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

Reviewer 1

I think the authors have addressed the issues raised in previous versions. I still find the conclusions as a whole not that much far-reaching, but the manuscript is useful as it presents a thorough analysis of an ensemble of simulations. This can be helpful to identify the reasons for the long persistence of cooling found in the different proxy data sets.

Thank you for your constructive comments. We have taken your comments into account in this round of revisions, as answered in detail below.

I have a few remarks regarding the formulation of some sentences:

1. Line 110 'To get a significant sample' The word significant is here not clear. A sample cannot be significant in the statistical sense - only a test statistics can be . I think the authors mean to ' increase the power of the test'.

Thanks, we have corrected this according to your comment.

2. Line 172 'Towards the end of the simulation, period the ensemble shows a larger spread in the ocean heat content than at the beginning of the simulations.' Delete comma after simulation.

Thanks, corrected.

3. Line 377 'The study from Zhong et al. (2011) about the onset of the LIA also concluded the response to be depended..' dependent

Thanks, corrected.

4. Line 416 'reveals a less good agreement (see Appendix A, Figure A3)' why not say 'worse agreement'?

Thanks for your comment. We rephrased the sentence.

5. Line 445 'Taking into account the entire range of ensemble members is therefore important' The authors mean here, I think, an ensemble of simulations. However, the paragraph is about ensemble of reconstructions (citing Büntgen et al.), and therefore, the sentence is unclear

Thanks for your comment. We mean both ensembles here. Büntgen et al. (2021) show that using different statistical methods on the same data gives different possible temperature reconstructions, which can be approached in the same manner as an ensemble of realizations of a specific model experiment. Therefore, it is important to take into account the range of the ensemble from both climate model simulations and proxy reconstructions. This has been added to the manuscript for clarification.

‘*Taking into account the entire range of ensemble members from both climate model simulations and proxy reconstructions is therefore important.’*

References:

Büntgen, U., Allen, K., Anchukaitis, K. J., Arseneault, D., Boucher, É., Bräuning, A., ... & Esper, J. (2021). The influence of decision-making in tree ring-based climate reconstructions. *Nature communications*, *12*(1), 1-10.

Reviewer 3

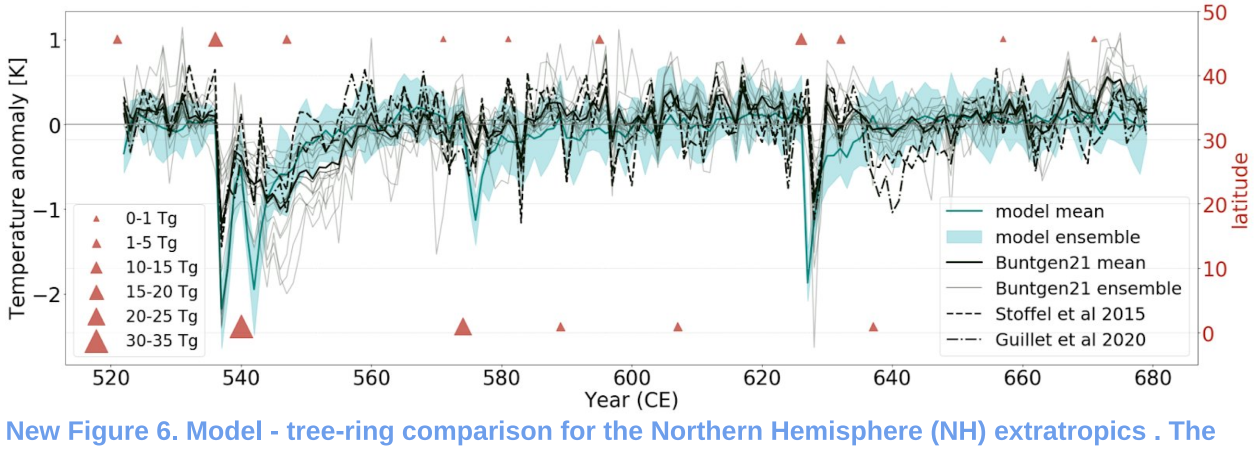
This is a substantially improved revision and the new organization and clear motivation and intent of the study makes it easier to follow the chain of logic and evidence. I thank the authors for addressing these.

Thank you for your constructive comments. We have taken your comments into account in this round of revisions, as answered in detail below.

