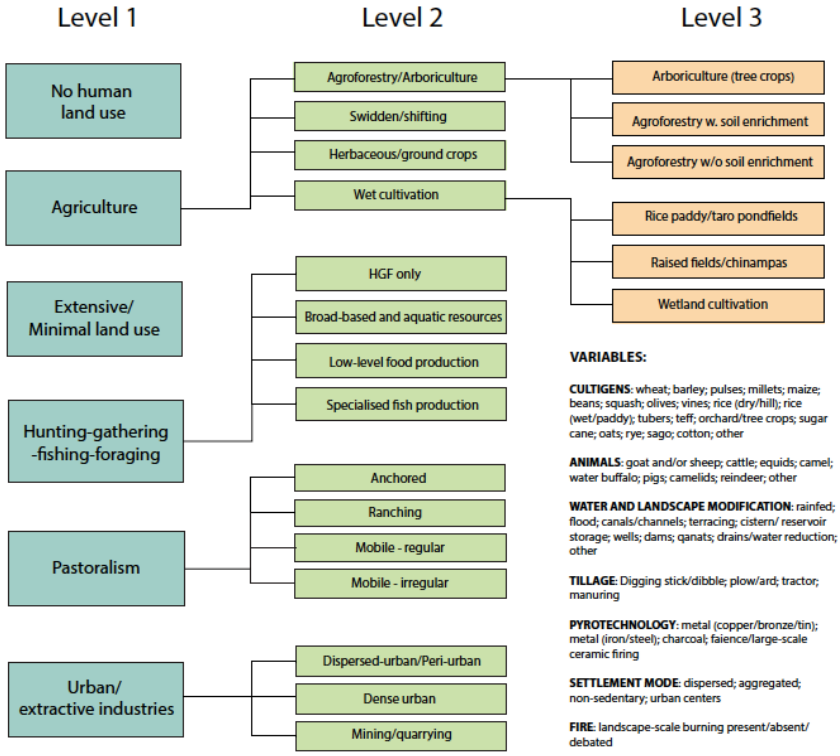

Figure 4: *The hierarchical scheme of land-use classes used for global mapping in LandCover6k (updated from Morrison et al, 2018).*

Deleted: *a*

There is considerable variation in how intensely land is used both for crops and for grazing within broad land-use categories both geographically and through time (Ford and Clarke, 2015; Styring et al., 2017). Maps of land-use types do not provide direct information on the intensity of farming practices or how they translate into per-capita land use. Archaeological data about agricultural yields, combined with information from analogous contemporary cultures, historical information (e.g. Pongratz et al., 2008) and theoretical estimates of land use required to meet dietary and energy requirements (e.g. Hughes et al., 2018), can be used to provide regional estimates of per-capita land use for specific land-use categories. LandCover6k will synthesise this information to allow regionally specific estimates of per-capita land use to be derived from the global land-use maps.

Information about the extent of grazing land is an important input for the development of revised LULC scenarios but, from a carbon-cycle modelling perspective, the amount of biomass removed by grazing is also a key parameter. Biomass loss varies not only with population size but also with the type of animal being reared (Herrero et al., 2013; Phelps & Kaplan, 2017) and thus information about what animals were present at a given location and estimates of population sizes are needed for improving the existing LULC scenarios. Although the conditions of bone preservation vary across the globe due to factors such as soil acidity, animal bones are routinely excavated (Lyman, 2008;

Deleted: to

Reitz & Wing, 2008). Morphometric analysis of bones, along with collateral information such as age-related culling patterns, make it possible to determine whether these are the remains of domesticated species. We thus have a relatively precise idea of when livestock were introduced into a region, what types of animal were being reared at a given time, and can also make informed estimates of population size. Although the level of detail will vary geographically, this information can be used to produce global livestock maps.

Deleted: and

The harvesting of wood for domestic fires, building, and for industrial activities such as transportation, pottery-making and metallurgy is an important aspect of human exploitation of the landscape in the pre-industrial period (McGrath et al., 2015). It has been argued that even Mesolithic hunter-gatherer communities shaped their environment through wood harvesting (Bishop et al., 2015). Approaches have been developed to quantifying the wood harvest associated with archaeological settlements at specific times based on the evidence of types of wood use, household energy requirements, population size, and calorific value of the wood used (see e.g. Marston, 2009; Janssen et al., 2017). However, quantitative information on ancient technology and lifestyle is sparse and direct estimates of the amount of wood harvest through time are likely to remain highly uncertain (Marston et al., 2017; Veal, 2017). Nevertheless, combining evidence-based inferential approaches with improved estimates of population size should allow changes in wood harvesting to be taken into account in constructing revised LULC scenarios.

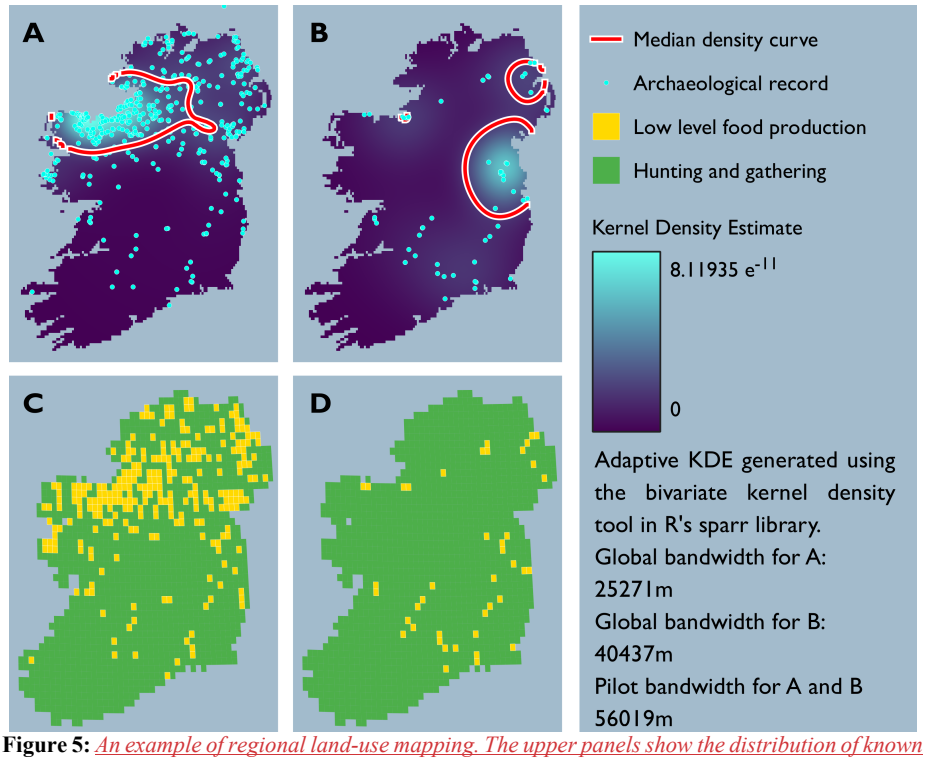

Figure 5: *An example of regional land-use mapping. The upper panels show the distribution of known archaeological sites superimposed on kernel density estimates of the extent of land-use based on the density of observations, and the lower panels show these data superimposed on the LandCover6k*

land-use classes 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). The assigned land-use classes are inferred from archaeological material from one (or more) sites within the grid box. It should not be assumed that the whole gridcell was being used for agriculture during the Middle and Early Neolithic. Informed assessment suggests that agricultural land (crop growing and grazing combined) probably occupied between 10-15% of the total grid area in the low-level food production regions of the eastern and western coastal areas, whilst agricultural land likely represents 5% or less of the total grid cell area in inland areas.

Deleted: *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 cal BC, 5600-5400 BP) (left panels) and the Early Neolithic (3750-3600 cal BC, 5750-5600 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….*

Deleted: -

4. Incorporation of archaeological data in LULC scenarios

The existing LULC scenarios are substantially dependent on historical regional population estimates at key times, which are then linearly interpolated to provide a year-by-year estimate of population. Estimates of regional population growth based on suitably screened ^{14}C data can be used to modify existing population growth curves (Figure 6), both in terms of establishing the initial date of human presence and by modifying a linear growth curve to allow for intervals of population growth and decline.

Information on the timing of the first appearance of agriculture at specific locations can be used to constrain the temporal record of LULC changes in the scenarios. This information can also be used to allocate LULC changes geographically across regions (Figure 6). Global land-use maps can be used to identify areas where there was no permanent agricultural activity at a given time (e.g. either unsettled areas or areas occupied by hunter-gatherer communities) and provide a further constraint on the geographic extent of the LULC changes given by the scenarios (Figure 6). The type of agriculture, including whether the region was predominantly used for tree or annual crops or for pasture, modifies the area of open land specified in the LULC scenarios. Information on the extent of rain-fed versus irrigated agriculture, as indicated by the presence of irrigation structures associated with archaeological sites, can also be used to refine the distribution of these classes in the LULC scenarios. Per-capita land-use estimates and their changes through time (see e.g. Hughes et al., 2018; Weiberg et al., 2019) provide a further refinement of the LULC scenarios, allowing a better characterization of the distinction between e.g. areas given over to extensive versus intensive animal production (rangeland versus pasture in the HYDE 3.2 terminology). There will remain areas of the world for which this kind of fine-grained information is not available. Nevertheless, by incorporating information where this exists, the LandCover6k products will contribute to a systematic refinement of existing LULC scenarios. Iterative testing of the revised scenarios will ensure that they are robust.

