Comment:

I do not fully agree with some of your responses on this manuscript but I am going to conclude that this amounts to differences between communities. I do not think your methods for statistical testing are really ideal but I won't labour that point.

The one things I do want to hold firm on is the ensembles question. I simply do not think you can waive this away with “It would of course be ideal to have more ensemble members”. Of course you will argue that your results are robust, but my assertion is that you cannot know this without ensembles. As a reviewer I assert that you are mistaken that your results are robust, and different ensemble members could lead to different conclusions. In my view, you need to demonstrate (not claim) that I am wrong. In short, my view is that ensembles are a necessity and you risk publishing a flawed paper if you restrict yourself here.

I am willing to accept the issue with a limited experimental design linked to the lack of boundary conditions. I would not ask you to re-run to create more boundary conditions. However, there are three ways to perturb a climate models ensemble – initial conditions, parameterizations and parameters. You can vary initial conditions. You can vary parameters – roughness, albedo, moisture holding capacity, and so on. You could vary parameterisations although I accept this is challenging. So, you could test the robustness of your conclusions to variations in two major sets of uncertainty. I do note your defence based on Strandberg – but this is one (good) line of defence ratehr depending on your own work! There are alternative arguments that highlight large differences depending on the methodology.

Ultimately, based on a quick review of the paleoclimate literature I note your methodology appears to be quite traditional and you can call on many earlier papers that provide a defence of your methods. However, alternatively, from a climate modelling and land surface science perspective there are many papers that highlight the risks of using single member ensembles. I would reject this paper on the grounds that in my view you are at very high risk of publishing a misleading paper, and I would challenge you to prove the contrary. However, an alternative view would be that you are using a method that is not uncommon in you field. On one side it is harsh of me to hold you to a higher than normal standard. On the other side, the field you work in really should update its methods and that has to start somewhere.

Reply:

We thank referee 1 for this new comment. We understand the referee’s concern about robustness of our results and worries that we might publish a flawed and misleading paper. We agree that we should further emphasize the uncertainties in both the discussion and conclusions of our paper. We therefore provide further clarifications in this response and suggest additional revisions to those we already made after the first round of comments.

To start with we wish to recall the aim and constraints of our study. We are testing the null hypothesis that anthropogenic land-cover change (i.e., land-cover change due to land use, LULCC) at 2.5 ka BP (Bronze Age in Europe) did not influence climate by using regional climate models (RCMs) and the best possible reconstruction of plant cover (land cover) to date. The scientific view on biogeophysical forcings from historical deforestation using Global Climate Models (Earth System Models, ECMs) is well formulated in the conclusions and prospects of the large study by de Noblet-Coudré et al. (2012): “The appropriate question is not whether anthropogenic deforestation has a globally averaged significant impact but is rather whether it has an impact on regions that have undergone intensive deforestation”. In this context, our study is new and appropriate. The current understanding of biogeophysical forcing from land-cover change on regional climate is still limited in comparison with the large-scale carbon cycle feedbacks (Jackson, 2008, Gaillard et al., 2015). This is due to the fact that studies on biogeophysical forcing requires the use of RCMs (Strandberg et al., 2014; Wramneby et al., 2010) and a detailed representation of past vegetation cover (Wramneby et al., 2010; Smith et al., 2001) as, in turn, biogeophysical forcing is operating on a regional spatial scale and influences primarily the regional climate via climate-land cover interactions at the detailed land-cover scale. The climate-land cover interactions and the processes governing them are described and represented in various ways in different RCMs. In the literature dealing with land use as a climate forcing, it has often been pointed out that the largest uncertainties in such studies reside in how the models respond to LULCC (e.g. Davin et al 2020). Therefore, ensemble RCM simulations are recommended in the same way as is currently the rule for ESM simulations studies (e.g., by the PMIP community).

In this context, we wish to clarify the use of the terms “climate ensemble simulation” and “climate model sensitivity tests” in our paper. Of the five major types of climate ensembles applied today (e.g., The Multi-Model Ensemble Approach. AR4 WGI Chapter 10: Global Climate Projections), we used both a “multi-model ensemble” and “a forcing ensemble”. The “perturbed physics ensemble” and “Initial condition ensemble” usually comprise sensitivity tests where individual climate models are tested for their sensitivity to different parameter settings (physical but also other parameters related to e.g., atmospheric chemistry, carbon cycle, etc.). The different types of climate ensembles are often produced in individual studies and presented in separate papers. In our study, both of our ensembles are small as they comprise only two regional climate models (RCMs) and two alternative land-cover forcing i.e., natural vegetation (climate-induced) and actual vegetation (climate- and human-induced), respectively.