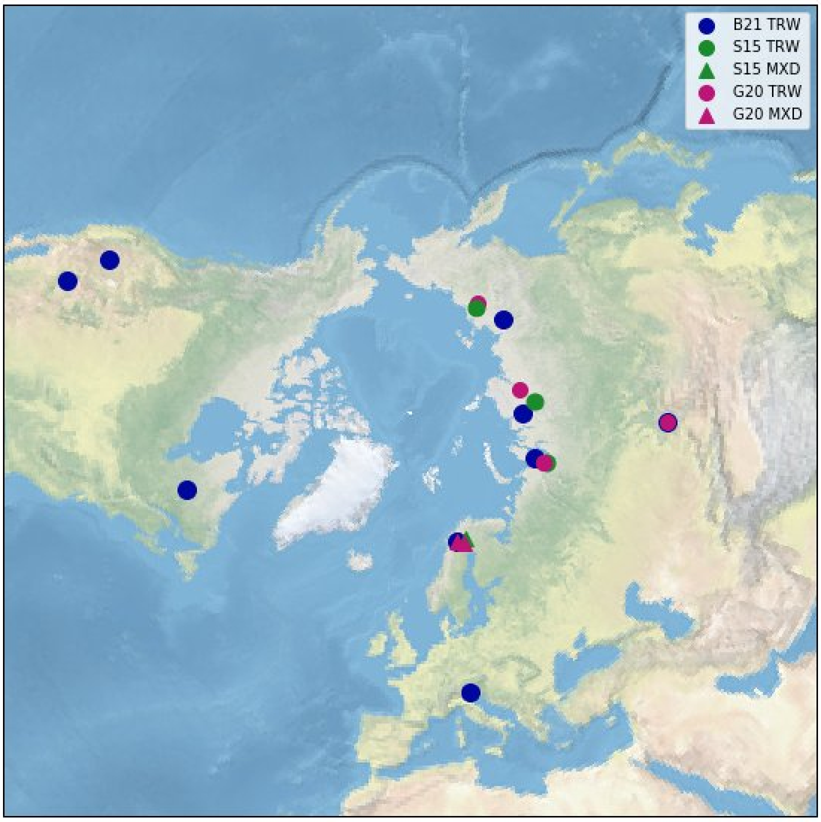
A few substantial comments:

1. The authors elected in this revision to focus their NH comparison strictly on the Büntgen et al. 2021 reconstruction. I think a stronger manuscript would still have included comparison to other reconstructions - like those of Stoffel and Guillet -- that also go back to the 6th century. I have no desire to cause additional work for the authors, but considering there are multiple reconstructions for this period, evaluating those (as well as taking the range of the Buntgen reconstruction ensemble reconstruction into account) would help better define how uncertainties in reconstruction temperatures in the first millennium affect the comparison with the ensemble of climate models.

Thanks for your comment. We updated Figure 6 in the manuscript, and the corresponding map in the appendix (Fig. A1, see below) taking Büntgen et al. (2021), Stoffel et al (2015), and Guillet et al (2020) into account, which highlights the uncertainties in reconstruction temperatures in the first millennium compared with the ensemble of the climate model. In the Appendix (Figures A2 and A3), we also show a comparison of the modeled temperature response to other proxy reconstructions from trees Esper (2012, Northern Scandinavia), Büntgen et al (2016, Alps and Altai), Luterbacher et al (2016, Europe) and PAGES2k Neukom et al (2019, NH).



New Figure 6. Model - tree-ring comparison for the Northern Hemisphere (NH) extratropics . The model 2m air temperature anomalies are taken for land only, JJA and 40-75N. Climate model mean and the spread of the model ensemble, tree-ring data for NH1 (Stoffel et al., 2015), N-VOLCv2 NH reconstruction (Guillet et al., 2020), and the mean and the ensemble of Büntgen et al. (2021). Anomalies are wrt 1-1850 CE except for the Guillet et al. data which are wrt 500-1850 CE.



New Figure A1. Locations of the tree-ring sites used in this study from Stoffel et al. (2015, green dots), Guillet et al. (2020, purple dots), and Büntgen et al. (2021, blue dots). Details can be found in Table A2. Tree ring width data is represented by the circles and maximum latewood density by triangles.

2. On a related note, I think more emphasis could be placed on looking at the range of responses in both reconstruction and simulation ensembles. The authors address this somewhat tangentially in Line 442 to Line 445 -- that 'real life' is just one iteration of possible climate states, just as the members of the simulation ensemble are possible trajectories the climate system could take. As e.g. Zambri et al. 2019 showed for Laki, initial conditions and internal variability can play an important role in how the forced signal is expressed - thus the existence of both an ensemble of climate model simulations AND an ensemble of reconstructions from Buntgen provides the unique opportunity to consider the range of possible temperature responses resulting from the mixing of internal and forced behavior in the climate system and the uncertainties in the reconstruction process together.