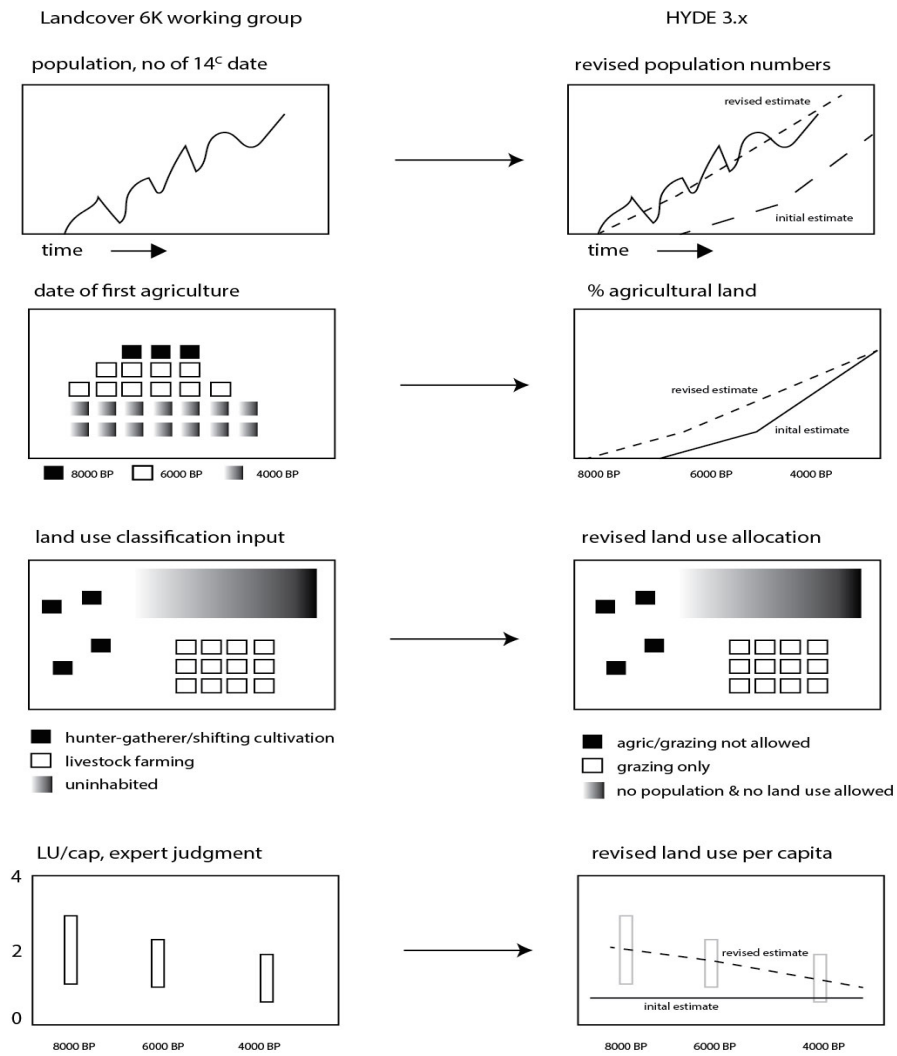

Figure 6: *Schematic illustration of the proposed implementation of ^{14}C-based population estimates, date of first agriculture, land-use maps, and land-use per capita information in the HYDE model (here indicated as HYDE3.x). The archaeological data are represented as values for a grid cell in geographic 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.*

Deleted: *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….*

5. Using pollen-based reconstructions of land cover changes to evaluate LULC scenarios

Pollen-based vegetation reconstructions can be used to corroborate archaeological information on the date of first agriculture from the appearance of cereals and agricultural weeds. These reconstructions can also be used to test the LULC reconstructions, either using relative changes in forest cover or reconstructions of the area occupied by different land cover types. LandCover6k uses the REVEALS pollen source-area model (Sugita, 2007) to estimate vegetation cover from fossil pollen assemblages. REVEALS predicts the relationship between pollen deposition in large lakes and the abundance of individual plant taxa in the surrounding vegetation at a large spatial scale (ca. 100 km x 100 km; Hellman et al., 2008a, b) using models of pollen dispersal and deposition. REVEALS can also be used with pollen records from multiple small lakes or peat bogs (Trondman et al., 2016) although this results in larger uncertainties in the estimated area occupied by individual taxa. The estimates obtained for individual taxa are summed to produce estimates of the area occupied by either plant functional (e.g. summer-green trees, evergreen trees) or land cover (e.g. open land, grazing land, cropland) types.

Deleted:

The geographic distribution of pollen records is uneven. There are also many areas of the world where environments that preserve pollen (i.e. lakes, bogs, forest hollows) are sparse. Site-based reconstructions of land cover are therefore interpolated statistically to produce spatially continuous reconstructions (Nielsen et al., 2012; Pirzamanbein et al., 2014; Pirzamanbein et al., 2018). LandCover6k uses a 1˚ resolution grid and all available pollen records in each grid cell to produce an estimate of land cover per grid cell through time. The more pollen records per grid cell and pollen counts per time window, the smaller the estimated error on the land-cover reconstruction. The uncertainties on the pollen-based REVEALS reconstructions are partly expressed by their standard errors (SEs). These SEs take into account the SE on the relative pollen productivity (RPP) of each plant taxon included in the REVEALS reconstructions and the variability between the site-specific REVEALS reconstructions (e.g. Trondman et al., 2015). The uncertainties on the pollen-based land cover reconstructions are taken into account when these reconstructions are compared with LULC scenarios (Kaplan et al., 2017).

Deleted: ¶

The REVEALS approach has already been used to produce gridded reconstructions of changes in the amount of open land through time across the northern extratropics (Figure 7; Dawson et al., 2018) These reconstructions provide mean plant cover for time slices of 500 years through the Holocene until 0.7ka BP, and three historical time windows (modern–0.1ka BP, 0.1–0.35ka BP, and 0.35–0.7ka BP). The more pollen samples per time interval and pollen records per grid cell, the more years within the 500 yrs time slice will be represented in the reconstruction. This implies that the number of years represented in a time-slice reconstruction varies in space and time.

A major limitation in applying REVEALS globally is the requirement for information about the relative pollen productivity (RPP) of individual pollen taxa, which is currently largely lacking for the tropics. However, LandCover6k has been collecting RPPs for China, South-East India, Cameroon, Brazil and Argentina and pollen-based land-cover reconstructions will be available for sufficient parts of the tropics to allow testing of the LULC scenarios. Another limitation of the REVEALS reconstructions is that RPP estimates are available for cultivated cereals but not for other cultivars or cropland weeds, so the LandCover6k pollen-based reconstructions will generally underestimate cropland cover (Trondman et al., 2015). It may also be possible to use alternative pollen-based reconstructions of land cover changes, such as the Modern Analogue Approach (MAT: e.g. Tarasov et al., 2007; Zanon et al. 2018); pseudo-biomization (e.g. Fyfe et al., 2014) or STEPPS (Dawson et

al., 2016). While none of these methods require RPPs, MAT and STEPPS can only be applied in regions where the pollen datasets have dense coverage (such as Europe and North America) and pseudo-biomization is affected by the non-linearity of the pollen-vegetation relationship that the REVEALS approach is designed to remove.

Comparison of the reconstructions of the extent of open land with the LULC deforestation scenarios will provide a first evaluation of the realism of the revised LULC scenarios (e.g. Kaplan et al., 2017). Underestimation or overestimation of open land in the LULC scenarios is not necessarily an indication that these scenarios are inaccurate because (a) pollen-based reconstructions cannot distinguish between anthropogenic and climatically determined natural open land (e.g. natural grasslands, steppes, wetlands) and (b) REVEALS underestimates cropland cover because there are no RPP estimates for cultivars other than cereals. However, overestimation of the area of open land in the LULC scenarios might suggest problems either in the archaeological inputs or their implementation, especially for times or regions when other evidence indicates cereals were the major crop. In this sense, despite potential problems, the LandCover6k pollen-based reconstructions of land cover will provide an important independent test of the revised LULC scenarios.

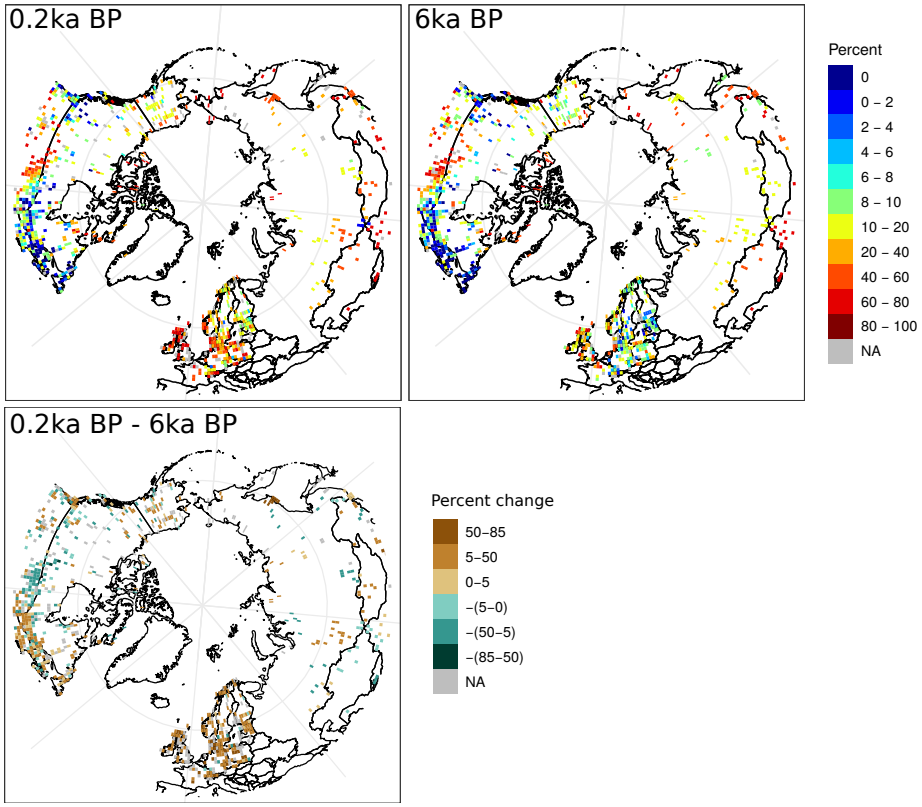

Figure 7: *Northern extratropical (>40°N) mean fractional cover of open land at 6000 years ago (6ka BP: top right panel) and 200 years ago (0.2ka BP: top left panel) estimated using REVEALS, and the*

Deleted: *at 6ka BP (left panel) and 0.2ka BP (centre panel)*

difference in fractional cover between the two periods (lower panel), where red indicates an increase in open land and blue a decrease (after Dawson et al., 2018).

