Responses to reviewer 1

Comment: Before I start I need to emphasise I do not work in the palaeo area so I cannot comment at all on any aspect of this paper relating to reconstructions. I'm a land surface modeller with lots of LULCC experience and in particular I am interested in the robustness of the methodology.

Reply: We thank you for your review and constructive comments.

Methodological issues

Comment: I am not familiar with Climate of the Past so these comments might be unfair, but in LULCC simulations one wants to tease out a signal from noise associated with model parameters, model parameterisations and internal model variability. That implies two fundamental needs. First, you need to run a reasonable number of ensemble members, perturbing parameters and parameterisations. For example, if you use other equally legitimate LULCC reconstructions (I mean, you propogate uncertainty in the reconstruction you use, or how you translate this reconstruction into model parameters, or how LPJ Guess simulates land cover etc) are your results robust, or are your results basically the consequence of a single set of choices? Second, you need to robustly assess the statisitcal significance of yourresults and t-tests are inadequate. Have a look at:

Wilks, D.S., 2016, The stippling shows statistically significant grid points: How research results are routinely overstated and overinterpreted, and what to do about it, Bulletin of the American Meteorological Society, doi: 0.1175/BAMS-D-15-00267.1

but, in addition look at:

Lorenz, R., A.J. Pitman, and S.A. Sisson, 2016, Does Amazonian deforestation cause global effects; can we be sure?, J. Geophysical Research, 121, 5567-5584, doi:10.1002/2015JD024357.

Note the need to test for field significance as well as t-tests.

Reply: There is a long history of controlling for multiple testing in statistics including the Bonferoni (1936) correction used in this paper and extensions including, (Benjamini & Hochberg, 1995; Holm, 1979), and the references listed in the papers provided by the reviewer above. Essentially two main principles have emerged, controlling for family-wise-error-rate (FWER) and false-discovery-rate (FDR). FWER tests (Bonferoni & Holm and others) aim to ensure that the probability for one or more false detections remains at the desired significance level (e.g. 5%), while FDR tests (Benjamini & Hochberg and others) try to ensure that the proportion of false detections is <5% of the total number of detections. As such the FWER test used in our paper is the most conservative multi- comparison correction and, as pointed out in Wilks, this test might be somewhat stricter than needed.

At lines 171-174 we mentioned the Bonferoni correction. We now added sentences (emphasized below in bold) to clarify and provide additional references:
Differences in temperatures and heat fluxes between the L and R simulations are tested using a student's t-test with Bonferoni (1936) correction for multiple testing. ***Accounting for multiple tests is important to reduce the risk of incorrect conclusions when testing for effects***

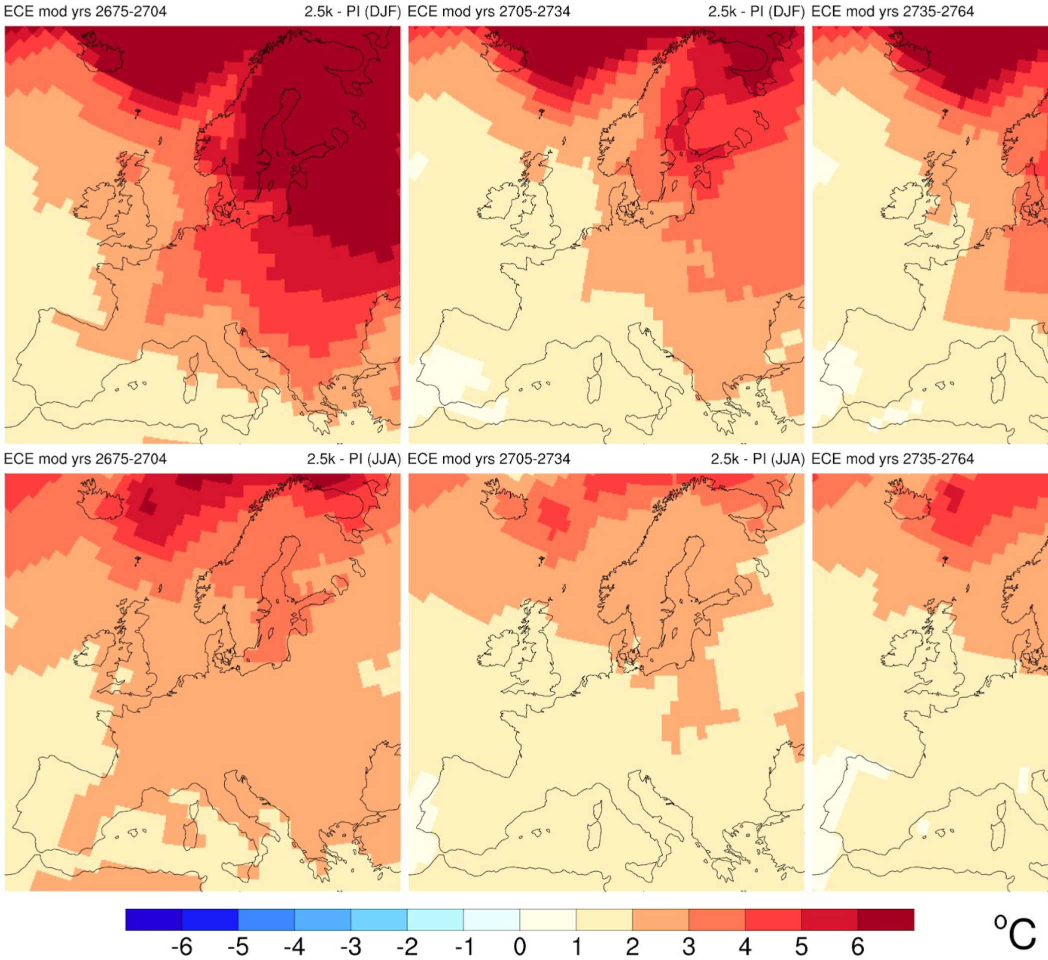
across all grid cells (Wilks, 2016). *The Bonferoni procedure has a 5 % family-wise error rate* ***(FWER)****, i.e. the probability of one or more false positives among all grid cells is 5 % instead of the 5 % false positive rate for each grid cell obtained when no correction is applied.* ***An alternative would be to use a procedure that corrects for the fraction of false positives, the false discovery rate (FDR) (e.g. Benjamini & Hochberg, 1995).*** **However,** ***FWER tests are more conservative than the FDR alternatives.***