Thanks for your comment. We agree this is an important point and we have elaborated on this in the discussion. In addition, our follow up paper van Dijk et al. (2022, in review) focuses more on the different individual ensemble members, which we also included in the discussion.

‘***Moreover, the study from Büntgen et al. (2021) shows that it is important to also use an ensemble when it comes to tree-ring reconstructions, as the different statistical methods used to analyze the data give different results. Even though the ensemble mean shows a discrepancy with the model simulation after the 536/540 CE eruptions, some ensemble members fall within the range of the model ensemble spread. Overall, the model ensemble shows less variability in particular to the 536/540 CE response but also to other volcanic eruptions than the tree-ring reconstruction ensemble. The reality can be viewed as one iteration of what could have happened under different initial conditions, ocean states and internal variability. Taking into account the entire range of ensemble members from both climate model simulations and proxy reconstructions is therefore important. In our follow***

up paper (van Dijk et al., 2022, in review) the individual members are analyzed in more detail for Scandinavia.'

A few minor points and some suggestions for the authors, particularly with respect to ensuring the introduction is clear:

3. Line 8: 'lasting up to 20 years is simulated' - this is a bit unclear as written - 20 years cooling after each eruption? Can you clarify this?

Thank you for your comment. By 'up to 20 years' we mean that the response time after all of the eruptions combined is ~20 years. In line 162 in the results section the response time for each eruption is given. We have clarified this in the sentence in line 8.

'***After the four large eruptions in 536, 540, 574 and 626 CE, a significant mean surface climate response in the NH lasting up to 20 years is simulated.'***

4. Line 24: It may be worth adding that volcanic eruptions are also our only natural analog for solar radiation management, so understanding (and being able to model them) is critical for this as well.

Thank you for your comment. We feel one has to be a bit careful here, as the dynamical response in the stratosphere can be of opposite sign, so we do not want to simplify too much. We have added the following sentence to the introduction:

'***In order to assess what potential impact they could have on future climate, it is important to understand what the climate response to volcanic forcing was in the past, and which mechanisms are involved. In addition, volcanic eruptions are often studied as a natural analog for solar radiation management (Robock et al., 2013).'***

5. Line 26: what is the difference between a cluster eruption and a double eruption? I have heard people suggest that 'double eruption' is potentially misleading (suggesting multiple eruptions from one source), so best to be clear in this sentence.

Thank you for your comment. We agree that one has to be careful with the phrase 'double eruption', and so we have changed this throughout the manuscript to 'volcanic double event' to clarify it was not necessarily the same volcanic system.

6. Line 30: I think this sentence is unclear as written: 'a shift in ice-core records lead ...' -- suggest rewording to: ' ... and updated ice core chronologies reveal two sulphur peaks that correspond to eruptions in 536 and 540 CE (Sigl et al. 2015)'.

Thank you for your comment. We have taken this up together with the comments from reviewer 4 and changed it to:

"Furthermore, historic documents reported a dimming of the sun in 536 CE (Stothers, 1984; Rampino et al., 1988). Revised ice core chronologies reveal two nearby sulfate peaks that correspond to eruptions in 536 CE and 540 CE followed by two large eruptions in 574 CE and 626 CE (Baillie, 2008; Sigl et al., 2015)."

7. Line 32: suggest replacing 'Thus' with 'Based on these records,'.

Thanks, we have taken this up in our text.

8. Line 35: suggest removing the sentence 'This all lines up to a cold period that was initiated by volcanic eruptions' and moving the following sentence (' Indeed, four large volcanic eruptions occurred in 536, 540, 574, and 626 CE (Sigl et al., 2015).') earlier in the paragraph, after the sentence 'Thus, this period was called the Late Antiquity Little Ice Age (LALIA)'.

Thanks for your suggestion, we have removed the sentence and altered the other sentences accordingly, see our answer to comment No. 6.

9. Line 104 and elsewhere: There is no 'Year 0' (the Gregorian calendar goes from 1 BCE to 1 CE).

Thanks for your comment. You are correct and we have changed this throughout the manuscript.

10. Line 252-259: It would be important here to talk, however, about the ensemble range from Buntgen et al. - for 536 and the following decade in particular, there is large range between the reconstruction ensemble members. Most useful would be to talk about the overlapping (or not) ranges of the ensemble climate model simulations and the ensemble reconstruction.