6. Testing the reliability of improved scenarios using climate-model simulations

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, 6 ka 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, 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 reconstructed 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 whether improvements to the LULC scenarios produces more realistic simulations of climate. Such an evaluation would need to go beyond the global comparison made here (Figure 8) to regional comparisons to identify whether improvements in simulated climate in regions where there is a large anthropogenic impact on land cover do not result in a degradation in the simulated climate elsewhere.

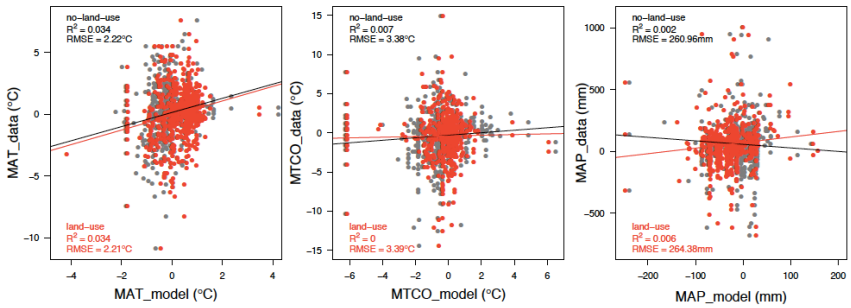

Figure 8: *Quantitative comparison of the change in climate between the mid-Holocene (6ka) and the pre-industrial period as shown by pollen-based reconstructions gridded to 2 x 2° resolution to be compatible with the model resolution (from Bartlein et al., 2011) and in simulations with and without the incorporation of land-use change (from Smith et al., 2016). This figure illustrates the approach that will be taken to evaluate the impact of new LULC scenarios on climate. 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.*

Deleted: *right*

Deleted: and

Deleted: observed

Deleted: and they therefore provide a good focus for testing improvements to the LULC scenarios….

Deleted: *Quantitative comparison of the change in climate between the mid-Holocene (6ka) and the pre-industrial period as shown by pollen-based reconstructions (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 6ka 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). Although the incorporation of land use produces somewhat warmer and wetter climates in these simulations, overall the incorporation of land-use produce no improvement of the simulated climates at sites with pollen-based reconstructions.*

Formatted: Font: Not Bold

Formatted: Font: Italic

7. Testing the reliability of improved scenarios using carbon-cycle models

Carbon-cycle modelling will be used as a further test of the realism of the improved LULC scenarios. Two constraints are available for testing the realism of past LULC scenarios. First, reconstructions of LULC history 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 change thus reduces to allocating a fixed total amount of land conversion from natural to agricultural use over time. More conversion in earlier periods implies less conversion in later periods. At the continental to global scale, cumulative LULC emissions scale linearly with the agricultural area. LULC scenarios that converge to the present-day state also converge to within a small range of cumulative historical emissions (Stocker et al., 2011; Stocker et al., 2017). Deviations from a linear relationship between extent and emissions are due to differences in biomass density in potential natural and agricultural vegetation of different regions affected by anthropogenic LULC. Differences in cumulative emissions for alternative LULC reconstructions with an identical present-day state are due to the long response time of soil carbon content following a change in carbon inputs and soil cultivation. Conserving the total extent of LULC (and allocating a fixed total expansion over time) is thus approximately equivalent to conserving cumulative historical LULC emissions. Thus, more LULC CO_2 emissions in earlier periods imply less CO_2 emissions in more recent periods.

The total C budget of the terrestrial biosphere provides a second constraint on LULC emissions through time. The net C balance of the land biosphere, which reflects the sum of all natural and anthropogenic effects on terrestrial C storage, can be reconstructed from ice-core data of past CO_2 concentrations and $\delta^{13}C$ composition (Elsig et al. 2009). Providing that all of the natural contributions to the land C inventory (e.g. the build-up of natural peatlands: Loisel et al., 2014) can be specified from independent evidence, the anthropogenic sources can be estimated as the difference between the total terrestrial C budget and natural contributions (Figure 9) at any specific time.

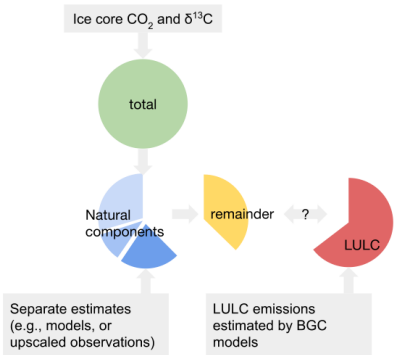

Figure 9: *Illustration of the terrestrial C budget approach to evaluate LULC. The total terrestrial C balance (green circle 'total') is constrained by ice core records of CO_2 and its isotopic signature ($\delta^{13}C$). Estimates for C balance changes of different natural land carbon cycle components (e.g., peatlands, permafrost, forest expansion/retreat, desert greening) can are estimated independently (blue slices 'Natural components') either from empirical upscaling of site-scale observations or from model-based analyses (BGC models forced with varying climate).The remainder (yellow slice 'remainder') is then calculated as the total terrestrial C balance (green circle 'total') minus the sum of the separate estimates of the natural components (blue slices 'Natural components') The remainder*

Deleted: reconstructions

Deleted: which is known from satellite land-cover observations and national statistics on the amount of land under use…

is effectively the emissions resulting from LULC changes, and can therefore be compared to LULC CO_2 emission estimates by carbon-cycle models.

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 scenarios, by comparing results obtained with prescribed LULC changes through time against a baseline simulation without imposed LULC. This will necessitate making informed decisions about the fraction of land under cultivation that is abandoned or left fallow each year, and the maximum extent of land affected by such episodic cultivation. We envisage using several different offline carbon-cycle models for this purpose in order to take account of uncertainties associated with differences between the carbon-cycle models. 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.

8. Implementation of LULC in Earth System Model simulations

We propose a series of simulations to examine the impact of LULC, using the revised LULC scenarios from LandCover6k and building on climate-model experiments that are currently being run either in CMIP6-PMIP4 (*midHolocene, past1000*) or within PMIP although not formally included as CMIP6-PMIP4 experiments.

The *mid-Holocene* (and its corresponding *piControl*) simulation 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 climate-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 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. These new mid-Holocene simulations would allow for a better understanding of the relationship between climate changes and land-surface feedbacks (including snow albedo feedbacks), and the role of water recycling at a regional scale. 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.

Table 1: 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.

Boundary conditions		1850CE (DECK *piControl*)	6ka (*midHolocene*)	6ka LULC (*midHoloceneLULC*)
Orbital parameters	Eccentricity	0.016764	0.018682	0.018682
	Obliquity	23.459	24.105	24.105

Deleted: *Illustration of the terrestrial C budget approach. The remainder (yellow slice 'remainder') is calculated as the total terrestrial C balance (green circle 'total') minus the sum of separate estimates of natural components (blue slices 'Natural components'), and can be compared to LULC CO_2 emission estimates by carbon-cycle models. …*

Deleted: changes through time

Deleted: The simulations

Deleted:

Deleted: is

Deleted: then the LULC forcing should be imposed within this framework but vegetation changes should be simulated in regions with no LULC change. …

Deleted: Thus, modelling groups who are running the *midHolocene* experiment with a fully interactive carbon cycle could also run the LULC experiment as a "free atmospheric CO_2" simulation.…

Deleted: Boundary conditions for CMIP6-PMIP4 and the mid-Holocene LULC experiments…

	Perihelion – 180	100.33	0.87	0.87
	Vernal equinox	Noon, 21 March	Noon, 21 March	Noon, 21 March
Greenhouse gases	Carbon dioxide (ppm)	284.3	264.4	264.4
	Methane (ppb)	808.2	597.0	597.0
	Nitrous oxide (ppb)	273.0	262.0	262.0
	Other GHG	DECK *piControl*	0	0
Other boundary conditions	Solar constant	TSI: 1360.747	As *piControl*	As *piControl*
	Palaeogeography	Modern	As *piControl*	As *piControl*
	Ice sheets	Modern	As *piControl*	As *piControl*
	Vegetation	Interactive	Interactive	pasture and crop distribution prescribed from a revised scenario
		DECK *piControl*	As *piControl*	pasture and crop distribution prescribed from a revised scenario
	Aerosols	interactive	Interactive	Interactive
		DECK *piControl*	As *piControl*	As *piControl*

The real strength of the revised LULC scenarios is to provide boundary conditions for transient climate-model simulations. The CMIP6-PMIP4 simulation of 850-1850 CE (*past1000*) already incorporates LULC changes as a forcing (Jungclaus et al. 2017), based on a harmonized data set that provides LULC changes from 850 through to 2015 CE (Hurtt et al., 2017), which in turn draws on output from the HYDE3.2 scenario (Klein Goldewijk et al., 2017a). The *past1000* protocol (Jungclaus et al., 2017) acknowledges that this default land-use data set is at the lower end of the spread in estimates of early agricultural area indicated by other LULC scenarios and recommends that modelling groups run additional sensitivity experiments using alternative maximum and minimum scenarios. The revised LULC scenarios created by LandCover6k could be used as an alternative to these maximum and minimum scenarios. Other than the substitution of the LandCover6k scenario, the specifications of other forcings would then follow the recommendations for the CMIP6-PMIP4 *past1000* simulation.