Comment: For your paper to be robust, you have to run ensembles. I know this is not likely what you want to hear but you are only running 30 years at 50 km for one region and this is really pretty cheap so it is not overly challenging. You can then apply robust statistical testing to the results to determine whether you have a real signal from LULCC or whether you are actually finding something that is effectively determined by the parameters, or the boundary conditions, you have chosen to use. You should be able to chose multiple 30 year periods from EC-Earth and thoroughly test the dependency of your results on boundary conditions.

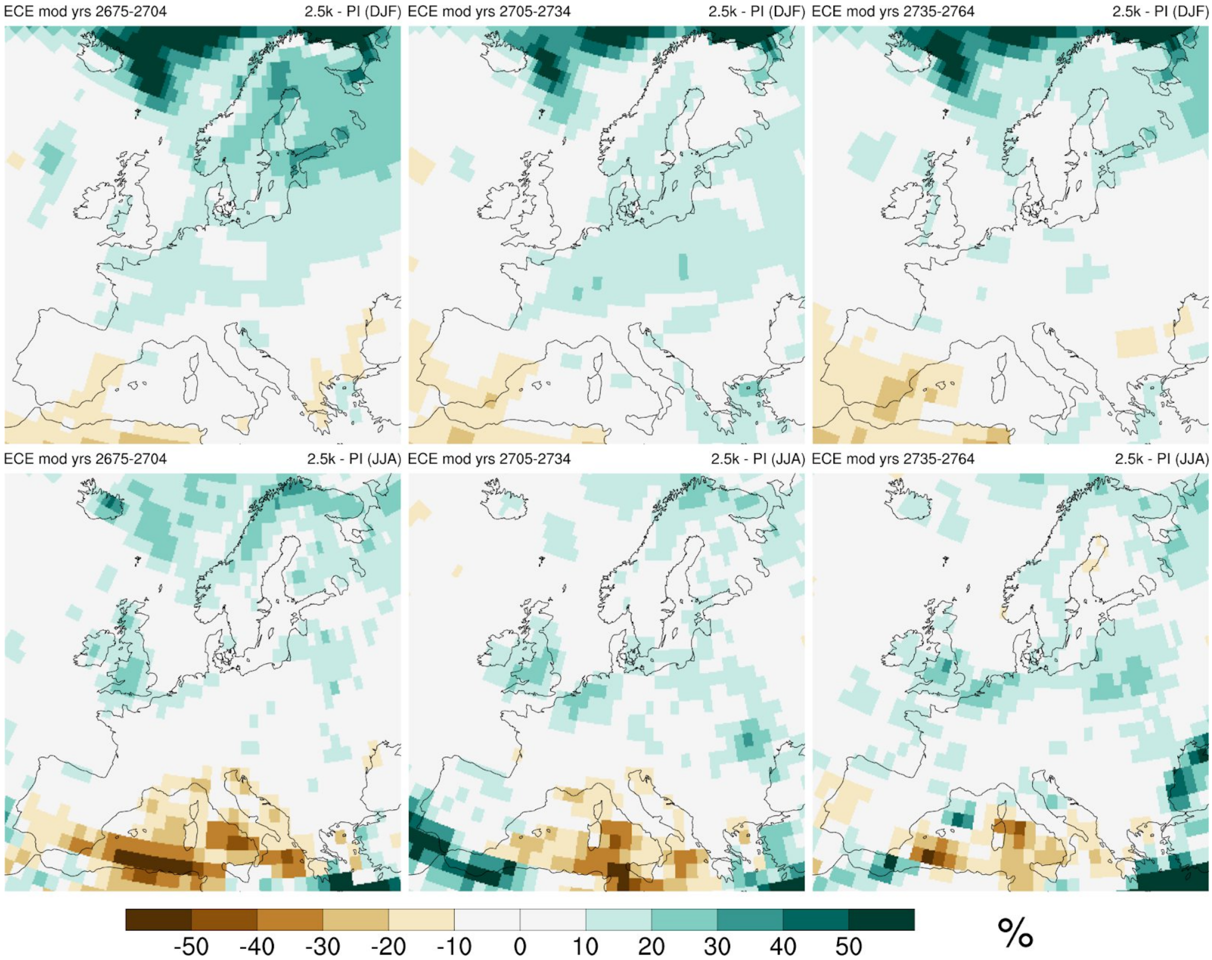
All this is going to do is change the patterns and the magnitudes of the changes in Figure 4, almost certainly remove the changes in Figure 5, except for a few places and these may be below field significance, and put uncertainty estimates on the lines in Figure 9. I bet you will find that the lines in Figure 9 are mostly statistically insignificantly different, but I am happy to be proven wrong. I also bet a great deal of hte pattern in Figure 4 is driven by the boundary conditions and if you used a different 30 year period you would get very different results. As a reviewer, I want to know if the results are robust, or are simply a reflection of the methodology. I am worried that a lot of the patterns and magnitudes you find are the result of your specifc choices on methodology and could therefore be atypical - and for me at least this needs to be resolved.