Thanks for your suggestion. We agree this is an important point and have added a description of both ensemble spreads to the results section, taking the new Figure 6 (see your comment 2 above) into account. We discuss the ensemble ranges in the discussion section, see also our reply to comment #2.

11. Line 260: omit 'that catches attention'

Thanks, we have omitted this part of the sentence.

12. Line 264 suggest 'which may be due to ...'

Thanks, this has been changed according to your comment.

13. Line 296: suggest replacing 'A few of the debated points are about discussing' with 'These include'

Thanks, we have followed your suggestion.

14. Line 300: 'which could dampen the signal' - can you say more about this? Wouldn't this suggest the signal is unforced then (part of internal variability)?

Thanks for your question. As we have a mix of tropical and extratropical eruptions including a double event, we probably can not expect a significant signal in the volcanic mean NAO response (Schneider et al., 2009). We have clarified this and added the following instead:

‘*which may explain the lack of a significant NAO response’.*

15. Line 328: suggest replacing 'large noise' with 'large internal variability'

Thanks, we have taken up your comment and changed the sentence accordingly.

16. Lines 409-410: I'm not sure how to reconcile the observation that the match is good for 'the recovery time of ~20 year after the peak cooling' but also that the match is bad for 'the lag after the 536/540 CE eruptions'? These seem like two contradictory statements?

Thank you for the comment, we see your point that this is a contradiction. We have changed the sentence according to the new Figure 6 above.

‘***The model - tree-ring comparison with the Büntgen et al. (2021) tree-ring ensemble reconstruction (Fig. 6) shows a very good agreement in the timing of the peak cooling of the 2 m air temperature anomalies for the NH extratropics land only JJA. The mismatches that are still present in this NH comparison, like the strength of the peak cooling, as well as the lag after the 536/540 CE eruptions, include potential deficiencies and uncertainties regarding the method. For example, for the Büntgen et al. (2021) tree ring reconstructions TRW was used, which is known to give a lagged and smoothed response (e.g. Esper et al., 2015; Zhang et al., 2015; Lücke et al., 2019; Zhu et al., 2020). This could explain the lag after the volcanic double event for Büntgen et al. (2021), whereas the other two reconstructions, Stoffel et al. (2015) and Guillet et al. (2020), are more in line with the model simulations.’***

17. Line 412: ' Reconstructions are becoming more uncertain ...' - this is true in general, although not exactly relevant to the current paper, where the number of chronologies in Buntgen is the same going back through the 6th century.

Indeed. Here we refer to reconstructions in general, which are becoming more uncertain back in time, to explain why we chose not to include certain other reconstructions in the main manuscript. We have clarified this point by adding the following:

‘***... This could explain the lag after the double eruption event for Büntgen et al. (2021), whereas the other two reconstructions, Stoffel et al. (2015) and Guillet et al. (2020), are more in line with the model simulations. As reconstructions are becoming more uncertain further back in time due to the sparseness of available proxy records, which mainly rely on tree ring records. This is the case especially before 1200 CE (Masson-Delmotte, 2013; Esper et al., 2018; Neukom et al., 2019) and therefore, we have chosen to use the Büntgen et al. (2021) reconstruction as it uses the same number of tree ring records throughout the entire CE. Additionally, we have chosen to include the reconstructions by Stoffel et al. (2015) and Guillet et al. (2020) as they both consist of a mix of TRW and MXD records. Testing a comparison of the model results with multiproxy temperature reconstructions for the entire NH annual mean (Neukom et al., 2019) reveals a worse agreement (see Appendix A, Figure A3).’***

18. Line 428: 'do not agree as for' - I'm unclear what this means? Does this mean the specific comparison is not as good as the NH comparison, or that they are similarly bad? Please clarify.

Thanks, here we mean that the specific comparison is not as good as for the NH comparison. We have clarified this in the text.

‘***Comparing these specific sites, the model and tree-ring reconstruction do not agree as well as they do for the NH tree ring ensemble reconstructions …’***

19. Line 449: As in the my previous review, I don't think this is at all relevant - if anything the cooling would have made the trees better recorders of temperature and there is no reason in this case to suspect a role for moisture anomalies specifically in the local mismatch between models and trees.

After our first round of revision, we 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 show a slight increase in simulated precipitation for the 20 year mean, but they also have an increase in evaporation at the same time. We find the atmospheric circulation separation at 40/45°N latitude very interesting, and thus would like to leave it in the paper to stimulate further work and discussion within the paleo climate proxy record community.

References:

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

Büntgen, U., Myglan, V. S., Ljungqvist, F. C., McCormick, M., Di Cosmo, N., Sigl, M., ... & Kirdyanov, A. V. (2016). Cooling and societal change during the Late Antique Little Ice Age from 536 to around 660 AD. *Nature geoscience*, *9*(3), 231-236.

