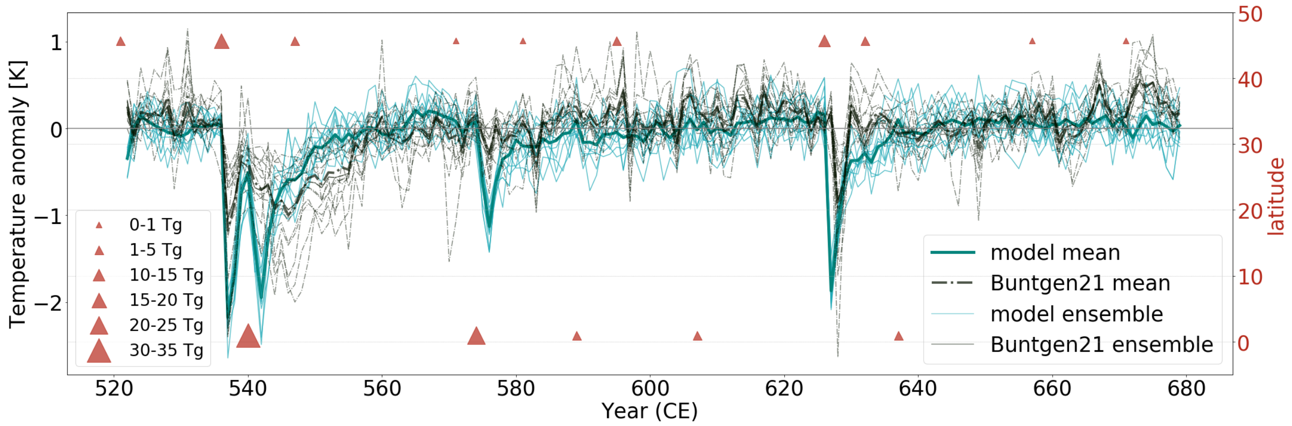
Dear editor and reviewers, thank you for the detailed review of the paper. We have compiled a point by point response going through all comments of the four reviews. Our comments are given in blue and bold text.

General reply:

We have considerably revised the manuscript by rewriting the abstract, introduction and summary and conclusion, as well as separating the discussion from the result section. We updated the references in the introduction and discussion. As suggested, we now use the Büntgen et al. (2021) tree-ring reconstruction ensemble for the model - tree-ring comparison (see figure below) and have rewritten the comparison section accordingly, focusing on the Northern Hemisphere temperature anomaly.



Reviewer 3:

This paper develops and analyzes an ensemble of climate model simulations covering the period of 4 large eruptions in the 5th and 6th century as well as the decades following. These modeling results are really very exciting because they provide brand new and much needed insight into the potential behavior of the climate system following large and important eruptions (including two closely-spaced eruptions in the 6th century) in the first millennium and the potential for new paleoclimate proxy/model comparisons of this important but still sparsely known period of the Common Era. These results will therefore be of great interest in understanding the range of responses to volcanic eruptions and relevant for both modelers and proxy paleoclimatologists.

My primary critical observation is that the manuscript is excessively descriptive. I think a stronger manuscript would result from placing the model (and the proxy-model) comparison directly in the larger context of uncertainties about first millennium climate and response to large eruptions, isolating the model behavior that is most interest to understanding this unique time period, a more complete treatment of the proxy data, and more accounting for uncertainty (the real strength of the model ensemble is the range of possibly behaviors of the ocean-atmosphere system that can be observed and used to quantify variability in the response). My major comments below are mostly about these issues, and some minor comments and suggestions follow.

Thank you for your constructive comments. We have taken your comments into account in the revised manuscript, as further answered above in the general comments and in detail below.

We have revised the manuscript so that it is less descriptive by separating the result and the discussion parts, and we elaborated on the uncertainties about first millennium climate reconstructions and model behaviour. We now use the Büntgen et al. (2021) tree-ring reconstruction ensemble for the model - tree-ring comparison, which consists of an ensemble of reconstructions, which reduces the uncertainty of individual temperature reconstructions.

Major comments:

1. Abstract and Introduction: I think the paper (and readers) would benefit from a re-writing of the abstract and a re-framing of the importance of this work.

Thank you for your suggestion. The abstract was rewritten, re-framing the importance of this work by taking your and reviewer 2' suggestion into account. For example:

'While the onset of this cold period can be clearly connected to the volcanic eruptions in 536 and 540 Common Era (CE), the duration, extent and magnitude of the cold period is uncertain.'

2. As it currently stands, one of the particularly interesting aspects of the 6th/7th century eruptions (that persistence or not of the cooling and therefore the duration of the 'Late Antique Little Ice Age') is mentioned only rather vaguely, and much of the abstract is given over to a simple description of the model phenomena.