Reply: It would of course be ideal to have more ensemble members. There are, however, practical limitations that reduce the number of simulations possible. These concern availability of necessary data and costs of computer resources. We explain these limitations below and argue that, given the limitations, our approach is the soundest possible.

A first, major limitation is that we don't have boundary conditions from EC-Earth for any other years representing 2.5k than those we used in our simulations. Saving boundary data is expensive. This means that it is unfortunately impossible to downscale other time slices. We have analysed three adjacent 30-year periods in EC-Earth to estimate the variability in EC-Earth at 2.5 ka (figures are attached to this comment). As expected, the results are different but they show some common features.Therefore, we conclude that the difference between 2.5ka and PI is robust, at least in terms of the general conclusions we draw from the results. We now acknowledge this issue by adding a new section 4.1 about "Differences between PI and 2.5 ka climate" to the Discussion:
"The simulations of PI climate were made to put the 2.5 ka climate in context, although the scope of this study is the response to LULCC rather than the difference between 2.5 ka and PI climates. The natural variability in Europe is large (e.g. Deser et al., 2020), which influences the comparison. Three adjacent 30-year periods in EC-Earth, considered here to represent the studied 2.5ka climate, differ by up to 1°C. However, the difference between '2.5 ka' and PI for the three adjacent periods show common features, which indicates that the results are sufficiently robust to draw general conclusions from them. A more detailed description of the 2.5 ka climate would require ensemble simulations including more time slices. The difference in climate between 2.5 ka and PI is a result of orbital forcing and greenhouse gas concentrations (Singh et al., 2023). Since our model experiment does not include sensitivity runs changing one forcing at a time, we do not know what forcing plays the largest role."



Temperature difference (°C) between 2.5 ka and PI (T 2.5k-R – PI) for winter (DJF, top row) and summer (JJA, bottom row) as simulated by EC-Earth for three adjacent periods representing 2.5 ka.



Precipitation difference (%) between 2.5 ka and PI (P 2.5k-R – PI) for winter (DJF, top row) and summer (JJA, bottom row) as simulated by EC-Earth for three adjacent periods representing 2.5ka.

The next question is to what degree these changes in reference climate translate into significant vegetation differences, and if these differences in potential vegetation have an effect on the response to LULCC. Based on previous experience (Strandberg et al., 2022) we know that it is not worth the effort of running several DVM simulations with small differences in the driving climate (e.g. climate variability). In Strandberg et al. (2022), different representations of the 6 ka climate differ with up to 2°C. These temperature differences translate into different vegetation, e.g. a warmer climate yields more deciduous forests and less bare ground. The response to LULCC is, however, not that sensitive to the potential vegetation. Instead, the response is to a larger degree determined by the choice of climate model. We motivated this in our paper at lines 132-136. We added some text (emphasised in bold) to clarify, as follows:
"Although some studies indicate that more than one iteration between a climate model and the vegetation model are needed to reach equilibrium (e.g. Velasquez et al., 2021), and simulated land cover is sensitive to the representation of climate used and vice versa (Strandberg et al., 2022), we chose to use only one iteration. This is because the difference between reconstructed (based on pollen data) and simulated (vegetation-model based) land cover is larger than the differences between alternative vegetation-model simulations of land cover. ***Thus, differences in potential vegetation (L) between RCM run iterations would not have a large impact on the response to "actual" land cover (R). That response is to a large degree determined by the scale of land-cover changes and the climate model used (Davin et al., 2020; Strandberg et al., 2022). "***
The purpose of this study is to simulate and reconstruct 2.5k vegetation, and to investigate the effect on climate from LULCC (i.e. the difference between reconstructed and simulated land

cover). We put less emphasis on describing the 2.5k climate including its variability. Therefore, we argue that a two-RCM ensemble using one 2.5k time slice is more informative than a one-RCM ensemble using several time slices. The assumption we make is that the sensitivity to LULCC is larger between climate models used than between adjacent time slices used. We find support in the literature for this statement (e.g. Strandberg et al., 2011; Strandberg & Kjellström, 2019; Velasquez et al., 2021; Strandberg et al., 2022). The recent literature on RCMs and LULCC also makes the assumption that the response to LULCC is not significantly affected by natural variability. Instead more emphasis is put on using several models (e.g. Davin et al., 2020; Breil et al., 2020; Daloz et al., 2022; Mooney et al., 2022). We should mention that earlier pollen-based reconstructions of past land cover in Europe over the Holocene using the same method as the one applied in this study are robust, i.e. they are little influenced by differences in the model-based potential vegetation depending on differences in the climate forcing (due to the climate model used) (Pirzamanbein et al., 2020). In this case also, the differences between the pollen-based reconstructed and the simulated potential land cover is larger than the differences between the alternative pollen-based land-cover reconstructions.

Regarding Fig 9, it is used to understand how LULCC affects climate. If all other boundary conditions are kept constant, changing the vegetation will change the climate. The question is what processes are behind that effect. Answering this question would be difficult if we introduce natural variability, as we would not be able to distinguish the effect of LULCC from the effect of natural variability. It is true that the effect of LULCC might be small compared to the effect of natural variability, but the effects shown in Fig 7 & 8 are obviously caused by LULCC, which was the question to be answered in our study.

More minor comments (first number is the line number of the manuscript)

23 - using two RCMs does not allow you to study how sensitive climate models can be - you have a sample size of 2. Please rephrase.