Büntgen, U., Allen, K., Anchukaitis, K. J., Arseneault, D., Boucher, É., Bräuning, A., ... & Esper, J. (2021). The influence of decision-making in tree ring-based climate reconstructions. *Nature communications*, *12*(1), 1-10.

Esper, J., Büntgen, U., Timonen, M., & Frank, D. C. (2012). Variability and extremes of northern Scandinavian summer temperatures over the past two millennia. *Global and Planetary Change*, *88*, 1-9.

Esper, J., Schneider, L., Smerdon, J. E., Schöne, B. R., & Büntgen, U. (2015). Signals and memory in tree-ring width and density data. *Dendrochronologia*, *35*, 62-70.

Guillet, S., Corona, C., Ludlow, F., Oppenheimer, C., & Stoffel, M. (2020). Climatic and societal impacts of a “forgotten” cluster of volcanic eruptions in 1108-1110 CE. *Scientific reports*, *10*(1), 1-10.

Lücke, L. J., Hegerl, G. C., Schurer, A. P., & Wilson, R. (2019). Effects of memory biases on variability of temperature reconstructions. *Journal of Climate*, *32*(24), 8713-8731.

Luterbacher, J., Werner, J. P., Smerdon, J. E., Fernández-Donado, L., González-Rouco, F. J., Barriopedro, D., ... & Zerefos, C. (2016). European summer temperatures since Roman times. *Environmental research letters*, *11*(2), 024001.

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

Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., ... & von Gunten, L. (2019). Consistent multi-decadal variability in global temperature reconstructions and simulations over the Common Era. *Nature geoscience*, *12*(8), 643.

Robock, A., MacMartin, D. G., Duren, R., & Christensen, M. W. (2013). Studying geoengineering with natural and anthropogenic analogs. *Climatic Change*, *121*(3), 445-458.

Schneider, D. P., Ammann, C. M., Otto-Bliesner, B. L., & Kaufman, D. S. (2009). Climate response to large, high-latitude and low-latitude volcanic eruptions in the Community Climate System Model. *Journal of Geophysical Research: Atmospheres*, *114*(D15).

Stoffel, M., Khodri, M., Corona, C., Guillet, S., Poulain, V., Bekki, S., ... & Masson-Delmotte, V. (2015). Estimates of volcanic-induced cooling in the Northern Hemisphere over the past 1,500 years. *Nature Geoscience*, *8*(10), 784-788.

van Dijk, E., Mørkestøl Gundersen, I., de Bode, A., Høeg, H., Loftsgarden, K., Iversen, F., ... & Krüger, K. (2022). Climate and society impacts in Scandinavia following the 536/540 CE volcanic double event. *Climate of the Past Discussions*, 1-55.

Zhang, H., Yuan, N., Esper, J., Werner, J. P., Xoplaki, E., Büntgen, U., ... & Luterbacher, J. (2015). Modified climate with long term memory in tree ring proxies. *Environmental Research Letters*, *10*(8), 084020.

Zhu, F., Emile-Geay, J., Hakim, G. J., King, J., & Anchukaitis, K. J. (2020). Resolving the differences in the simulated and reconstructed temperature response to volcanism. *Geophysical Research Letters*, *47*(8), e2019GL086908.

Reviewer 4

Dear authors,
Thank you for the revisions which have improved the paper. For the submission to become acceptable, more work is needed as some of the new text elements do not fit nicely into the previous version which makes it difficult to follow all ideas. I would also strongly encourage the authors to have their paper edited by a native speaker.
The most relevant comments are listed below:

Thank you for your thorough comments. We have arranged the flow of the text elements and the manuscript was checked by a native speaker for English grammar.

We had some problems with your line numbering listed below, as your comments do not seem to match the lines that are referred to in those comments. We have tried our best to interpret to which section the comments apply, and responded accordingly below.

1. lines 41-42: One major double event leading to widespread and severe cooling occurred in 1108-10. Please add the following citation: Guillet, S., Corona, C., Ludlow, F.M., Oppenheimer, C., Stoffel, M., 2020. Climatic and societal impacts of a "forgotten" cluster of volcanic eruptions in 1108-1110 CE. Nature Scientific Reports 10, 6715.

Thanks for your suggestion. We have added this reference to the introduction.

2. line 46: Sigl et al. (2015) rely on tree-ring data that were - at least partly - published in Büntgen et al. (2011). I suggest to remove the Sigl reference here as that paper used ice cores primarily.