We have reframed the abstract in a way that the motivation of the work is more clear, and we have described the results in a way that they are more interconnected.

3. A stronger abstract would set the stage for the proxy disagreements (and note the sparseness of the proxy data as exacerbating the difficulty in understanding these uncertainties) and then frame the results in terms of this as well as the importance of understanding large events like the closely spaced eruptions of the 6th century.

Thanks for your very good suggestion. We have rewritten the abstract accordingly.

4. Likewise, the Introduction would benefit from using more recent publications about the Little Ice Age as well as a more structured framing of the uncertainty and motivation for the study.

The paper by Moreno-Chamarro et al. (2017) has been added to the introduction and discussion sections. The LIA is not the focus period of our study, and we do not aim to make a direct comparison, as we do not have this period in our model simulations. We therefore have added only one newer publication on the LIA period, as we do discuss the similar mechanism behind the cooling and thus agree it is important for the discussion part.

As described in our overall reply, we have revised the manuscript extensively, with new tree-ring data comparison, a rewritten abstract, introduction and summary and conclusions, and elaborated discussion sections.

5. Initial discussion of proxy data (Lines 53 to 59). This section needs to be enhanced, as an appreciation for the uncertainties and causes of uncertainties for first millennium climate reconstructions, particularly with the resolution needed to resolve volcanic eruptions precisely, is important for the comparisons that come later in the paper.

The paper of the Pages2k consortium (2017) has been added to address the uncertainties for first millennium climate reconstructions.

6. While Ahmed et al. 2013 (which is properly PAGES 2k Network 2013) was a distillation of temperature proxy data for the last IPCC, it is superseded by a number of papers, including PAGES2k Consortium 2017 (Emile-Geay et al. 2017).

We updated the section with the paper of Pages2k consortium (2017).

7. The authors might also want to cite Esper et al. 2018 (doi is 10.1016/j.dendro.2018.06.001) which analyzes tree-ring proxy uncertainties in the early part of the last millennium (and therefore these uncertainties will be even greater in the first millennium of the Common Era). Particularly here: there is a substantial body of literature now (some of it discussed later) about the ability of different tree-ring proxy measurements to resolve or 'smear' volcanic cooling - MXD vs tree-ring width.

Thanks for your suggestion. The study has been included and is now used to address the tree-ring uncertainties and sparsity in proxy data in the first millennium in the discussion section.

8. Similarly, multiproxy approaches that mix seasons, hemispheres, or are low resolution might not resolve the volcanic signal or may require additional post-processing of model data to make an appropriate comparison. Since the source of possible uncertainties in proxy reconstruction of

6th/7th century climate is important for the comparisons that come later, I think a more thorough and up-to-date discussion is warranted here in the introduction.

Thanks for your suggestion. We have added why a multiproxy approach would not be the best to use for this study in the introduction.

9. Proxy data (Section 2.2): As above, I think a more complete and clear description of the proxy data here would be useful for later in the paper when comparisons become important.

Since we use the data from Büntgen et al. (2021) instead we rewrote the methods section accordingly, also taking up your comment. We point to the corresponding papers for the details of the tree-ring data.

10. Table 2 lists the individual proxy data that are available (this is good to have this, since the representation of tree-ring width and MXD can sometimes be subsumed when using a reconstruction), but the wording in Section 2.2 is confusing - for instance, what does 'The first four sites combined are the "NH land" compilation by Stoffel et al. (2015)' mean? Does this mean that the 4 sites listed first in the Table were also used by Stoffel? This isn't clear.

Thanks for bringing this to our attention. The first four sites were indeed used by Stoffel et al. (2015). Since we changed the tree-ring data to Büntgen et al. (2021), the table is updated accordingly and moved to the appendix.

11. By the time one arrives at Figure 5 and the associated text, it isn't clear what/which of each of these proxies is going into the comparison, so a more thorough discussion of the proxy data used and what each reconstruction in Figure 5 contains is necessary.

The tree-ring part in the results and discussion section has been rewritten according to the new ensemble reconstructions.

12. The following line says 'The data sets 135 contain a mix of tree ring width (TRW) and maximum latewood density (MXD).' and this is true, but only the NSCAN MXD data are available for the 6th and 7th century - the rest are tree-ring width and subject to the potential problems described in the following lines. Again, this section seems rather sparse and is not clearly organized, and yet limitation of the proxy data (or their particular time series properties) will become important later in the paper. Since the authors prepared this manuscript, there as been a new ensemble reconstruction of Common Era temperature (Buntgen et al. 2021, doi is 10.1038/s41467-021-23627-6) - while I realize this paper actually came out after this manuscript was submitted and the authors cannot have been expected to use these reconstructions (of course!) I would encourage them

to at least consider them (formally analytically or informally as comparison, since the early part of the LALIA is examined in the paper), at the discretion of the editor.