Reply: To make that point clear we rephrase the sentence as follows:
"*Since the sensitivity to LULCC is dependent on the choice of climate model, we also use two RCMs.*"

25 - your 2-4C change in temperature is very large - it is of order the magnitude of the change people find from deforesting the Amazon. I'm not saying this is wrong, but it is a significant change that needs careful evaluation. Line 60 points to changes more like 0.5C which is less confronting.

Reply: The 2-4°C at line 25 refers to the difference between PI and 2.5ka climates, while the 0.5°C at line 60 refers to the effect of LULCC. It is expected that the difference between these two time periods with different climates is larger than the effect of LULCC.

164 - 50 km resolution is very coarse ... there is a very large potential benefit of 50 km - its computationally relatively cheap whcih allows multiple ensembles.

Reply: See reply above.

335 - this is confusing. You are right that 2 m temperatures are diagnostic, but I do not know what you mean by "surface temperature". If it is the lowest model air temperature then this is prognostic and comparable if it is at the same height across the models. Can you clarify?

Reply: With ‘surface temperature’ we refer to the variable ‘ts’ which, according to the CF convention has the standard name ‘surface_temperature’. This literally means temperature at the surface, as opposed to e.g. near surface temperature at 2 m, or at the lowest pressure or model levels.

390 - this paragraph reasonably focusses on albedo, but roughness length is also importnat - there is literature that claims differences in Z0 dominate the changes in albedo. I do not agree with this - I suspect it is an artifcact of the PBL schemes but you might not want to entirely ignore Z0.

Reply: Thanks for reminding us about this. We added the following text to the end of this section:
“The roughness length (Z0) is also changed by LULCC, therefore differences in Z0 could potentially explain some of the climatic responses. As Z0 differences are constant over the year, the same scale of vegetation changes would give the same Z0 difference across the domain. Consequently, roughness changes can not explain why the response to LULCC is different depending on the season or the region. We consider albedo and latent heat flux as the most likely explanations because they best correlate to temperature differences.”

415 - the text around here is fine - but you sort of discuss it, but then do not take these uncertainties through int the abstract and discussion. Similarly, you are clearly aware of the uncertainties due to small ensembles (lines 498, 500) and I do note you state that you only sample some of the uncertainty space but if you read this section and then re-read your abstract you get a totally different perspective.

Reply:
In the Abstract we replace the last sentence with:
“Although the results are model dependent, the relatively strong response implies that anthropogenic land-cover changes that had occurred during the Neolithic and Bronze Age could have already affected the European climate by 2.5 ka.”

In the Conclusions, we think that this is already covered since we write:
“*In summer, the RCMs used in this study respond somewhat differently to land-cover differences, showing that the choice of land cover and climate model is important for the resulting simulated climate. Summer temperatures are strongly related to differences in heat fluxes between the atmosphere and the ground. Since the response in heat fluxes to differences in land cover depends on model physics, it is more likely that models respond differently in summer than in winter. For summer, RCA4 responds more strongly to the imposed differences in land cover than HCLIM. This explains some differences between the climate conditions simulated by RCA4 and HCLIM. It is difficult to assess which model has the most realistic response. Model performance is dependent on many other factors including, but not limited to, large-scale circulation, parameterisations and resolution. The best way to understand this model uncertainty is to use several models to try to capture the range of possible climates.”*

References
Belušić, D., Fuentes-Franco, R., Strandberg, G.and Jukimenko, A., 2019: Afforestation reduces cyclone intensity and precipitation extremes over Europe. Environ. Res. Lett. 14,https://doi.org/10.1088/1748-9326/ab23b2

M. Breil, D. Rechid, E.L. Davin, N. de Noblet-Ducoudré, E. Katragkou, R.M. Cardoso, P. Hoffmann, L.L. Jach, P.M.M. Soares, G. Sofiadis, S. Strada, G. Strandberg, M.H. Tölle, K. Warrach-Sagi (2020): The opposing effects of re/af-forestation on the diurnal temperature cycle at the surface and in the lowest atmospheric model level in the European summer, Journal of Climate, 33,21, 9159-9179, DOI:https://doi.org/10.1175/JCLI-D-19-0624.1

Daloz, A. S., Schwingshackl, C., Mooney, P., Strada, S., Rechid, D., Davin, E. L., Katragkou, E., de Noblet-Ducoudré, N., Belda, M., Halenka, T., Breil, M., Cardoso, R. M., Hoffmann, P., Lima, D. C. A., Meier, R., Soares, P. M. M., Sofiadis, G., Strandberg, G., Toelle, M. H., and Lund, M. T.: Land–atmosphere interactions in sub-polar and alpine climates in the CORDEX flagship pilot study Land Use and Climate Across Scales (LUCAS) models – Part 1: Evaluation of the snow-albedo effect , The Cryosphere, 16, 2403–2419, https://doi.org/10.5194/tc-16-2403-2022, 2022.

Davin, E. L., Rechid, D., Breil, M., Cardoso, R. M., Coppola, E., Hoffmann, P., Jach, L. L., Katragkou, E., de Noblet-Ducoudré, N., Radtke, K., Raffa, M., Soares, P. M. M., Sofiadis, G., Strada, S., Strandberg, G., Tölle, M. H., Warrach-Sagi, K., and Wulfmeyer, V.: Biogeophysical impacts of forestation in Europe: first results from the LUCAS (Land Use and Climate Across Scales) regional climate model intercomparison, Earth Syst. Dynam., 11, 183–200, https://doi.org/10.5194/esd-11-183-2020, 2020.