The two-model ensemble allows us to see how much differences in model physics impact on the two RCMs responses to land-cover forcing (i.e., natural versus actual vegetation (see definitions above)). A follow-up study would be to investigate how each model responds to land-cover forcings, e.g. whether the effect mostly comes from differences in roughness, albedo, heat exchange, etc. Such sensitivity studies are often designed to analyse responses in one model (e.g., Belusic et al., 2019; Breil et al 2023).

We did not use large ensembles in our study for the following reasons, mostly imposed by time and financial constraints. There are inherent differences between palaeo studies (dealing with several centuries to millennia in the past) and ‘traditional’ studies (usually involving one to two centuries). A ‘traditional’ climate model study usually builds on community efforts (e.g., the CMIP or Cordex ensembles) providing a much larger data base than could possibly be produced within a single study performed within a specific research project with a limited budget. It is reasonable to expect multi-member ensemble sizes of 10-20 (or more) members in climate-change studies of the 21st century. There are only a few periods of special interest in the past (e.g. the Holocene climate optimum at 6ka) for which ensemble sizes comparable to ‘traditional’ studies’ are available. When targeting time periods that have not been/are not currently studied intensively by community efforts (e.g., 2.5 ka in our study), the size of model ensembles is very limited. Our choice of the 2.5 ka time window is motivated by the research question of our project based on the latest advances in Holocene land-cover reconstructions for Europe revealing a major anthropogenic continental-scale land-cover change. We are fully aware of the uncertainties related with the use of only two RCMs and emphasize this issue in the manuscript. Nevertheless, our study is unprecedented within the field of regional palaeo climate modelling and will hopefully be followed by similar studies using other RCMs in order to progressively build up larger multi-model ensembles for 2.5 ka and the land-cover forcing in the European Bronze Age.

*

Regarding "perturbed physics ensembles", the use of two RCMs with different physics does produce a two-member ensemble. Again, it is a small ensemble, but our results clearly show that the largest uncertainties are related to the models' physics and parameterisation (compared to the uncertainties related to a small multi-model ensemble), i.e., the two RCMs do not respond to LULCC in the same way (Figures 7 & 8). The differences between RC4 and HCLIM are thoroughly described and discussed in the manuscript. To better emphasize these uncertainties, we revised the wording in order to avoid any misunderstandings.   Abstract, lines 29-34:

"The results also **suggest** that LULCC at 2.5 ka impacted the climate in parts of Europe. /…/ **Although the results are model dependent**, the relatively strong response **implies** that anthropogenic land-cover changes /…/ **could have** affected the European climate by 2.5 ka. "

*

Regarding perturbing parameters, it is of course of interest to investigate how sensitive a RCM is to alterations in e.g. albedo, roughness length, water holding capacity. Changing parameters would obviously provide different results. However, such a study would not help us to decrease the uncertainties in our results on the response of the two RCMs to different vegetation cover under 2.5 ka conditions in terms of climate and vegetation. As argued above, the uncertainties in our results are due primarily to differences in the RCMs' physics.

*

Relating to the suggestion to change initial conditions this would indeed be a possibility to infer some natural variability into an ensemble. However, as we can't change the initial conditions in the global model we would still have the same lateral boundary conditions for the simulations and therefore only sample part of the natural variability without considering forcing from the large scale. Compared to the effect of boundary conditions, the effect of initial conditions disappears rather quickly. Rummukainen (2010) writes: "Compared to regional weather forecasting in which initial conditions are an important source of the uncertainty and also to shorter RCM simulations (months to seasons), in longer RCM simulations lateral boundary conditions are a more marked uncertainty source."

We rely on Velasquez et al. (2021) and Strandberg et al. (2022) because these studies are the most relevant ones to date in the context of our study. We refer to Strandberg et al. (2022) because that study uses the same RCMs than we do in the present study. We therefore did not find it useful to repeat the same calculations, that would lead to similar conclusions. Our study (and paper) focuses on the RCMs' response to different land (vegetation) cover, natural (LPJ-GUESS simulated) versus actual (pollen-based) vegetation, see above. The response to this difference in land cover is significantly larger than the response to small variations in the simulated natural vegetation.

Below we list revisions made to emphasise uncertainties better than we did in the first version. The most important passages are marked in bold.

*

Regarding changing the potential vegetation as an initial condition, we write (lines 137-143):

"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 (Strandberg et al., 2022)**. Thus, differences in potential vegetation (L) between RCM run iterations would probably 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)."

*

Regarding model physics and response, we write:

Lines 184-187:

"In this study, /…/RCA4/…/ and /…/HCLIM/…/ are run /…/ across Europe /…/. They have different model physics and cannot be expected to respond to LULCC in the same way."