We agree with this comment. Thus, the new data set from Büntgen et al. (2021) has been used in the revised manuscript, and the tree-ring section has been rewritten, as mentioned above.

13. Finally, it would be worth I think mentioning why multiproxy reconstruction like PAGES2k, LMR etc. are likely to be unsuitable for this comparison and why the authors rely (and rightly I think) on tree-ring data.

Thank you for your suggestion. Why a multiproxy reconstruction like PAGES2k etc. is not suitable for this study has been added to the introduction.

14. Proxy-model comparison: This section is unclear in places and speculative without support in others; for instance, (Line 355) I'm not sure what 'The temperature anomalies from the model simulations and the 2 sigma variability range fall within the 2 sigma variability of the NH of the model simulations' means? I also find it to be too qualitative - what does 'good agreement' mean and how to measure it?

Thanks, this was a writing error, it has been corrected.

As for the qualitative 'good agreement', we compare the range of variability of the model to the range of variability of the tree-ring ensemble.

'... the reconstructed NH temperatures from the ensemble mean fall within the spread of the model simulations.'

The proxy-model comparison section was rewritten, as we now use the tree-ring reconstruction ensemble from Büntgen et al. (2021). We have taken your comment into account in the comparison, by using absolute numbers and the standard deviation.

15. In Line 362, 'More deviation is visible' is also vague. I think this section would benefit from a more straightforward and quantitative exploration of the proxy-model comparisons.

The comparison section has been rewritten according to the new Büntgen et al. (2021) tree-ring reconstruction ensemble we are now using. We have taken your comment into account in the revision, see our previous comment.

16. In Line 365, I'm not sure how something could be both 'less good ... but still remarkable'? However, also the full range of the model ensemble should really be considered in the comparison -

the 'real world' is just simply one iteration of what could have happened under different initial conditions, forcing uncertainty, feedbacks, and interactions and stochastic variability. So the comparison is not simply to the multimodel ensemble mean or even peak cooling, but taking into account the full range of ensemble variability and seeing the tree-ring data as one 'ensemble member' of possible actual and physically plausible atmosphere-ocean states.

Thank you for your suggestion. As for the above comment, we have made the comparison discussion more quantitative, by describing the comparison in terms of significance and anomaly values. In addition, the ensemble range has been taken into account in the rewritten comparison discussion.

17. Later in Line 376, the authors write that 'There is a good agreement between the tree-ring temperatures and the model temperatures after normalization' - but again this lack of precision doesn't do justice to the comparison - indeed there appears to be reasonable association for the major eruptions for NH temperature from Stoffel (including some MXD) and the NSCAN MXD, but for Alps and Altai the lag recovery is longer. So simply saying there is a good agreement masks interesting differences.

We have taken your comment into account and have elaborated the discussion of the differences between the model and the tree-ring ensemble. Since we are now using the Büntgen et al. (2021) tree-ring reconstruction ensemble and to be able to go more in detail when it comes to agreements and discrepancies, we decided to focus on the NH in the main manuscript and to move the comparison on the individual sites to the appendix.

18. In Line 381, this seems very highly speculative: 'could be due to the prescribed volcanic forcing in the model, and that the 547 eruption might have had a stronger impact on NH land than the model simulates.' - why wouldn't the same apply to Stoffel or NSCAN then? There would need to be some support for this to claim it as a source of the discrepancy.

Since we now use the Büntgen et al. (2021) tree-ring reconstruction ensemble, the focus in the main manuscript is on the Northern Hemisphere. The local tree-ring based reconstructions figures have been shifted to the appendix and the corresponding text has been rewritten accordingly. There is a lag after the 536/540 CE eruptions visible in all tree-ring reconstructions, in Nscan as well, though not as pronounced. This could be due to the MXD data that was used for this site. We have taken up your comment and elaborated the explanation of the 547 CE eruption in the discussion.

19. On Line 390, again this seems highly speculative: 'Perhaps the century long lasting cooling may be only apparent in the Alps and Altai tree-ring records, as the cooling is a local feature occurring at high altitude of the mid-latitudes.'