Mooney, P. A., Rechid, D., Davin, E. L., Katragkou, E., de Noblet-Ducoudré, N., Breil, M., Cardoso, R. M., Daloz, A. S., Hoffmann, P., Lima, D. C. A., Meier, R., Soares, P. M. M., Sofiadis, G., Strada, S., Strandberg, G., Toelle, M. H., and Lund, M. T.: Land–atmosphere interactions in sub-polar and alpine climates in the CORDEX Flagship Pilot Study Land Use and Climate Across Scales (LUCAS) models – Part 2: The role of changing vegetation, The Cryosphere, 16, 1383–1397, https://doi.org/10.5194/tc-16-1383-2022, 2022.

Pirzamanbein, B., Poska, A., and Lindström, J. (2020). Bayesian reconstruction of past land cover from pollen data: Model robustness and sensitivity to auxiliary variables. Earth and Space Science, 7, e2018EA00057. https://doi.org/10.1029/2018EA000547

Strandberg, G. and E. Kjellström, 2019: Climate Impacts from Afforestation and Deforestation in Europe.*Earth Interact.,*23, 1–27, https://doi.org/10.1175/EI-D-17-0033.1

Strandberg, G., Brandefelt, J., Kjellström, E. and Smith, B. 2011: High-resolution regional simulation of last glacial maximum climate over Europe. Tellus 63A, 107-125.DOI: 10.1111/j.1600-0870.2010.00485.x

Strandberg, G., Lindström, J., Poska, A., Zhang, Q., Fyfe, R., Githumbi, E., Kjellström, E., Mazier, F., Nielsen, A.B., Sugita, S., Trondman, A.-K., Woodbridge, J., Gaillard, M.-J., 2022: Mid-Holocene European climate revisited: New high-resolution regional climate model simulations using pollen-based land-cover, Quaternary Science Reviews, Volume 281, 107431, ISSN 0277-3791, https://doi.org/10.1016/j.quascirev.2022.107431.

Velasquez, P., Kaplan, J.O., Messmer, M., Ludwig, P., and Raible, C.C.: The role of land cover in the climate of glacial Europe. Clim. Past 17, 1161e1180. https://doi.org/10.5194/cp-17-1161-2021, 2021.

Responses to reviewer 2

Comment: In this manuscript, the authors describe a comprehensive study to simulate the climate of Europe at 2.5ka and quantify the effects of anthropogenic land cover change on during this period. Using a GCM, two RCMs, and land cover descriptions from both a vegetation model and a statistical reconstruction, the authors demonstrate that not only was European climate different at 2.5ka as compared to the latest preindustrial time, but also that anthropogenic land cover change, expressed as an increase in the amount of open land over the default scenario of potential natural land cover, led to significant changes in climate. The source of this climate change is biogeophysical feedbacks between land and atmosphere, chiefly differences in springtime albedo, and summertime evapotranspiration. Using two different RCMs having different land surface parameterizations, the authors show that there are substantial differences among models, particularly in the importance of land cover change on summertime climate. Overall, this is an excellent study, truly impressive in its comprehensiveness and transparent about its limitations. I commend the authors for their thorough work; this paper will be an important contribution to the field and a model for regional studies elsewhere in the world. I have only minor comments that should be addressed before publication.

Reply: We thank the reviewer for a thorough review, the constructive comments and the very encouraging evaluation of our study.

Comment: The one thing I really miss is some kind of discussion about the mechanisms causing the 2.5k-PI changes in climate. The results are interesting, and they are highlighted in the abstract and described in detail in section 3.2 and Figures 4-6, but there does not appear to be any further discussion of the causes of these climate changes. Are the changes in circulation caused by orbital forcing, greenhouse gases, land cover differences, or some combination of the above? With the climate models showing a stronger wintertime meridional pressure gradient and a stronger blocking high over northeastern Europe in summer (at least in the RCMs), the results are reminiscent of those presented by Mauri et al. (2014). Because, as the authors note, these simulations for 2.5ka are rather unique with most of the focus of Holocene paleoclimate modeling on 6ka, I would appreciate a few sentences in the discussion section speculating on just generally what caused the large-scale simulated climate changes between 2.5ka and PI.

Reply: This is a relevant question that we perhaps somehow failed to mention. The purpose of this study is, however, to simulate and reconstruct 2.5k vegetation, and to investigate the effect on climate from LULCC (i.e. the difference between reconstructed and simulated land cover). We put therefore less emphasis on a detailed description of the 2.5k climate. The differences in forcing are orbital forcing and greenhouse gases. Thus, the 2.5 ka must be a result of that, but since we didn’t do simulations changing only greenhouse gases and orbital forcing we can’t say which is the dominant one. We can only speculate about the reasons. Nevertheless we have added a new section 4.1 about “Differences between PI and 2.5 ka climate” to the Discussion:

"The simulations of PI climate were made to put the 2.5 ka climate in context, although the scope of this study is the response to LULCC rather than the difference between 2.5 ka and PI climates. The natural variability in Europe is large (e.g. Deser et al., 2020), which influences the comparison. Three adjacent 30-year periods in EC-Earth, considered here to represent the studied 2.5ka climate, differ by up to 1°C. However, the difference between '2.5 ka' and PI for the three adjacent periods show common features, which indicates that the results are sufficiently robust to draw general conclusions from them. A more detailed description of the 2.5 ka climate would require ensemble simulations including more time slices. The difference in climate between 2.5 ka and PI is a result of orbital forcing and greenhouse gas concentrations (Singh et al., 2023). Since our model experiment does not include sensitivity runs changing one forcing at a time, we do not know what forcing plays the largest role."

Minor comments:

Line 40: The citation to Ridgwell et al. (2003) is missing from the bibliography. Please check this and all the other cited references.

Reply: The reference is added to the list. We also have cross-checked all references.