A transient climate simulation for a longer period of the Holocene would provide a more stringent test of the impact of LULC on the coupled earth system. We suggest that this transient simulation (*holotrans*) should start from the pre-existing *midHolocene* simulation to capitalise on the fact that the *midHolocene* simulation has 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). In order to be consistent with the CMIP6-PMIP4 *midHolocene* protocol (Otto-Bleisner et al., 2017), changes in orbital forcing should be specified from Berger and Loutre (1991) and year-by-year changes in CO_2, CH_4 and N_2O should be specified following Joos and Spahni (2008). LULC changes should be implemented by imposing crop and pasture area through time as specified in the revised LULC scenarios; elsewhere, the simulated vegetation should be active. It will be necessary

Deleted: LC6k

Deleted: imposed

Deleted: LC6k

Deleted: imposed

Deleted: data set

Deleted: We suggest that this transient simulation (*holotrans*) should start from the pre-existing *midHolocene* simulation to ensure that the ocean and land carbon cycle is in equilibrium (Table 2).…

to run the Holocene transient climate simulation in two steps. A first simulation (*holotrans LULC*) should be run using prescribed atmospheric CO_2 concentration even though the carbon cycle is fully interactive, because this will establish the consistency of the carbon cycle in the land surface model. However, once this is done it will be possible to re-run the simulations with interactive CO_2 emissions. Table 3 provides a summary of the proposed ESM simulations.

Table 2: Boundary conditions for baseline PMIP Holocene transient (6 ka BP to 1850 CE) and LULC transient simulations

		Mode	**Source/Value**	**LULC experiment**
Orbital parameters		transient		As baseline simulation
Greenhouse gases	CO_2	transient	Dome C	As baseline simulation
	CH_4		Combined EPICA & GISP record	As baseline simulation
	N_2O		Combined EPICA NGRIP, & TALDICE record	As baseline simulation
Solar forcing		transient	Steinhilber et al. (2012)	As baseline simulation
Volcanic forcing		transient	To be determined	As baseline simulation
Palaeogeography		Constant at PI values	Modern	As baseline simulation
Ice sheets		Constant at PI values	Modern	As baseline simulation
Vegetation		interactive		LC6k transient pasture and crop distribution imposed
Aerosols		Constant at PI values		As baseline simulation

Unlike the situation for the mid-Holocene, where there is a global climate benchmark data set (Bartlein et al., 2011) which provides reconstructions of multiple bioclimatic variables of seasonal temperature and moisture, the opportunities for quantitative evaluation of the *holotrans* simulated climate are more limited. Seasonal temperature reconstructions are available for Europe (Davis et al., 2003) and North America (Viau et al, 2006; Viau and Gajewski, 2009). Although there is a new global data set that provides global temperature reconstructions for the Holocene (Kaufman et al., in press), it is based on only 472 terrestrial records worldwide and the results for zonally averaged temperature changes are therefore likely to be more robust than the regional details. There are also 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). Furthermore, the simulated time-course of CO_2 emissions can be compared to the ice core records.

Table 3: Summary of proposed simulations.

Name	Mode	Purpose
piControl	equilibrium	Standard CMIP6-PMIP4 simulation

Deleted: prescribed in the atmosphere

Formatted: Normal, Left

Deleted: Unlike the situation for the mid-Holocene, where there is a global climate benchmark data set (Bartlein et al., 2011), quantitative evaluation of the *holotrans* simulated climate can only be made for key regions. Quantitative climate reconstructions through the Holocene are currently only available for Europe (Davis et al., 2003) and North America (Viau et al, 2006; Viau and Gajewski, 2009). However, 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).…

Formatted: Font: (Default) +Body (Helvetica Neue), Underline colour: Auto, Font colour: Auto, Pattern: Clear

midHolocene	equilibrium	Standard CMIP6-PMIP4 simulation
midHoloceneLULC	equilibrium	Sensitivity to LULC changes
holotrans	transient	Baseline fully transient simulation from 6ka onwards, with no LULC
holotrans_LULC	transient	Fully transient simulation from 6ka onwards, with LULC imposed

The CMIP6-PMIP4 *mid-Holocene* simulations are stylized experiments, lacking several potential forcings (in addition to LULC), including changes in atmospheric dust loading, in solar irradiance, and volcanic forcing. We suggest that additional sensitivity tests could be run to take these additional forcings into account. In the case of solar and volcanic forcing, this would also ensure that the transient *holotrans* simulations mesh seamlessly with the *past1000* simulation. Changes in solar variability during the Holocene should be specified from Steinhilber et al. (2012). There are records of volcanic forcing for the past 2000 years (Sigl et al., 2015; Toohey and Sigl, 2017), and these are used in the *past1000* simulation. Observationally constrained estimates of the volcanic stratospheric aerosol for Holocene are currently under development (M. Sigl, pers comm.) and could be implemented as an additional sensitivity experiment when available. Changes in atmospheric dust loading are not included in the *past1000* simulation but are important during the earlier part of the Holocene (Pausata et al., 2016; Tierney et al., 2017; Messori et al., 2019). Although continuous reconstructions of dust loading through the Holocene are not available, it would be possible to use estimates for particular time-slices (Egerer et al., 2018) to test the sensitivity to this forcing.

Outcomes and Perspectives

LandCover6k has developed a scheme for using archaeological information to improve existing scenarios of LULC changes during the Holocene, specifically by using archaeological data to provide better estimates of regional population changes through time, better information on the date of initiation of agriculture in a region, more regionally specific information about the type of land use, and more nuanced information about land-use per capita than currently implemented in the LULC scenarios generated by HYDE and KK10. While the final global data set are still in production, fast-track priority products have been created and their impact on current scenarios is being tested.

Although the work of LandCover6k will provide more solid knowledge about anthropogenic modification of the landscape, some information will inevitably be missing and some key regions will be poorly covered. There will still be large uncertainties associated with revised LULC scenarios, even though these will be based on more solid evidence than the existing LULC scenarios. Documenting the uncertainties in the archaeological inputs and their impacts on the revised scenarios is an important goal of the LandCover6k project. We propose using the information about the uncertainties in the archaeological data sources to generate multiple LULC scenarios comparable to the "low-end", "high-end" scenarios used for e.g. in future projections. Furthermore, we have proposed a series of tests that will help to evaluate the realism of the final scenarios, based on independent evidence from pollen-based reconstructions of land cover, reconstructions of climate, and carbon-cycle constraints. These tests should help in identifying which of the potential LULC reconstructions are most realistic and in constraining the sources of uncertainty.

Deleted: We have therefore proposed

We have proposed the use of offline carbon-cycle simulations solely as a test of the realism of the revised LULC scenarios. Quantifying the LULC contribution to CO_2 emissions during the Holocene would require additional simulations in which other forcings (climate, atmospheric CO_2, insolation) are kept constant. The difference in simulated total terrestrial C storage between these simulations and LULC simulations provides an estimate of *primary emissions* (Pongratz et al., 2014) and avoids additional model uncertainty regarding the sensitivity of land C storage to atmospheric CO_2 or climate being included in emission estimates. There are other sensitivity tests that would be useful. For example, vegetation-carbon-cycle models differ in their ability to account for gross land use transitions within grid cells (Arneth et al., 2017). This is critical for simulating effects of non-permanent agriculture where land is simultaneously abandoned and re-claimed within the extent of a model grid cell. Such shifting cultivation-type agriculture implies forest degradation in areas recovering from previous land use and leads to substantially higher LULC emissions compared to model estimates where only net land-use changes are accounted for (Shevliakova et al., 2009). It would therefore be interesting to run additional simulations accounting for net land use change, and indeed separating out the effects of wood harvesting and shifting cultivation.

We anticipate that it will be possible to incorporate realistic LULC scenarios for the mid-Holocene as part of the climate-model sensitivity experiments planned during PMIP4. Such experiments will complement the CMIP6-PMIP4 baseline model experiments, by providing insights into whether discrepancies between simulated and observed 6 ka climate could be the result of incorrect specification of the land-surface boundary conditions. However, the incorporation of archaeological information into LULC scenarios clearly makes it possible to target other interesting periods for such experiments, for example to explore if land-use changes played a role in abrupt events such as the 4.2 ka event, or to examine the impact of population declines in the Americas as a consequence of European colonisation (1500-1750 CE) or the changes in land use globally during the Industrial era (post 1850 CE).

In addition to providing a protocol for the PMIP 6ka sensitivity experiments, we have devised a protocol for implementing the optimal LULC reconstructions for the Holocene in transient climate-model or ESM experiments. The goal here is to provide one of the necessary forcings that could be used for transient simulations in future phases of PMIP. This will allow an assessment of LULC in these simulations, and therefore help address issues that are a focus for other MIPs e.g. LUMIP or LS3MIP. When these new forcings are created, they will be made available through the PMIP4 website (https: //pmip4.lsce.ipsl.fr/doku.php/exp_design:lgm, PMIP4 repository, 2017) and the ESGF Input4MIPS repository (https://esgf-node.llnl.gov/projects/input4mips/, with details provided in the "input4MIPs summary" link). Modelling groups who run either equilibrium or transient climate-model experiments following this protocol are encouraged to follow the standard CMIP protocol of archiving their simulations through the ESFG.

Code and Data Availability

The data used for Figure 1 are publicly available. The HYDE3.2 data can be downloaded https://doi.org/10.17026/dans-25g-gez3. The KK10 data can be downloaded from https://doi.org/10.1594/PANGAEA.871369. The code and data used to generate Figure 1 are available from https://github.com/jedokaplan/ALCC_comparison_figure. The data and code used to generate Figure 3 are available from https://github.com/mavdlind/GMD. The data and code used to generate Figure 5 are available from ?????. The European pollen-based reconstructions used in Figure 7 are available https://doi.org/10.1594/PANGAEA.897303. The pollen data used to generate the Siberian reconstructions is available from https://doi.org/10.1594/PANGAEA.898616. An earlier version of this Figure was published in Dawson et al., 2018. The code used to generate 7 Figure is

Deleted: vegetation-

Deleted: In addition to providing a protocol for the PMIP 6ka sensitivity experiments, we have devised a protocol for implementing the optimal LULC reconstructions for the Holocene in transient experiments. The goal here is to provide one of the necessary forcings that could be used for transient simulations in future phases of PMIP. This will allow an assessment of LULC in these simulations, and therefore help address issues that are a focus for other MIPs e.g. LUMIP or LS3MIP.¶

available from https://doi.org/10.5281/zenodo.3604328. The pollen-based reconstructions used in the generation of Figure 8 are available from 10.5281/zenodo.3601028. The climate model outputs used to generate Figure 8 are available from 10.5281/zenodo.3601040. The code used to generate Figure 8 is available from 10.5281/zenodo.3601011.