Thank you for your suggestion. We have removed the Sigl et al. (2015) in this context.

'***One of the coldest decades of the last 2000 years in the NH and Europe is visible in tree-ring records during the mid-sixth century (Larsen et al., 2008; Büntgen et al., 2011).'***

3. line 47ff: This is a bit messy in my view. The ice-core data need to be put in context more carefully. Also, the reference to four eruptions (lines 54-55) come somewhat late.

Thanks for your comment. We have restructured the paragraph including Reviewers 1 and 3 comments as well as follows:

…"**Several cluster eruptions and volcanic double events occurred in the last 2000 years as recorded in the ice core records, coinciding with cold periods in Northern Hemisphere (NH) tree-ring records (Briffa et al., 1998; Sigl et al., 2013). One of the coldest decades of the last 2000 years in the NH and Europe is visible in tree-rings during the mid-sixth century, which was preceded by two volcanic eruptions as recorded in ice cores (Larsen et al., 2008; Büntgen et al., 2011). Furthermore, historic documents reported a dimming of the sun in 536 CE (Stothers, 1984; Rampino et al., 1988). Revised ice core chronologies reveal two sulfate peaks that correspond to eruptions in 536 CE and 540 CE followed by two large eruptions in 574 CE and 626 CE (Baillie, 2008; Sigl et al., 2015). Reconstructed tree-ring temperatures from the Alps and Altai show a century-long cooling that might have exceeded that of the Little Ice Age (LIA) during the 14th-19th century (Büntgen et al., 2016). Based on these**

records, this period was called the Late Antiquity Little Ice Age (LALIA). wever, other studies reveal contrasting results35
on how long lasting the surface cooling in the NH extratropics was, varying from multi-decadal to centennial cooling. These results are based on tree ring reconstruction methods and tree ring record updates, as well as ice-core records and documentary evidence (e.g. Esper et al 2012b; Matskovsky and Helama, 2014; Helama et al., 2017; Guillet et al., 2020; Büntgen et al., 2021; Helama et al., 2021). Thus, the duration of this volcanic induced cooling event remains open."...

4. lines 170-1: Setting all eruptions to January will have an influence on the cooling and the persistence of cold conditions. This has to be said more clearly and - ideally - remediated. Using different seasonalities would certainly help to determine the spread of cooling and uncertainties.

We are using the volcanic forcing data set eVolv2k (Toohey and Sigl, 2017), which has been recommended for CMIP6/PMIP4 past2k simulations (Jungclaus et al., 2017)). This volcanic forcing sets January as the standard eruption month for unknown eruptions. Upcoming CMIP7/PMIP5 volcanic forcing datasets will remediate this, where our study is contributing to. Indeed, the eruption season is important for understanding the climate response, as clearly mentioned in the discussion in line 405. For the 536 CE eruption a winter eruption is consistent with the timing of the dust veil observed over Europe in March (Stothers, 1984). Sensitivity experiments with different seasons have been carried out elsewhere (Toohey et al., 2011). We have clarified this in the discussion.

'***In addition, the eruption season is also important, as different eruption seasons give different atmospheric circulation patterns and therefore influence the transport of the sulfate aerosol, leading to different surface responses (Toohey et al., 2011).'***

5. lines 179ff: Relying exclusively on tree-ring width data (as presented in Büntgen et al. 2021) is problematic. Several of the NH reconstructions extend back to 500 CE and should be kept as they contain latewood density data which are less affected by memory in tree rings.

Yes we agree. Thus, we have added the NH reconstructions by Stoffel et al. (2015) and Guilett et al. (2020), which also include MXD data, to Figure 6, as also suggested by Reviewer 3. In addition, we have added a discussion to the reconstructions from Matskovsky and Helama, 2014 and Helama et al., 2021 as suggested by you below as well. The text has been adapted accordingly in the manuscript including MXD data in the introduction, method, results and the discussion sections.