Lines 52-54: For relevant studies on the biogeophysical effects of anthropogenic land cover change on the paleoclimate of Europe, please also consider citing Dermody et al. (2012) and Gilgen et al. (2019)

Reply: We added the following sentence at line 55:

"There is also evidence of anthropogenic impact on climate via land-use changes during the Roman Period (Dermody et al., 2012; Gilgen et al., 2019)."

Line 107: As all of Earth history before the Industrial Revolution was "pre-industrial" I would request adding the qualifier "latest" or "late" before the term pre-industrial at the beginning of this line, e.g., "… latest pre-industrial (1850 CE, hereafter PI)…"

Reply: Good point. Changed as suggested.

Line 159-160: Just to be clear, the LPJ-GUESS simulated land cover in 1850 CE used as the boundary condition for the EC-Earth simulations does not include anthropogenic land cover change? Strandberg et al. (2014) showed large differences in the 1850 CE climate of Europe comparing simulations using potential natural vegetation to those where a land use mask was imposed; what are the implications of using potential vegetation as the baseline boundary condition for the climate simulations against which all the other simulations are compared?

Reply: This is indeed a bit unclear. We clarify it by revising the text as follows (new text in bold):

*"The land cover used in the PI and 2.5 ka simulations is prescribed using simulated **potential** land cover for 1850 CE from LPJ-GUESS implemented offline. **The quasi-equilibrium representing the 2.5 ka climate conditions is reached after 200 years of simulation. Note that this LPJ-GUESS simulation is not part of the model runs performed specifically for this study (for more details, see Zhang et al., 2021). Because the choice of vegetation plays a role in the resulting simulated PI (1850 CE) and 2.5ka climates, it will in turn have an impact on the comparison between PI and 2.5 ka climates. However, the expected climate response of using 'actual' rather than potential PI vegetation in the EC-Earth simulations is much smaller than the difference in climate between PI and 2.5 ka. Strandberg et al. (2014) show a response to anthropogenic land cover (instead of potential land cover) at 0.2 ka of maximum +0.5°C. Therefore, we assumed that the choice of PI vegetation (either potential or 'actual') in the EC-Earth simulations should not affect the simulated 2.5 ka vegetation or climate to such a degree that it would significantly influence the difference in climate between 2.5 ka and PI. Moreover, given that the major focus of our paper is on the difference in the 2.5 ka simulated regional climate when either potential vegetation or 'actual pollen-based vegetation' is used in the RCMs runs, we considered that the use of the same potential land-cover in the EC-Earth climate runs for both PI and 2.5 ka would have little impact on the boundary conditions used for the RCMs runs at 2.5 ka BP.**"*

In order to avoid misunderstandings we have also improved Figure 2 and Table 1.

Line 201: Check the spelling of "Krumhar[d]t"

Reply: Corrected

Line 284-285: Here is says that EC-Earth uses "prescribed land cover for 1850 CE", is this potential-natural, or is there a land use mask imposed? As noted in the comment above, it is unclear.

Reply: It is potential vegetation (see response above). We therefore revised the sentence as follows:

*"Note however that the EC-Earth simulations do not use the land-cover data R for 2.5 ka, but prescribed **potential** land cover for 1850 CE (see Methods for more details).***"**

Paragraph starting on line 288: As noted in the general comments above, it would be great to have some explanation of the drivers of these differences in climate.

Reply: See reply above.

Lines 386-387: (a) It is noted that LPJ-GUESS simulates “too large abundances of Picea” in northern Europe at 2.5ka. On what basis is this statement made? A citation or reference back to an earlier section of the manuscript would be helpful here.

(b) It is well known that LPJ-type DVMs tend to simulate too much woody cover at the expense of herbaceous and bare ground overall and globally. The authors note that wetlands and other edaphic controls on vegetation cover, along with processes such as migrational lag are a limitation to the DVM simulations. Could the over-simulation of tree cover in boreal Europe be part of the reason why there are relatively large anthropogenic land-cover induced climate anomalies in Northern Europe, particularly in the Scandes Mountains?

Reply (a): The statement on *Picea* indeed needs references.

We are now providing two references on the *Picea* issue in an additional sentence (highlighted below in bold) after the statement, and add a statement and reference on total tree cover in LPJ-GUESS simulations as follows:

“Lower ET values in R than in L land cover in most of northern Europe at 2.5 k BP are partly due to overly large abundances of Picea in the LPJ-GUESS simulation. ***The latter is known from earlier comparisons between LPJ-GUESS simulated vegetation and records of pollen accumulation rates (Miller et al., 2008) or pollen-based REVEALS reconstructions of plant cover (Marquer et al., 2017) over the Holocene.”***

Reply (b): We are not aware of strong published evidence of a general overestimation of tree cover by LPJ-GUESS in “boreal Europe”. The latest study comparing LPJ-GUESS-simulated tree cover with pollen-based REVEALS estimates of tree cover (Dallmeyer et al., 2023, preprint) shows that LPJ-GUESS mean total tree cover (MTC) is in good agreement with REVEALS MTC in Boreal Europe at 2 k BP when standard deviation is considered. The disagreement between LPJ-GUESS and REVEALS estimates is found instead in the relationship between evergreen and deciduous tree cover where LPJ-GUESS simulates a larger cover of evergreen trees than the REVEALS estimated cover (seen in Fig. S1 for e.g. the Scandes and other regions of northern Europe). Note that in mountainous regions such as the Scandes, larger REVEALS cover than LPJ cover of open land (Fig. S1) does not necessarily imply that the REVEALS open land represents anthropogenic land-cover change, it can also be due to LPJ-GUESS underestimating natural herb vegetation at higher altitudes.