Acknowledgements. LandCover 6k is a working group of the Past Global Changes (PAGES) programme, which in turn received support from the Swiss Academy of Sciences. We thank PAGES for their support for this activity. The land use group also received funding under the Holocene Global Landuse International Focus Group of INQUA. SPH acknowledges funding from the European Research Council for "GC2.0: Unlocking the past for a clearer future". MJG thanks the Swedish Strategic Research Area MERGE (Modelling the Regional and Global Earth System Model) and Linnaeus University's faculty of Health and Life Sciences (Kalmar, Sweden) for financial support. We thank Joy Singarayer for providing the climate model outputs that were used to generate Figure 8 and Guangqi Li for assistance in producing this figure. BDS was funded by ERC H2020-MSCA-IF-2015, grant number 701329. The dataset for Figure 5 was generated from the 'Cultivating Societies: Assessing the Evidence for Agriculture in Neolithic Ireland' project, supported by the Heritage Council, Ireland under the INSTAR programme 2008–2010 (Reference 16682 to Whitehouse, Schulting, Bogaard and McClatchie).

Deleted: **Code and data availability**. All the climate model forcing data sets, their references, and their code will be provided on the PMIP4 website (https://pmip4.lsce.ipsl.fr/doku.php/exp_design:lgm, PMIP4 repository, 2017). All the pollen-based REVEALS land-cover reconstructions, and the land use data and associated code will be archived in PANGAEA (https://www.pangaea.de). ¶
¶
¶

Deleted: .

Deleted: .

Deleted: .

Author Contributions. SPH, MJG, BDS, MVL, KKG wrote the first draft. SPH, PB, FSRP contributed to the design of the climate model experiments, BS and TK to the design of the carbon-cycle simulations. The figures were contributed by JK (Fig. 1), BS (Fig. 2, Fig 9.), MVL (Fig. 3), OB (Fig. 4), NJW (Fig. 5), KKG (Fig. 6), AD (Fig. 7), SPH (Fig. 8). All authors contributed to the final version of the paper.

References

Arneth, A., Denton, F., Agus, F., Elbehri, A., Erb, K., Elasha, B.O., Rahimi, M., Rounsevell, M., Spence, A., and Valentini, R.: IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse gas fluxes in Terrestrial Ecosystems, 2019.

Arneth, A., Sitch, S., Pongratz, J., Stocker, B.D., Ciais, P., Poulter, B., Bayer, A.D., Bondeau, A., Calle, L., Chini, L.P., Gasser, T., Fader, M., Friedlingstein, P., Kato, E., Li, W., Lindeskog, M., Nabel, J.E.M.S., Pugh, T.A.M., Robertson, E., Viovy, N., Yue, C., and Zaehle, S.: Historical carbon dioxide emissions caused by land-use changes are possibly larger than assumed, Nature Geosci., 10, 79-84, doi: 10.1038/ngeo2882, 2017.

Balsera, V., Díaz-del-Río, P., Gilman, A., Uriarte, A., and Vicent, J.M.: Approaching the demography of late prehistoric Iberia through summed calibrated probability distributions (7000-2000 cal BC), Quat. Int., 208-211, doi:10.106/j.quaint.2015.06.022, 2015.

Bartlein, P. J., Harrison, S. P., Brewer, S., Connor, S., Davis B. A. S., Gajewski, K., Guiot, J., Harrison-Prentice, T. I., Henderson, A., Peyron, O., Prentice, I. C., Scholze, M., Seppä, H., Shuman, B., Sugita, S., Thompson, R. S., Viau, A., Williams, J., and Wu, H.: Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis, Clim. Dyn., 37, 775-802, doi: 10.1007/s00382-010-0904-1, 2011.

Bartlein, P.J., Harrison, S.P., and Izumi, K.: Underlying causes of Eurasian mid-continental aridity in simulations of mid-Holocene climate, Geophys. Res. Lett., 44, doi: 10.1002/2017GL074476, 2017.

Barton, C.M., Ullah, I.I., and Bergin, S.: Land use, water and Mediterranean landscapes: modelling long-term dynamics of complex socio-ecological systems, Phil. Trans. R. Soc., A368, 5275-5297, doi: 10.1098/rsta.2010.0193, 2010.

Berger, A., and Loutre, M-F.: Insolation values for the climate of the last 10 million of years, Quat. Sci. Rev., 10, 297-317, https://doi.org/10.1016/0277-3791(91)90033-Q, 1991.

Bishop, R.R., Church, M.J., and Rowley-Conwy, P.A.: Firewood, food and human niche construction: the potential role of Mesolithic hunter–gatherers in actively structuring Scotland's woodlands, Quat. Sci. Rev., 108, 51-75, 2015.

Bocquet-Appel, J.-P., Naji, S., Vander Linden, M., and Kozłowsi, J.K.: Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of ^{14}C dates, J. Arch. Sci., 36, 807-820, doi: 10.1016/j.jas.2008.11.004, 2009.

Crema, E.R., Habu, J., Kobayashi, K., and Madella, M.: Summed probability distribution of ^{14}C dates suggests regional divergences in the population dynamics of the Jomon period in eastern Japan, PlosOne, 11, e0154809, doi: 10.1371/journal.pone.0154809, 2016.

Davis, B.A.S., Brewer, S., Stevenson, A.C., Guiot, J., and Juggins, S.: The temperature of Europe during the Holocene reconstructed from pollen data, Quat. Sci. Rev., 22, 1701–1716, 2003.

Davison, K., Dolukhanov, P., Sarson, G. R., and Shukurov, P.: The role of waterways in the spread of the Neolithic, J. Arch. Sci., 33, 641-652, doi: 10.106/j.jas.2005.09.017, 2006.

Dawson, A., Paciorek, C.J., McLachlan, J.S., Goring, S., Williams, J.W., and Jackson, S.T.: Quantifying pollen-vegetation relationships to reconstruct ancient forests using 19th-century forest composition and pollen data, Quat. Sci. Rev., 137, 156-175, doi: 10.1016/j.quascirev.2016.01.012, 2016.

Dawson, A., Cao, X., Chaput, M., Hopla, E., Li, F., Edwards, M., Fyfe, R., Gajewski, K., Goring, S.J., Herzschuh, U., Mazier, F., Sugita, S., Williams, J.W., Xu, Q., and Gaillard, M-J.: Finding the magnitude of human induced Northern Hemisphere land-cover transformation between 6 and 0.2 ka BP, PAGES Mag., 26, 34-35, https://doi.org/10.22498/pages.26.1.34, 2018.

Egerer, S., Claussen, M., and Reick, C.: Rapid increase in simulated North Atlantic dust deposition due to fast change of northwest African landscape during the Holocene, Clim. Past, 14, 1051-1066, https://doi.org/10.5194/cp-14-1051-2018, 2018.

Ellis, E.C., Kaplan, J.O., Fuller, D.Q., Vavrus, S., Klein Goldewijk, K., and Verburg, P. H.: Used planet: A global history, Proc. Nat. Acad. Sci., 110, 7978-7985, 2013.

Elsig, J., Schmitt, J., Leuenberger, D., Schneider, R., Eyer, M., Leuenberger, M., Joos, F., Fischer, H., and Stocker, T. F.: Stable isotope constraints on Holocene carbon cycle changes from an Antarctic ice core, Nature, 461, 507-510, 2009.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fischer, N., and Jungclaus, J.H.: Evolution of the seasonal temperature cycle in a transient Holocene simulation: orbital forcing and sea-ice, Clim. Past, 7, 1139-1148, https://doi.org/10.5194/cp-7-1139-2011, 2011.

Ford, A., and Clarke, K.C.: Linking the past and present of the ancient Maya: lowland land use, population distribution, and density in the Late Classic Period, in: The Oxford Handbook of Historical Ecology and Applied Archaeology, Isendahl, C. and Stump, D. (eds.), doi: 10.1093/oxfordhb/9780199672691.013.33, 2015.

Freeman, J., Baggio, J.A., Robinson, E., Byers, D.A., Gayo, E., Finley, J.B., Meyer, J.A., Kelly, R.L., and Anderies, J.M.: Synchronisation of energy consumption by human societies throughout the Holocene, Proc. Nat. Acad. Sci., 115, 9962-9967, doi: 10.1073/pnas.1802859115, 2018.