There are other signs that the cooling lasted longer in the Alps, for example an advance in glacier fronts. The LIA was spatially heterogeneous in duration and timing and so the same can be true for the LALIA. Records from Greenland ice-cores for the study period (Sigl et al. 2015) agree with the tree-ring records from the Alps (Büntgen et al. 2017). Interestingly, our model simulations reveal a spatial pattern with a hemispheric dipole dividing the NH response at around 45 N. Further tree ring based studies on spatial patterns are needed in the future.

We have clarified this point accordingly in the revised manuscript.

20. Again, in Line 395 the authors speculate that 'The change in hydro-climate corresponds to the soil moisture availability for the trees and thus could have impacted tree ring growth', but again this is just speculation, and indeed for the Alps, which have the longest lag at odds with the model, the 20 years summer precipitation anomaly (Figure 3c) is positive and the winter signal is mixed.

We have added the papers of Basset et al. (1964) and Müller et al. (2016) to support the theory that moisture availability could have impacted the tree-ring growth. The Alps do indeed have a slight increase in precipitation for the 20 year mean, but they also have an increase in evaporation at the same time. We find the atmospheric circulation separation interesting, and think it is worth further investigation.

21. Potentially the most parsimonious answer is that tree-ring width has a tendency to increase the 'tail' of the post-volcanic cooling and change the timing of recovery to baseline. But the authors give this only a brief mention in Line 371.

Thank you for your suggestion, we have elaborated on this in the discussion of the model tree-ring comparison.

22. Summary and Conclusion: This section is largely a restatement of the paper, but would be stronger with a distillation of the main points of the article and major conclusions.

The Summary and conclusions have been shortened and rewritten.

Additional Comments:

23. Line 12, Line 13 'land–sea'

This has been corrected.

24. Line 50: perhaps 'multidecadal cooling, as has been reconstructed by Buntgen et al. (2016).'
The sentence has been altered according to the comment:

'...multidecadal cooling, as has been reconstructed (Büntgen et al., 2016; Helama et al., 2017)...'

25. Line 55: 'Common Era'
'Common Era' has been introduced in the abstract, after which it is addressed as 'CE'.

26. Line 101-102: this sentence seems out of place? 'A common paleo-model set-up is to use 1850 pre-industrial conditions, due to model simulation limitations'.
We have changed the sentence to the following:

'In contrast to other studies, where the initial state was simply taken from a pre-industrial control simulation, our approach allows us to include the climate and forcing history of the previous decades and centuries as well as their integrative effects (e.g., Gleckler et al., 2006).'

27. Line 103: This is confusing as written - but why use standard deviations instead of the temperature deviations in the bootstrap?

We are not sure what is meant exactly by this comment. We calculated the standard deviations for the bootstrap for each of the variables, and used this to calculate the significance.

28. Figure 1: are these data from Jungclaus et al. and Toohey and Sigl? Best to include a citation with the figure caption as well.

The figure caption has been revised:

'...Global zonal mean volcanic forcing (Aerosol Optical Depth, AOD) for the study period (520-680 CE) and b) zonal mean accumulated AOD (520-680 CE) in 15 year bins from the reconstructed volcanic forcing of Toohey and Sigl (2017)...'

29. Table 1: I assume the months (January) are assigned because the actual month of the eruption is not known for these? It would be worth mentioning this (and some of the potential consequences, e.g. Stevenson et al. 2017) in the methods section as well
Thank you for the suggestion, we have taken it up and added it to the method section.

'... The eruptions are set to January in the model forcing as the actual eruption month is unknown.'

As for the discussion on eruption season, we have added the Toohey et al. (2011) paper.

30. Line 152: '2 K' - relative to which baseline? I presume the 521-680 CE mean mentioned in Line 142, but please clarify

The 2K cooling is relative to the 0-1850 CE mean, which is taken as the baseline throughout the paper. This has now been clarified in the text as follows:

'...with a peak cooling of the NH 2m air temperature of ~2 K … compared to the 0-1850 CE mean.'

31. Line 157: 'decreased for ∼20 years' - it is difficult to see this in Figure 2 because of the scale - can you provide a range of the actual return to baseline periods? Particularly since the closely-spaced 536/540 eruptions would be expected (I think) to collectively show a longer recovery time than the individual eruptions in the 570s and 626 event

The actual return periods have been added to the text:

'...reveals a maximum cooling in the first and second year after the eruption and is decreased for 20 years after the 540 CE, for 30 years after the 574 CE and for 14 years after 626 CE eruptions, much longer than the direct response of the AOD.'

32. Line 169-175: Some other more recent papers to consult (and cite as appropriate) might include Lehner et al. 2013 and Slawinska and Robock 2018

These papers have been taken into consideration and we added the paper of Lehner et al. (2013) to the discussion of ocean - sea-ice feedbacks.