Tree ring data (Methods):
'***For the model-tree ring temperature comparison, different tree ring data and reconstructions are used. The tree ring data used are based on tree ring width (TRW) and maximum latewood density (MXD). TRW is known to have biological memory and gives a lagged and smoothed response to volcanic eruptions. In contrast, MXD is based on cell density, which gives a better representation of volcanic induced surface cooling (Anchukaitis et al., 2012; Esper et al., 2015; Zhu et al., 2020; Ludescher et al., 2020). MXD data is therefore the preferred target for our model comparison if available, but the data is sparse during the first millennium. Thus, both tree ring methods are taken into account using the reconstructions from Stoffel et al. (2015), Guillet et al. (2020), and Büntgen et al. (2021), next to others (see Appendix A). Stoffel et al. (2015) and Guillet et al. (2020) consist of a mix of MXD and TRW. The tree ring ensemble reconstruction from Büntgen et al. (2021) is based on TRW only and is taken from 9 sites over the NH covering the past 2000 years (Fig. A1). The raw data were distributed to 15 different dendrochronology groups. These***

groups all carried out their own statistical methods on the data, after which the now different data sets were combined to form a tree ring ensemble. More details about the data and the reconstruction method can be found in the corresponding publications.
To use the same reference period for the model and tree ring data, we subtracted the 1-1850 CE mean from the model and tree ring ensemble. For the model-tree ring comparison a land mask was applied to the model 2 m air temperature and we analyzed only the NH extratropics between 40◦and 75N. The tree ring data sets capture the boreal summer temperature during June, July and August (JJA) and were therefore compared to JJA 2 m air temperatures from the model.'

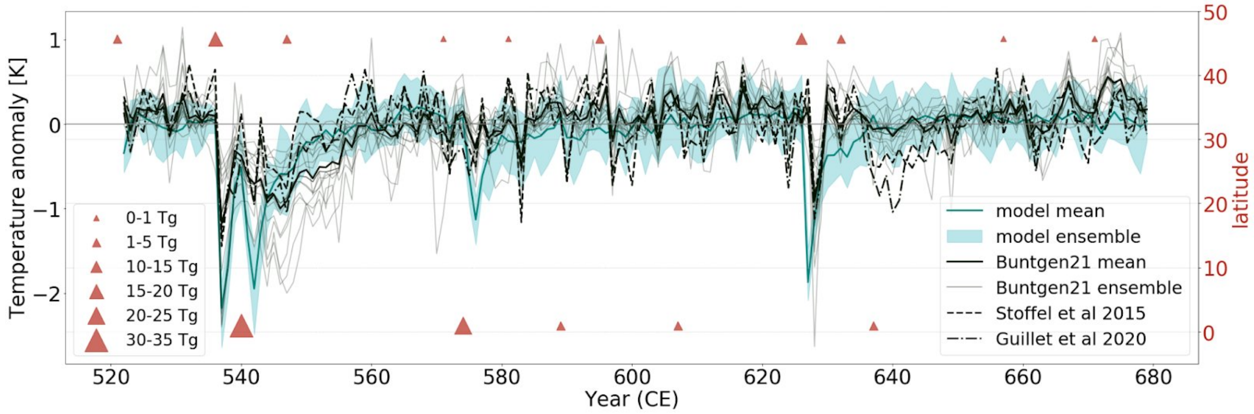
6. lines 188ff: Yes, you are right that MXD data is sparse but leaving it out completely would be the wrong decision. I also suggest that the authors check the following publication and discuss results therein with their findings: Helama, S., Stoffel, M., Hall, R.J., Jones, P.D., Arppe, L., Matskovsky, V.V., Timonen, M., Nöjd, P., Mielikäinen, K., Oinonen, M., 2021. Recurrent transitions to Little Ice Age-like climatic regimes over the Holocene. Climate Dynamics 6, 3817–3833.

Thanks for your suggestion and see our responses to include MXD data above. We have added a discussion of our results with regard to the Helama et al. (2021) and Matskovsky and Helama (2014) papers in the tree ring section (new line 454). Next, we have added Helama et al. (2021) to the discussion on atmospheric circulation patterns (new line 350).

'*In addition, Matskovsky and Helama (2014) and Helama et al. (2021) report a century long cooling for Northern Scandinavia from 530 to 650 CE based on MXD, TRW and δ13C data, which is not supported by our model simulations (Fig. A2a). Other proxy-based studies (Larsen et al., 2008; Esper et al., 2012b; Luterbacher et al., 2016; Helama et al., 2017; Neukom et al., 2019) found a cooling up to 570 CE for Europe, Scandinavia, and the NH, based on tree ring, ice-core, lake sediment, and documentary records (Figs. A2a and d, and A3). Comparisons for the different tree ring data sets have been carried out by previous studies for the last millennium (Wilson et al., 2016; Lücke et al., 2019). However, not enough records exist for the first millennium to carry out a similar comparison for this period yet. From the perspective of our model results, the persistence of the cooling was not as long lasting as the tree ring sites from the Alps, Altai, and Northern Scandinavia suggest. … Another possibility is, that our model resolution is too coarse to fully capture the topography of the Alps, Altai, and Northern Scandinavia…*'

'*Helama et al. (2021) describe an East Atlantic pattern during the study period, corresponding to clear sky conditions over Northern Scandinavia, as obtained from tree ring proxies (TRW, MXD, and stable carbon isotope (δ13C)). This pattern of reduced cloudiness is consistent with the dry conditions simulated over Northern Europe in our model simulations.'*

7. line 325: what is the cooling obtained in NH reconstructions based on MXD or mixed MXD-TRW reconstructions? It would be nice to have these as well.