Fyfe R. M., Woodbridge, J. E., and Roberts, N.: From forest to farmland: pollen-inferred land cover change across Europe using the pseudobiomization approach, Glob. Change Biol., 21: 1197-

Deleted: Ellis, E.C., and Ramankutty, N.: Putting people in the map: Anthropogenic biomes of the world, Frontiers Ecol. Environ., 6, 439–447, doi: 10.1890/070062, 2008.¶ Ellis, E.C., Klein Goldewijk, K., Siebert, S., Lightman, D., and Ramankutty, N.: Anthropogenic transformation of the biomes, 1700 to 2000, Glob. Ecol. Biogeog., 19, 589-606, 2010.¶

Deleted: Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N., and Snyder, P. K.: Global consequences of land use, Science, 309, 570-574, 2005.¶

1212, doi:10.1111/gcb.12776, 2014.
Gaillard, M.-J., Sugita, S., Mazier, F., Kaplan, J.O., Trondman, A.-K., Brostroem, A., Hickler, T., Kjellstroem, E., Kunes, P., Lemmen, C., Olofsson, J., Smith, B., and Strandberg, G.: Holocene land-cover reconstructions for studies on land-cover feedbacks, Clim. Past, 6, 483-499, https://doi.org/10.5194/cp-6-483-2010, 2010.
Gaillard, M-J., Whitehouse, N., Madella, M., Morrison, K., and von Gunten, L: Past land use and land cover, PAGES Mag. ,26, 1-44, doi:10.22498/pages.26.1, 2018.
Harrison, S.P., Bartlein, P.J., Izumi, K., Li, G., Annan, J., Hargreaves, J., Braconnot, P.B., and Kageyama, M.: Evaluation of CMIP5 palaeo-simulations to improve climate projections, Nature Clim. Change, 5, 735-743, 2015.
He, F., Vavrus, S.J., Kutzbach, J.E., Ruddiman, W.F., Kaplan, J.O., and Krumhardt, K.M.: Simulating global and local surface temperature changes due to Holocene anthropogenic land cover change, Geophys. Res. Lett., 41, 623–631, 2014.
Hellman, S., Gaillard, M.-J., Broström, A., and Sugita, S.: The REVEALS model, new tool to estimate past regional plant abundance from pollen data in large lakes: validation in southern Sweden, J. Quat. Sci., 22, 1-22, 2008a.
Hellman, S., Gaillard M.-J., Broström, A., and Sugita, S.: Effects of the sampling design and selection of parameter values on pollen-based quantitative reconstructions of regional vegetation: a case study in southern Sweden using the REVEALS model, Veg. History Archaeobot., 17, 445-460, 2008b.
Herrero, M., Havlík, P., Valin, H., Notenbaert, A., Rufino, M.C., Thornton, P.K., Blümmel, M., Weiss, F., Grace, D., and Obersteiner, M.: Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems, Proc. Nat. Acad. Sci., 110, 20888-20893, 2013.
Hughes, R.E., Weiberg, E., Bonnier, A., Finne, M., and Kaplan, J.O.: Quantifying land use in past societies from cultural practice and archaeological data, Land, 7, 9, doi.org/10.3390/land/7010009, 2018.
Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Calvin, K., Fujimori, S., Klein Goldewijk, K., Hasegawa, T., Havlik, P., Lawrence, D., Lawrence, P., Popp, A., Stehfest, E., van Vuuren, D., and Zhang, X.: Harmonization of global land-use change and management for the period 850–2100, Geosci. Model Dev. Discuss., 2017.
Janssen, E., Poblome, J., Claeys, J., Kint, V., Degryse, P., Marinova, E., and Muys, B.: Fuel for debating ancient economies. Calculating wood consumption at urban scale in Roman Imperial times, J. Arch. Sci.: Reports, 11, 592-599, 2017.
Joos, F., and Spahni, R.: Rates of change in natural and anthropogenic radiative forcing over the past 20,000 years, Proc. Nat. Acad. Sci., 105, 1425–1430, 2008.
Joos, F., Gerber, S., Prentice, I.C., Otto-Bliesner, B.L., and Valdes, P.J.: Transient simulations of Holocene atmospheric carbon dioxide and terrestrial carbon since the last glacial maximum, Global Biogeochem. Cy., 18, GB2002, doi:10.1029/2003GB002156, 2004.
Jungclaus, J.H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., Egorova, T., Evans, M., González-Rouco, J.F., Goosse, H., Hurtt, G.C., Joos, F., Kaplan, J. O., Khodri, M., Klein Goldewijk, K., Krivova, N., LeGrande, A.N., Lorenz, S. J., Luterbacher, J., Man, W., Maycock, A.C., Meinshausen, M., Moberg, A., Muscheler, R., Nehrbass-Ahles, C., Otto-Bliesner, B.I., Phipps, S.J., Pongratz, J., Rozanov, E., Schmidt, G.A., Schmidt, H., Schmutz, W., Schurer, A., Shapiro, A.I., Sigl, M., Smerdon, J.E., Solanki, S.K., Timmreck, C., Toohey, M., Usoskin, I.G., Wagner, S., Wu, C.-J., Yeo, K.L., Zanchettin, D., Zhang, Q., and Zorita, E.: The PMIP4 contribution to CMIP6 – Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations, Geosci. Model Dev., 10, 4005–4033, https://doi.org/10.5194/gmd-10-4005-2017, 2017.
Kageyama, M., Braconnot, P., Harrison, S.P., Haywood, A., Jungclaus, J., Otto-Bliesner, B.,

Peterschmitt, J.-Y., Abe-Ouchi, A., Albani, S., Bartlein, P., Brierley, C., Crucifix, M., Dolan, A., Fernandez-Donado, L., Fischer, H., Hopcroft, P., Ivanovic, R., Lambert, F., Lunt, D., Mahowald, N., Peltier, W.R., Phipps, S., Roche, D., Schmidt, G., Tarasov, L., Valdes, P., Zhang, Q., and Zhou, T.: The PMIP4 contribution to CMIP6 – Part 1: Overview and over-arching analysis plan. Geosci. Model Dev., 11: 1033-1057. https://doi.org/10.5194/gmd-11-1033-2018, 2018.

Kaplan, J.O., Krumhardt, K.M., Ellis, E.C., Ruddiman, W.F., Lemmen, C., and Klein Goldewijk, K.: Holocene carbon emissions as a result of anthropogenic land cover change, Holocene, 21, 775-791, 2011.

Kaplan, J.O., Krumhardt, K.M., Gaillard, M-J., Sugita, S., Trondman, A-K., Fyfe, R., Marquer, L., Mazier, F., and Nielsen, A.B.: Constraining the deforestation history of Europe: Evaluation of historical land use scenarios with pollen-based land cover reconstructions, Land, 6, 9, doi:10.3390/land6040091, 2017.

Kaufman, D., McKay, N., Routson, C., Erb, M., Davis, B., Heiri, O., Jaccard, S., Tierney, J., Dätwyler, C., et al.: A global database of Holocene paleo-temperature records, Scientific Data, in press.

Kay, A. U., Fuller, D. Q., Neumann, K., Eichhorn, B., Höhn, A., Morin-Rivat, J., Champion, L., Linseele, V., Huysecom, E., Ozainne, S., Lespez, L., Biagetti, S., Madella, M., Salzmann, U., and Kaplan, J. O.: Diversification, intensification, and specialization: Changing land use in western Africa from 1800 BC to AD 1500, J. World Prehistory, 32, 179–228, doi: 10.1007/s10963-019-09131-2. 2019, 2019.

Klein Goldewijk, K., Beusen, A., van Drecht, G., and de Vos, M.: The HYDE 3.1 spatially explicit database of human induced land use change over the past 12,000 years, Glob. Ecol. Biogeog., 20, 73-86, 2011.

Klein Goldewijk, K., Beusen, A., Doelman, J., and Stehfest, E.: Anthropogenic land-use estimates for the Holocene; HYDE 3.2, Earth Syst. Sci. Data, 9, 927-953, https://doi.org/10.5194/essd-9-927-2017, 2017a.

Klein Goldewijk, K., Dekker, S.C., and van Zanden, J.L.: Per-capita estimations of long-term historical land use and the consequences for global change research, J. Land Use Sci., 12, 313-337, https://doi.org/10.1080/1747423X.2017.1354938, 2017b.

Lawrence, D.M., Hurtt, G.C., Arneth, A., Brovkin, V., Calvin, K.V., Jones, A.D., Jones, C.D., Lawrence, P.J., de Noblet-Ducoudré, N., Pongratz, J., Seneviratne, S.I., and Shevliakova, E.: The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: rationale and experimental design, Geosci. Model Dev., 9, 2973-2998, https://doi.org/10.5194/gmd-9-2973-2016, 2016.

Le Quéré, C., Andrew, R.M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P.A., Korsbakken, J.I., Peters, G.P., Canadell, J.G., Arneth, A., Arora, V.K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L.P., Ciais, P., Doney, S.C., Gkritzalis, T., Goll, D.S., Harris, I., Haverd, V., Hoffman, F.M., Hoppema, M., Houghton, R.A., Hurtt, G., Ilyina, T., Jain, A.K., Johannessen, T., Jones, C.D., Kato, E., Keeling, R.F., Goldewijk, K.K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D.R., Nabel, J.E.M.S., Nakaoka, S.-I., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P.P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I.T., van der Werf, G.R., Viovy, N., Walker, A.P., Wiltshire, A.J., Wright, R., Zaehle, S., and Zheng, B.: Global Carbon Budget 2018, Earth Syst. Sci. Data, 10, 2141-2194, https://doi.org/10.5194/essd-10-2141-2018, 2018.

Loisel, J., Yu, Z., Beilman, D.W., Camill, P., Alm, J., Amesbury, M.J., Anderson, D., Andersson, S., Bochicchio, C., Barber, K., Belyea, L.R., Bunbury, J., Chambers, F.M., Charman, D.J., Vleeschouwer, F.D., Fiałkiewicz-Kozieł, B., Finkelstein, S.A., Gałka, M., Garneau, M.,

Deleted: in press

Deleted: https://doi.org/10.5194/essd-9-1-2017

Hammarlund, D., Hinchcliffe, W., Holmquist, J., Hughes, P., Jones, M.C., Klein, E.S., Kokfelt, U., Korhola, A., Kuhry, P., Lamarre, A., Lamentowicz, M., Large, D., Lavoie, M., MacDonald, G., Magnan, G., Mäkilä, M., Mallon, G., Mathijssen, P., Mauquoy, D., McCarroll, J., Moore, T.R., Nichols. J., O'Reilly, B., Oksanen, P., Packalen, M., Peteet, D., Richard, P.J., Robinson, S., Ronkainen, T., Rundgren, M., Sannel, A.B.K., Tarnocai, C., Thom, T., Tuittila, E.-S., Turetsky, M., Väliranta, M., and der Linden, M., van Geel B., van Bellen, S., Vitt, D., Zhao, Y., and Zhou, W.: A database and synthesis of northern peatland soil properties and Holocene carbon and nitrogen accumulation, Holocene, 24, 1028-1042, 2014.