Lines 364-372:

"HCLIM simulates larger R-L temperature differences than RCA4 in large parts of Scandinavia and western Europe, suggesting a stronger sensitivity of HCLIM than RCA4 to the difference in land-cover data./…/ In summer, RCA4 responds strongly to the inferred land-cover differences./…/ HCLIM, on the other hand, shows very small R-L temperature differences."

Lines 451-460):
"The areas with the largest R-L temperature differences in summer (Fig. 7c, d, g, h) correspond to the areas with the largest significant differences in latent heat flux as simulated by RCA4/../ In contrast, HCLIM shows few significant temperature differences, as a result of very small differences in latent heat flux (Fig. 8d). Davin et al. (2020) demonstrated that RCMs could disagree in their temperature response to idealised changes in land cover in summer. There is not a direct relationship between differences in land cover and simulated changes in EVT. EVT is determined in the models by the magnitude of land-cover differences, as well as the absolute amount and type of vegetation. **Furthermore, local conditions, such as available soil moisture, determine how much water is available for EVT. How these factors are translated into temperature responses depends on the individual models' physics and parameterisations.**"

Lines 470-473:
"**The largest discrepancies between RCA4 and HCLIM** are found in the annual cycles of EVT in WCE and IBP. In WCE, RCA4 simulates a larger amplitude and slightly higher values of EVT in the second half of the year. In IBP, it is HCLIM that simulates a larger amplitude and a longer period of high EVT. "

Lines 477-478:
"Such a pattern is seen in both RCA4 and HCLIM simulations, but it is clear that RCA4 tends to dry out more rapidly than HCLIM (Fig 9f)."
Lines 436-437:
"Even with identical forcing, the response to regional differences in land cover can differ significantly between model simulations."

Lines 441-443:

“The small response in latent heat flux and temperature in HCLIM in summer relative to that in RCA4 shows that **sensitivity to land-cover changes differs across models**, regions and seasons.”

Lines 554-558:
“**Our way of handling model uncertainty is to use two models with different model physics.** As we get a similar response as that obtained in studies of climate-land-cover interactions using idealised land-cover forcing under present climate conditions (Strandberg & Kjellström, 2019); Breil et al., 2020; Davin et al., 2020) and comparable studies of past climate (Russo et al., 2022), we ascribe most of the uncertainty associated with the response to differences in land-cover in our study to differences in the two models’ physics. **To improve the estimate of uncertainty of our results will require larger multi-model ensembles. Such ensembles are not available for 2.5 ka to date.”**

Lines 582-589:
“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 multi-model ensembles to capture the range of possible climates.**“
*
Regarding natural variability and GCM data as boundary conditions, we write (lines 375-384):
“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.”
*
Regarding misleading results, we write (lines 560-568):
“This study describes the climate and land cover 2500 years ago (at 2.5 ka) as simulated by one GCM, two RCMs, one DVM and using one reconstruction of 2.5 ka land cover based on pollen data, statistical interpolation methods and climate model results. The results **provide some insights on** how simulated climate is influenced by LULCC, and how sensitive RCMs are to differences in land cover.
The results presented here **suggest that** LULCC at 2.5 ka impacted the climate in parts of Europe. Simulations including LULCC give **up to** 1 °C higher seasonal mean temperature in parts of northern Europe in winter and **up to** 1.5 °C warmer in southern Europe in summer than simulations with potential land cover. This **relatively strong** response **suggests that** anthropogenic land-cover changes **may have** affected European climate during the Bronze Age.

This also **implies that** LULCC is important in future scenarios, especially in low-emissions scenarios where the greenhouse gas forcing is relatively small."

References

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

Breil, M., Weber, A., and Pinto, J. G.: The potential of an increased deciduous forest fraction to mitigate the effects of heat extremes in Europe, Biogeosciences Discuss. [preprint], https://doi.org/10.5194/bg-2023-18, in review, 2023.

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. Dyn. 11, 183-200, https://doi.org/10.5194/esd-11-183-2020., 2020.

Jackson RB 2008. Environ Res Lett 3, 044006.doi.org/10.1088/1748-9326/3/4/044006

De Noblet-Coudré et al. 2012. J Clim 25, 3261–3281.

Smith B et al. 2001. Global Ecol. Biogeogr. 10 (6), 621–637.

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., and Gaillard, M.-J.: Mid-Holocene European climate revisited: New high-resolution regional climate model simulations using pollen-based land-cover, Quaternary Science Reviews, https://doi.org/10.1016/j.quascirev.2022.107431., 2022.

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.

Gaillard et al. 2015**.** Causes of regional change: land use. In The BACC Author Team: Second Assessment of Climate Change for the Baltic Sea Basin. Springer, Berlin, Heidelberg, 503 pp, DOI 10.1007/978-3-319-16006

Wramneby A et al. 2010. J Geophys Res, 115, D21119, doi:10.1029/2010JD014307, 2010.