In order to clarify this point, we have added the following sentence:

“However, the latest comparison between LPJ-GUESS and REVEALS total tree cover in boreal Europe shows that LPJ-GUESS mean total tree cover (MTTC) is in good agreement with REVEALS MTTC in boreal Europe at 2 k BP when standard deviation is considered (Dallmeyer et al., 2023). At a subcontinental spatial scale and for PI, MTTC is larger in LPJ-GUESS simulations than in REVEALS estimates in the British Isles only, and it is lower in LPJ-GUESS than REVEALS in northernmost Europe and the alpine

region. In continental and boreal Europe, the difference in MTTC is insignificant when the standard deviations are considered. The largest disagreement between LPJGUESS and REVEALS estimates in large parts of northern Europe is found instead in the relationship between evergreen and deciduous tree cover where LPJGUESS simulates larger cover of evergreen trees than their REVEALS estimated cover (Fig. A1; Marquer et al., 2017). "

It is also true that an overrepresentation of trees in northern European mountains (simulated vegetation, L) would imply higher temperatures in winter and /or spring compared to reconstructed vegetation (R) due to the albedo effect, and cooler in summer due to the evapotranspiration effect. If the DVM would simulate less forest it would decrease the difference between L and R, which in turn would result in a smaller response of the climate to the difference in land cover. However, as mentioned above, we have no evidence that the total tree cover is significantly overestimated by LPJ-GUESS, but the evergreen trees are overestimated in parts of central and boreal Europe.

Comment: Along these lines, it would be interesting to compare the simulated PI land cover used as the boundary condition with an independent assessment of the potential natural vegetation of Europe, e.g., by using the generalized version of the Map of the Natural Vegetation of Europe (Bohn et al., 2000/2003). Even a qualitative comparison between the LPJ-GUESS simulated vegetation and the Bohn et al. map would be illustrative in terms of understanding where LPJ-GUESS simulates greater than expected tree cover. This could lead to a better understanding of how biases in the baseline climate simulation affects the subsequent assessments of climate feedbacks.

Reply: LPJ-GUESS simulates potential land cover (L) only for 2.5k and uses EC-Earth climate at 2.5k as forcing. We did not simulate PI land cover with LPJ-GUESS for this paper. The comparison suggested by the reviewer (with the purpose to understand where LPJ-GUESS simulates greater than expected tree cover at PI) would require that LPJ-GUESS is run with EC-Earth climate at PI. This would provide insights for PI only. As mentioned above, the best study on this issue so far is the study by Dallmeyer et al. (2023) that compares DVMs (LPJ-GUESS and JSBach) simulation results with pollen-based REVEALS reconstructions of plant cover in Europe for several time windows of the Holocene. This study shows that differences between LPJ-GUESS-simulated and REVEALS estimated tree cover vary through time. At PI the mean tree cover (MTC) is larger in LPJ-GUESS simulations than in REVEALS estimates only in the British Isles and lower in LPJ-GUESS than REVEALS in northernmost Europe and the alpine region, while the difference in MTC between the two is not different from zero in the other European regions when the standard deviations are considered.

Line 403, Figure 8: Why is there a small negative winter albedo anomaly in southern Europe, particularly in the southeast? Some comment on this would be interesting.

Reply: These negative anomalies are really small, in the order of 0.01 or less. We consider them to show no change (i.e. anomalies not different from zero), as do the positive anomalies in the same order. We added the following in bold:

"This agrees with earlier studies of idealised land-cover changes in RCMs showing that albedo explains a large part of the temperature signal in winter in Scandinavia (Strandberg & Kjellström, 2019; Davin et al., 2020). ***Note that differences in albedo can be small, but still significant, because the variability is small. Some of the differences in the range of ±0.1 in Figure 8a,b are < 0.05 and should be considered as not different from zero (i.e. no difference)."***

Lines 448-449: At this point it would also be worth noting that there is likely to be a spatial mismatch in comparing most paleoclimate reconstructions at point-scale with gridded climate model output. Although climate anomalies tend to have much more spatial autocorrelation than climate itself, with a 50-km grid even in the climate model output, modeled and reconstructed climate can be subject to a large spatial and elevational mismatch. That is, climate reconstructed at a site can reflect very localized conditions that would be averaged out or smoothed, or even reflecting a different level in the atmosphere when choosing the nearest model grid-box and comparing that to a site.

Reply: A very valid point. We add the following sentences, marked in bold:

"There is also a problem with the spatial scale of the reconstructed versus simulated climate. Proxy data may represent local climate conditions that do not necessarily match the spatial scale of the model simulated climate. *Making an in-depth model-proxy comparison would require an assessment of each data point (what they represent, uncertainty ranges etc.), which is outside the scope of this paper.* ***Future studies could improve the model-proxy comparison by e.g. compensating for differences in altitude between model and proxy data (e.g. Strandberg et al., 2011) "***

Lines 459-460: Along the lines of the previous comment, was any attempt made to spatially downscale the GCM and/or RCM output to better match the site conditions, especially in the vertical? If not, the authors could at least speculate that this could be possible in future studies and be a priority for more detailed work comparing modeled and reconstructed paleoclimate anomalies.

Reply: We did not attempt to make such corrections (see our response above). The proxy comparisons should be seen only as a complement providing some indication, but it is not part of the major scope of our study. We added a sentence to suggest future improvements (see reply above).

Line 490: Can the authors say anything about the RCM sensitivity to uncertainty in the boundary conditions? It would be highly desirable to see ensembles of the RCM simulations in order to better characterize the robustness of the simulated climate anomalies and influence of anthropogenic land cover change.