Lyman, R. L.: Quantitative Paleozoology. Cambridge University Press, 2008.

Maezumi, S.Y., Robinson, M., de Souza, J., Urrego, D.H., Schaan, D., Alves, D., and Iriarte, J.: New insights from pre-Columbian land use and fire management in Amazonian Dark Earth forests, Front. Ecol. Evol., 6, 111, doi: 10.3389/fevo.2018.00111, 2018.

Mahowald, N.M., Randerson, J.T., Lindsay, K., Munoz, E., Doney, S.C., Lawrence, P., Schlunegger, S., Ward, D.S., Lawrence, D., and Hoffman, F. M.: Interactions between land use change and carbon cycle feedbacks, Glob. Biogeochem. Cy., 31, 96–113, doi:10.1002/2016GB005374, 2017.

Marston, J.M.: Modeling wood aquisition strategies from archaeological charcoal remains, J. Arch. Sci., 36, 2192-2200, 2009.

Marston, J.M., Holdaway, S.J., and Wendrich, W.: Early- and middle-Holocene wood exploitation in the Fayum basin, Egypt, Holocene, 27, 1812-1824, 2017.

Mauri, A., Davis, B. A. S., Collins, P. M., and Kaplan, J. O.: The influence of atmospheric circulation on the mid-Holocene climate of Europe: a data–model comparison, Clim. Past, 10, 1925–1938, https://doi.org/10.5194/cp-10-1925-2014, 2014.

Mazoyer, M., and Roudart, L.: A History of World Agriculture: From the Neolithic to the Current Crisis. Earthscan, UK, 2006.

McGrath, M. J., Luyssaert, S., Meyfroidt, P., Kaplan, J. O., Burgi, M., Chen, Y., Erb, K., Gimmi, U., McInerney, D., Naudts, K., Otto, J., Pasztor, F., Ryder, J., Schelhaas, M. J., and Valade, A.: Reconstructing European forest management from 1600 to 2010, Biogeosci., 12, 4291-4316. doi:10.5194/bg-12-4291-2015, 2015.

McLaughlin, T. R., Whitehouse, N. J., Schulting, R. J., McClatchie, M., Barratt, P. and Bogaard, A.: The changing face of Neolithic and Bronze Age Ireland: A big data approach to the settlement and burial records. J. World Prehist., 29, 117-153. doi:10.1007/s10963-016-9093-0, 2016.

Messori, G., Gaetani, M., Zhang, Q., Zhang, Q., and Pausata, F. S. R.: The water cycle of the mid-Holocene West African monsoon: The role of vegetation and dust emission changes, Int. J. Climatol., 39, 1927– 1939, https://doi.org/10.1002/joc.5924, 2019.

Mitchell, L., Brook, E., Lee, J., Buizert, C., and Sowers, T.: Constraints on the late Holocene anthropogenic contribution to the atmospheric methane budget, Science, 342, 964–966, doi:10.1126/science.1238920, 2013.

Morrison, K.D., Hammer, E., Popova, L., Madella, M., Whitehouse, N., Gaillard, M.-J. and LandCover6k Land-Use Group Members: Global-scale comparisons of human land use: developing shared terminology for land-use practices for global changes, PAGES Mag., 26, 8-9, 2018.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura T., and Zhang, H.: Anthropogenic and natural radiative forcing. in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (T F Stocker, D Qin, G-K Plattner, M Tignor, S K Allen, J Boschung, A Nauels, Y Xia, V Bex and P M Midgley (eds), Cambridge: Cambridge University Press Cambridge, United Kingdom and New York, NY, USA, 2013.

Nakagawa, T., Tarasov, P.E., Nishida, K., Gotanda, K., and Yasuda, Y.: Quantitative pollen-based climate reconstruction in central Japan: application to surface and Late Quaternary spectra, Quat. Sci. Rev., 21, 2099–2113, 2002.

Nielsen, A.B., Giesecke, T., Theuerkauf, M., Feeser, I., Behre, K.-H., Beug, H.-J., Chen, S.-H., Christiansen, J., Dörfler, W., Endtmann, E., Jahns, S., de Klerk, O., Kühl, N., Latałowa, M., Odgaard, B.V., Rasmussen, P., Stockholm, J.R., Voigt, R., Wiethold, J., and Wolters, S.: Quantitative reconstructions of changes in regional openness in north-central Europe reveal new insights into old questions, Quat. Sci. Rev., 47, 131–149, 2012.

Oh, Y., Conte, M., Kang, S., Kim, J., and Hwang, J.: Population fluctuation and the adoption of food production in prehistoric Korea: using radiocarbon dates as a proxy for population change, Radiocarbon, 59, 1761-1770, doi: 10.1017/RDC.2017.1.22, 2017.

Ortega-Rosas, C.I., Guiot, J., Penalba, M.C., Ortiz-Acosta, M.E.: Biomization and quantitative climate reconstruction techniques in northwestern Mexico—with an application to four Holocene pollen sequences, Glob. Planet. Change, 61, 242–266, 2008.

Orton, D., Gaastra, J., and Vander Linden, M.: Between the Danube and the Deep Blue Sea: zooarchaeological meta-analysis reveals variability in the spread and development of Neolithic farming across the western Balkans, Open Quat., 2, doi: 10.5334/oq.28, 2016.

Otto-Bliesner, B.L., Braconnot, P., Harrison, S.P., Lunt, D.J., Abe-Ouchi, A., Albani, S., Bartlein, P.J., Capron, E., Carlson, A.E., Dutton, A., Fischer, H., Goelzer, H., Govin, A., Haywood, A., Joos, F., Legrande, A.N., Lipscomb, W.H., Lohmann, G., Mahowald, N., Nehrbass-Ahles, C., Pausata, F.S.R., Peterschmidt, J-Y., Phipps, S.J., Renssen, R., and Zhang, Q.: The PMIP4 contribution to CMIP6 – Part 2: Two interglacials, scientific objective and experimental design for Holocene and Last Interglacial simulations. Geosci. Mod. Dev., 10, 3979-4003, https://doi.org/10.5194/gmd-10-1-2017, 2017.

Pausata, F.S.R., Messori, G., Zhang, Q.: Impacts of dust reduction on the northward expansion of the African monsoon during the Green Sahara period, Earth Planet. Sci. Lett., 434, 298-307,https://doi.org/10.1016/j.epsl.2015.11.049, 2016.

Perugini, L. Caporaso, L., Marconi, S., Cescatti, A., Quesada, B., de Noblet-Ducoudré, N., House, J.I., and Arneth, A.: Biophysical effects on temperature and precipitation due to land cover change, Environ. Research Lett., 12, 053002, https://doi.org/10.1088/1748-9326/aa6b3f, 2017.

Phelps, L.N., and Kaplan, J.O.: Land use for animal production in global change studies: Defining and characterizing a framework, Glob Chang Biol, 23, 4457-4471, 10.1111/gcb.13732, 2017.

Pirzamanbein, B., Lindström, J., Poska, A., Sugita, S., Trondman, A., Fyfe, R., Mazier, F., Nielsen, A.B., Kaplan, J.O., Bjune, A.E., Birks, H.J.B., Giesecke, T., Kangur, M., Latałowa, M., Marquer, L., Smith, B., and Gaillard, M.-J.: Creating spatially continuous maps of past land cover from point estimates: A new statistical approach applied to pollen data, Ecol. Complexity, 20, 127-141, 2014.

Pirzamanbein, B., Lindström, J., Poska, A., and Gaillard, M.-J.: Modelling spatial compositional data: Reconstructions of past land cover and uncertainties, Spatial Stat., 24, 14–31, 2018.

Pongratz, J., Reick, C., Raddatz, T., and Claussen, M.: A reconstruction of global agricultural areas and land cover for the last millennium, Glob. Biogeochem. Cy., 22, 2008.

Pongratz, J., Reick, C.H., Raddatz, T., and Claussen, M.: Biogeophysical versus biogeochemical climate response to historical anthropogenic land cover change, Geophys. Res. Lett., 37, L08702, doi:10.1029/2010GL043010, 2010.

Pongratz, J., Reick, C. H., Houghton, R. A., and J. I. House, J. I.: Terminology as a key uncertainty in net land use and land cover change carbon flux estimates, Earth Syst. Dynam., 5, 177–195, www.earth-syst-dynam.net/5/177/2014/ doi:10.5194/esd-5-177-2014, 2014.

Ramankutty, N., and Foley, J.A.: Estimating historical changes in global land cover: Croplands from 1700 to 1992, Glob. Biogeochem. Cy., 13, 997-1027, 1999.

Reitz, E. J., and Wing E.S.: Zooarchaeology. Cambridge University Press, 2008.

Rick, J.W.: Dates as data: an examination of the Peruvian Preceramic radiocarbon record, Am. Antiq., 52, 55–73, 1987.

Robinson, E., Zahid, H.J., Codding, B.F., Haas, R., and Kelly, R.L.: Spatiotemporal dynamics of prehistoric human population growth: radiocarbon 'dates as data' and population ecology models, J. Arch. Sci., 101, 63-71, 2019.

Ruddiman, W. F.: The anthropogenic greenhouse era began thousands of years ago, Clim. Change, 61, 261–293, doi:10.1023/B:CLIM.0000004577.17928.fa, 2003.

Russell, T., Silva, F., and Steele, J.: Modelling the spread of farming in the Bantu-speaking regions of Africa: an archaeology-based phylogeography, PlosONe, 9, e87584, doi: 10.1371/journla.pone.0087584, 2014.

Shennan, S., Downey, S.S., Timpson, A., Edinborough, K., Colledge, S., Kerig, T., Manning, K., and Thomas, M.G.: Regional population collapse followed initial agriculture booms in mid-Holocene Europe, Nat. Comms., 4, 248,. doi: 10.1038/ncomms3486, 2013.