33. Line 270: Perhaps useful to consult Fischer et al. 2007 (doi is 10.1029/2006GL027992, 2007) and Rao et al. 2017 (doi is 10.1002/2017GL073057) with respect to Mediterranean (and Europe) response to eruptions in historical and paleoclimate data.

Thanks for the suggestion, we have consulted these papers with respect to the Mediterranean response and included them in our discussion.

34. Figure 5 caption: 'Fennoscandia'?

The term Fennoscandia has been replaced by 'Northern Scandinavia (Nscan)'.

35. Line 372: as well as estimate of the forcing, spatial distribution of AOD anomalies, feedbacks, uncertainty in timing of the eruption (Stevenson et al.) - so, lots of potential uncertainties.

There are several sources of uncertainties, which comes with using models and proxy data. Uncertainties are described in the new discussion section of the manuscript.

36. Line 377-: I'm not sure this requires any further explanation - the models and data have different references periods and likely different means, but what is of interest is the volcanic signal, so a renormalization isn't that remarkable

Thanks for your suggestion, we have taken up your comment. We have now used 0-1850 CE as a reference period for all data series, so that it is consistent throughout the manuscript. In the new discussion we do not go in detail on this realigning of the data.

References

Bassett, J.R., 1964. Tree growth as affected by soil moisture availability. *Soil Science Society of America Journal*, *28*(3), pp.436-438.

Büntgen, U., Myglan, V.S., Ljungqvist, F.C., McCormick, M., Di Cosmo, N., Sigl, M., Jungclaus, J., Wagner, S., Krusic, P.J., Esper, J. and Kaplan, J.O., 2016. Cooling and societal change during the Late Antique Little Ice Age from 536 to around 660 AD. *Nature geoscience*, *9*(3), pp.231-236.

Büntgen, U., Myglan, V.S., Ljungqvist, F.C., McCormick, M., Di Cosmo, N., Sigl, M., Jungclaus, J., Wagner, S., Krusic, P.J., Esper, J. and Kaplan, J.O., 2017. Reply to 'Limited Late Antique cooling'. *Nature Geoscience*, *10*(4), pp.243-243.

Büntgen, U., Allen, K., Anchukaitis, K.J., Arseneault, D., Boucher, É., Bräuning, A., Chatterjee, S., Cherubini, P., Churakova, O.V., Corona, C. and Gennaretti, F., 2021. The influence of decision-making in tree ring-based climate reconstructions. *Nature communications*, *12*(1), pp.1-10.

Gleckler, P.J., AchutaRao, K., Gregory, J.M., Santer, B.D., Taylor, K.E. and Wigley, T.M.L., 2006. Krakatoa lives: The effect of volcanic eruptions on ocean heat content and thermal expansion. *Geophysical Research Letters*, *33*(17).

Lehner, F., Born, A., Raible, C.C. and Stocker, T.F., 2013. Amplified inception of European Little Ice Age by sea ice–ocean–atmosphere feedbacks. *Journal of climate*, *26*(19), pp.7586-7602.

Moreno-Chamarro, E., Zanchettin, D., Lohmann, K., Luterbacher, J. and Jungclaus, J.H., 2017. Winter amplification of the European Little Ice Age cooling by the subpolar gyre. *Scientific Reports*, *7*(1), pp.1-8.

Müller, M., Schwab, N., Schickhoff, U., Böhner, J. and Scholten, T., 2016. Soil temperature and soil moisture patterns in a Himalayan alpine treeline ecotone. *Arctic, Antarctic, and Alpine Research*, *48*(3), pp.501-521.

PAGES2k Consortium, 2017. A global multiproxy database for temperature reconstructions of the Common Era. *Scientific data*, *4*.

Sigl, M., Winstrup, M., McConnell, J.R., Welten, K.C., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D. and Fischer, H., 2015. Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature*, *523*(7562), pp.543-549.

Stoffel, M., Khodri, M., Corona, C., Guillet, S., Poulain, V., Bekki, S., Guiot, J., Luckman, B.H., Oppenheimer, C., Lebas, N. and Beniston, M., 2015. Estimates of volcanic-induced cooling in the Northern Hemisphere over the past 1,500 years. *Nature Geoscience*, *8*(10), pp.784-788.

Toohey, M., Krüger, K., Niemeier, U. and Timmreck, C., 2011. The influence of eruption season on the global aerosol evolution and radiative impact of tropical volcanic eruptions. *Atmospheric Chemistry and Physics*, *11*(23), pp.12351-12367.

Toohey, M. and Sigl, M., 2017. Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE. *Earth System Science Data*, *9*(2), pp.809-831.