New Figure 6. Model - tree-ring comparison for the Northern Hemisphere (NH) extratropics. The model 2m air temperature anomalies are taken for land only, JJA and 40-75N. Climate model mean and the spread of the model ensemble, tree-ring data for NH1 (Stoffel et al., 2015), N-VOLCv2 NH reconstruction (Guillet et al., 2020), and the mean and the ensemble of Büntgen et al. (2021). Anomalies are wrt 1-1850 CE except for the Guillet et al. data which are wrt 500-1850 CE.

Yes, we agree and we have updated Figure 6 in the manuscript (see new figure 6) and the corresponding text. See our previous comments to you and Reviewer 3. MXD records do not exist for the entire NH, only mixed TRW and MXD reconstructions are available. However, Wilson et al. (2016) as well as Lücke et al. (2019) provide a comparison between modeled, TRW, MXD and mixed reconstructions for the last millennium. For a comparison like this to be possible for the first millennium requires more investment in tree-ring records for the NH during this period. We have added these two references now to the discussion.

'Comparisons for the different tree ring data sets have been carried out by previous studies for the last millennium (Wilson et al., 2016; Lücke et al., 2019). However, not enough records exist for the first millennium to carry out a similar comparison for this period yet.'

8. line 326ff: TRW data have biological memory, so comparing persistence is of limited use. You would better compare with the available NH MXD records here and put the TRW, MXD, Neukom and model results into perspective.

Thank you for your comment. We have included this comparison now and updated Figure 6 in the manuscript and the corresponding text. See also our previous comments to you and Reviewer 3.

9. line 334: Why do you assume January for the eruption if documentary evidence suggests an autumn eruption?

We assume the reviewer is referring to the 626 CE eruption and the reason for the mismatch between the model simulations and the tree-ring records. See also our answer to your comment #4 above. For the volcanic forcing the eVolv2k data set (Toohey and Sigl, 2017) was used following the CMIP6/PMIP4 protocol (Jungclaus et al., 2017), January is used as the standard eruption month when the eruption is unknown. Upcoming CMIP7/PMIP5 volcanic forcing datasets will remediate this, where our study is contributing to.

References:

Guillet, S., Corona, C., Ludlow, F., Oppenheimer, C., & Stoffel, M. (2020). Climatic and societal impacts of a "forgotten" cluster of volcanic eruptions in 1108-1110 CE. *Scientific reports*, *10*(1), 1-10.

Helama, S., Stoffel, M., Hall, R. J., Jones, P. D., Arppe, L., Matskovsky, V. V., ... & Oinonen, M. (2021). Recurrent transitions to Little Ice Age-like climatic regimes over the Holocene. *Climate dynamics*, *56*(11), 3817-3833.

Jungclaus, J. H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., ... & Zorita, E. (2017). The PMIP4 contribution to CMIP6–Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations. *Geoscientific Model Development*, *10*(11), 4005-4033.

Lücke, L. J., Hegerl, G. C., Schurer, A. P., & Wilson, R. (2019). Effects of memory biases on variability of temperature reconstructions. *Journal of Climate*, *32*(24), 8713-8731.

Matskovsky, V. V., & Helama, S. (2014). Testing long-term summer temperature reconstruction based on maximum density chronologies obtained by reanalysis of tree-ring data sets from northernmost Sweden and Finland. *Climate of the Past*, *10*(4), 1473-1487.

Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., ... & Woodruff, T. E. (2015). Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature*, *523*(7562), 543-549.

Stoffel, M., Khodri, M., Corona, C., Guillet, S., Poulain, V., Bekki, S., ... & Masson-Delmotte, V. (2015). Estimates of volcanic-induced cooling in the Northern Hemisphere over the past 1,500 years. *Nature Geoscience*, *8*(10), 784-788.

Stothers, R. B. (1984). Mystery cloud of AD 536. *Nature*, *307*(5949), 344-345.

Toohey, M., Krüger, K., Niemeier, U., & 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), 12351-12367.

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

Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D'arrigo, R., ... & Zorita, E. (2016). Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context. *Quaternary Science Reviews*, *134*, 1-18.