Shevliakova, E., Pacala, S. W., Malyshev, S., Hurtt, G. C., Milly, P. C. D., Caspersen, J. P., Sentman, L. T., Fisk, J. P., Wirth, C., and Crevoisier, C.: Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink, Glob. Biogeochem. Cy., 23, GB2022, doi:10.1029/2007GB003176, 2009.

Sigl, M., Winstrup, M., McConnell, J.R., Welten, K.C., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., Fischer, H., Kipfstuhl, S., Kostick, C., Maselli, O.J., Mekhaldi, F., Mulvaney, R., Muscheler, R., Pasteris, D.R., Pilcher, J.R., Salzer, M., Schüpbach, S., Steffensen, J.P., Vinther, B.M., and Woodruff, T.E.: Timing and climate forcing of volcanic eruptions for the past 2,500 years, Nature, 523, 543–549, https://doi.org/10.1038/nature14565, 2015.

Silva, F., and Steele, J.: New methods for reconstructing geographical effects on dispersal rates from large-scale radiocarbon databases, J. Arch. Sci., 52, 609-620, doi: 10.1016/j.jas.2014.04.021, 2014.

Silva, F., and Vander Linden, M.: Amplitude of travelling front as inferred from ^{14}C predicts levels of genetic admixture among European early farmers, Sci. Reports, 7, 11985, doi: 10.1038/s41598-017-12318-2, 2017.

Silva, F., Stevens, C.J., Weisskopf, A., Castillo, C., Qin, L., Bevan, A., and Fuller, D.Q.: Modelling the geographical origin of rice cultivation in Asia using the Rice Archaeological Database, PlosOne, 10, e0137024, 2015.

Singarayer, J.S., Valdes, P.J., Friedlingstein, P., Nelson, S., and Beerling, D.J.: Late Holocene methane rise caused by orbitally controlled increase in tropical sources, Nature, 470, 82– 85, doi:10.1038/nature09739, 2011.

Smith, M.C., Singarayer, J.S., Valdes, P.J., Kaplan, J.O., and Branch, N.P.: The biogeophysical climatic impacts of anthropogenic land use change during the Holocene, Clim. Past, 12, 923-941, doi: https://doi.org/10.5194/cp-12-923-2016, 2016.

Steinhilber, F., Abreu, J. A., Beer, J., Brunner, I., Christl, M., Fischer, H., Heikkilä, U., Kubik, P. W., Mann, M., McCracken, K.G., Miller, H., Miyahara, H., Oerter, H., and Wilhelms, F.: 9400 years of cosmic radiation and solar activity from ice cores and tree rings, Proc. Natl. Acad. Sci., 109, 5967–5971, 2012.

Stocker, B. D., Strassmann, K., and Joos, F.: Sensitivity of Holocene atmospheric CO2 and the modern carbon budget to early human land use: analyses with a process-based model, Biogeosci., 8, 69-88, doi:10.5194/bg-8-69-2011, 2011.

Stocker, B.D., Yu, Z., Massa, C., and Joos, F.: Holocene peatland and ice-core data constraints on the timing and magnitude of CO_2 emissions from past land use, Proc. Natl. Acad. Sci., 114, 1492-1497, doi:10.1073/pnas.1613889114, 2017.

Styring, A., Rösch, M., Stephan, E., Stika, H.-P., Fischer, E., Sillmann, E., and Bogaard, A.: Centralisation and long-term change in farming regimes: comparing agricultural practices in

Neolithic and Iron Age south-west Germany, Proc. Prehist. Soc., 83: 357-381, doi: 10.1017/ppr.2017.3, 2017.

Sugita, S.: Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition, Holocene, 17, 229–241, 2007.

Tarasov, P., Williams, J.W., Andreev, A., Nakagawa, T., Bezrukova, E., Herzschuh, U., Igarashi, Y., Müller, S., Werner, K., and Zheng, Z.: Satellite- and pollen-based quantitative woody cover reconstructions for northern Asia: Verification and application to late-Quaternary pollen data, Earth Planet. Sci. Lett., 264, 284–298, 2007.

Tauger, M.B: Agriculture in World History, Routledge, 2013.

Tierney, J.E., Pausata, F.S.R., and deMenocal, P.B.: Rainfall regimes of the Green Sahara, Sci. Advan., 3, e1601503, 2017.

Timpson, A., Colledge, S., Crema, E., Edinborough, K., Kerig, T., Manning, K., Thomas, M. G. & Shennan, S.: Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method, J. Arch. Sci., 52, 549-557, doi: 10.1016/j.jas.2014.08.011, 2014.

Toohey, M. and Sigl, M.: Volcanic stratospheric sulphur injections and aerosol optical depth from 500 BCE to 1900 CE, Earth Syst. Sci. Data, 9, 809-831, https://doi.org/10.5194/essd-9-809-2017, 2017.

Trondman, A. K., Gaillard, M-J., Mazier, F., Sugita, S., Fyfe, R., Nielsen, A.B., Twiddle, C., Barratt, P., Birks, H.J.B., Bjune, A. E., Björkman, L., Broström, A., Caseldine, C., David, R., Dodson, J., Dörfler, W., Fischer, E., van Geel, B., Giesecke, T., Hultberg, T., Kalnina, L., Kangur, M., van der Knaap, P., Koff, T., Kuneš, P., Lagerås, P., Latałowa, M., Lechterbeck, J., Leroyer, C., Leydet, M., Lindbladh, M., Marquer, L., Mitchell, F.J. G., Odgaard, B.V., Peglar, S.M., Persson, T., Poska, A., Rösch, M., Seppä, H., Veski, S., and Wick, L.: Pollen-based quantitative reconstructions of Holocene regional vegetation cover (plant-functional types and land-cover types) in Europe suitable for climate modelling, Glob. Change Biol., 21, 676-697, doi:10.1111/gcb.12737, 2015.

Trondman, A.-K., Gaillard, M.-J., Sugita, S., Björkman, L., Greisman, A., Hultberg, T., Lagerås, P., and Lindbladh, M.: Are pollen records from small sites appropriate for REVEALS model-based quantitative reconstructions of past regional vegetation? An empirical test in southern Sweden, Veget. Hist. Archaeobot., 25, 131–151, doi: 10.1007/s00334-015-0536-9, 2016.

van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S.I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M.J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D.M., Weedon, G.P., Ellis, R., Hagemann, S., Mao, J., Flanner, M.G., Zampieri, M., Materia, S., Law, R.M., and Sheffield, J.: LS3MIP (v1.0) contribution to CMIP6: the Land Surface, Snow and Soil moisture Model Intercomparison Project – aims, setup and expected outcome, Geosci. Model Dev., 9, 2809-2832, https://doi.org/10.5194/gmd-9-2809-2016, 2016.

Vavrus, S., Ruddiman, W.F., and Kutzbach, J.E.: Climate model tests of the anthropogenic influence on greenhouse-induced climate change: the role of early human agriculture, industrialization, and vegetation feedbacks, Quat. Sci. Rev., 27, 1410-1425, 2008.

Veal, R., 2017. Wood and charcoal for Rome: towards an understanding of ancient regional fuel economics, In de Haas, T. & Gijs, T. (eds), *Rural communities in a globalizing economy: new perspectives on the economic integration of Roman Italy,* Brill, (New York and Leiden): pp.388-406.

Viau, A.E., and Gajewski, K.: Reconstructing millennial, regional paleoclimates of boreal Canada during the Holocene, J. Clim., 22, 316–330, 2009.

Viau, A., Gajewski, K., Sawada, M., and Fines, P.: Mean-continental July temperature variability in North America during the past 14,000 years, J. Geophys. Res. Atmos., 111, D09102, doi:10.1029/2005JD006031, 2006

Weiberg, E., Hughes, R. E., Finné, M., Bonnier, A., and Kaplan, J. O.: Mediterranean land use

systems from prehistory to antiquity: a case study from Peloponnese (Greece), J. Land Use Sci., 1-20, doi:10.1080/1747423x.2019.1639836, 2019.

Whitehouse, N., Schulting, R. J., McClatchie, M., Barratt, P., LcLaughlin, T.R., Bogaard, A., Colledge, S., Marchant, R., Gaffrey, J., and Bunting, M.J.: Neolithic agriculture on the European western frontier: the boom and bust of early farming in Ireland, J. Arch. Sci., 51, 181-205, doi: 10.1016/j.jas.2013.08.009, 2014.

Williams, A.: The use of summed radiocarbon probability distributions in archaeology: a review of methods. J. Archaeol. Sci., 39, 578–589, https://doi.org/10.1016/j.jas.2011.07.014, 2012.

Wilmshurst, J.M., McGlone, M.S., Leathwick, J.R., and Newnham, R.M.: A pre-deforestation pollen-climate calibration model for New Zealand and quantitative temperature reconstructions for the past 18000 years BP, J. Quat. Sci., 22, 535–547, 2007.

Wright, P.: Preservation or destruction of plant remains by carbonization. J.Arch.Sci 30, 577-583, doi: 10.1016/S0305-4403(02)00203-0, 2003.

Zahid, H.J., Robinson, E., and Kelly, R.L.: Agriculture, population growth, and statistical analysis of the radiocarbon record, Proc. Natl. Acad. Sci., 113, 931-935, doi: 10.1073/pnas.1517650112, 2016.

Zanon, M., Davis, B.A.S., Marquer, L., Brewer, S., and Kaplan, J.O.: European forest cover during the past 12,000 years: a palynological reconstruction based on modern analogs and remote sensing, Front. Plant Sci., 9, 253, doi: 10.3389/fpls.201800253, 2018.

Zimmermann, A., Wendt, K.P., and Hilpert, J.: Landscape archaeology in central Europe. Proc. Prehist. Soc., 75, 1-53, doi: 10.1017/S007949X00000281, 2009.

Zohary, D., Hopf, M., and Weiss, E.: Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in South-west Asia, Europe, and the Mediterranean Basin, 4th Edn., Oxford University Press, Oxford, 2012.