Reply: We are not sure to interpret correctly what is meant by the reviewer's "boundary conditions". We assume that the reviewer refers to the land cover. We did not produce such ensembles, but instead refer to Strandberg et al. (2022). There it was shown that it is not worth the effort of running several DVM simulations with small differences in the driving climate (e.g. climate variability). In Strandberg et al. (2022), different representations of the 6 ka climate differ with up to 2°C. These temperature differences translate into different vegetation, e.g. a warmer climate yields more deciduous forests and less bare ground. The response to LULCC is, however, not that sensitive to the potential vegetation. Instead, the response is to a larger degree determined by the choice of climate model. We added some text (emphasised in bold) to clarify, as follows:

"Although some studies indicate that more than one iteration between a climate model and the vegetation model are needed to reach equilibrium (e.g. Velasquez et al., 2021), and that simulated land cover is sensitive to the representation of climate used and vice versa (Strandberg et al., 2022), we chose to use only one iteration. This is because the difference between reconstructed ***(based on pollen data)*** *and simulated* ***(vegetation-model based)*** *land cover is larger than the differences between alternative vegetation-model simulations of land cover.* ***Thus, differences in potential vegetation (L) between RCM run iterations would not have a large impact on the response to "actual" land cover (R). That response is to a large degree determined by the scale of land-cover changes and the climate model used (Davin et al., 2020; Strandberg et al., 2022)."***

Lines 515-516: Again, it would be great to have some more comment here about why simulated 2.5ka climate was warmer than that for the PI, particularly in winter, in terms of some kind of dynamical explanation.

Reply: As said in a comment above, we can only speculate about this. We add the bold text to the first sentence:

*"The models simulate a 2.5 ka climate that was warmer than the PI climate****, as a result of different orbital forcing, and feedbacks."***

References:

Bohn, U., Neuhäusl, R., Gollub, G., Hettwer, C., Neuhäuslová, Z., Raus, T., Schlüter, H., & Weber, H. (2000/2003). Karte der natürlichen Vegetation Europas / Map of the Natural Vegetation of Europe. Maßstab / Scale 1 : 2 500 000. Münster: Landwirtschaftsverlag.

Dallmeyer, A., Poska, A., Marquer, L., Seim, A., and Gaillard-Lemdahl, M.-J.: The challenge of comparing pollen-based quantitative vegetation reconstructions with outputs from vegetation models

– a European perspective, Clim. Past Discuss. [preprint], https://doi.org/10.5194/cp-2023-16, in review, 2023.

Dermody, B. J., de Boer, H. J., Bierkens, M. F. P., Weber, S. L., Wassen, M. J., & Dekker, S. C. (2012). A seesaw in Mediterranean precipitation during the Roman Period linked to millennial-scale changes in the North Atlantic. Climate of the Past, 8(2), 637-651. doi:10.5194/cp-8-637-2012

Gilgen, A., Wilkenskjeld, S., Kaplan, J. O., Kühn, T., & Lohmann, U. (2019). Effects of land use and anthropogenic aerosol emissions in the Roman Empire. Climate of the Past, 15(5), 1885-1911. doi:10.5194/cp-15-1885-2019

Marquer, L., Gaillard, M.-J., Sugita, S., Poska, A., Trondman, A.-K., Mazier, F., Nielsen, A. B., Fyfe, R. M., Jönsson, A. M., Smith, B., Kaplan, J. O., Alenius, T., Birks, H. J. B., Bjune, A. E., Christiansen, J., Dodson, J., Edwards, K. J., Giesecke, T., Herzschuh, U., Kangur, M., Koff, T., Latałowa, M., Lechterbeck, J., Olofsson, J., and Seppä, H.: Quantifying the effects of land use and climate on Holocene vegetation in Europe, Quaternary Science Reviews, 171, 20–37, https://doi.org/10.1016/j.quascirev.2017.07.001, 2017.

Mauri, A., Davis, B. A. S., Collins, P. M., & Kaplan, J. O. (2014). The influence of atmospheric circulation on the mid-Holocene climate of Europe: a data-model comparison. Climate of the Past, 10(5), 1925-1938. doi:10.5194/cp-10-1925-2014

Paul A. Miller, Thomas Giesecke, Thomas Hickler, Richard H. W. Bradshaw, Benjamin Smith, Heikki Seppä, Paul J. Valdes and Martin T. Sykes: Exploring climatic and biotic controls on Holocene vegetation change in Fennoscandia. Journal of Ecology 2008, 96, 247 –259 doi: 10.1111/j.1365-2745.2007.01342.x;

Singh, R., Tsigaridis, K., LeGrande, A. N., Ludlow, F., and Manning, J. G.: Investigating hydroclimatic impacts of the 168–158 BCE volcanic quartet and their relevance to the Nile River basin and Egyptian history, Clim. Past, 19, 249–275, https://doi.org/10.5194/cp-19-249-2023, 2023.

Strandberg, G., Kjellstrom, E., Poska, A., Wagner, S., Gaillard, M. J., Trondman, A. K., Mauri, A., Davis, B. A. S., Kaplan, J. O., Birks, H. J. B., Bjune, A. E., Fyfe, R., Giesecke, T., Kalnina, L., Kangur, M., van der Knaap, W. O., Kokfelt, U., Kunes, P., Latalowa, M., Marquer, L., Mazier, F., Nielsen, A. B., Smith, B., Seppa, H., & Sugita, S. (2014). Regional climate model simulations for Europe at 6 and 0.2 k BP: sensitivity to changes in anthropogenic deforestation. Climate of the Past, 10(2), 661-680. doi:10.5194/cp-10-661-2014