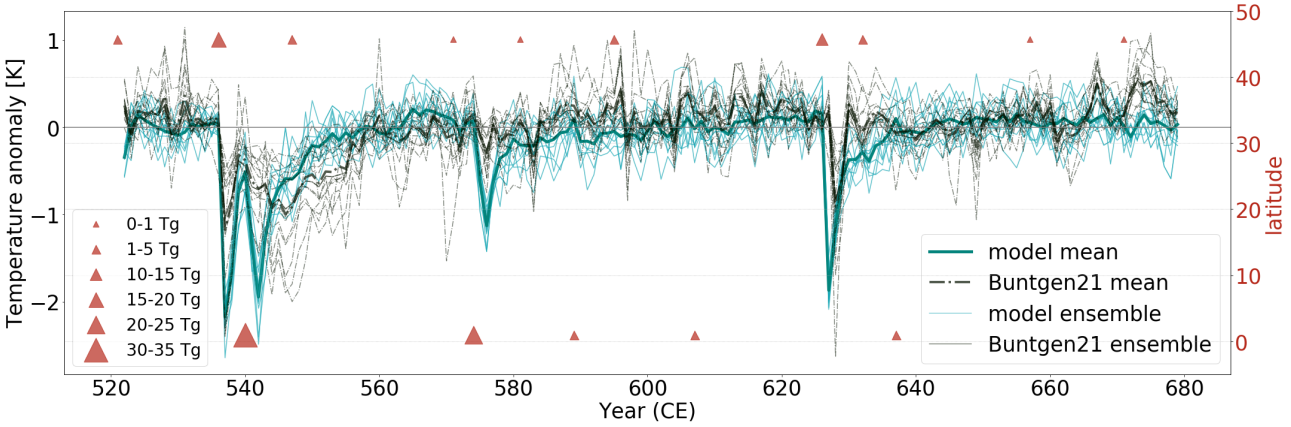
Dear editor and reviewers, thank you for the detailed review of the paper. We have compiled a point by point response going through all comments of the four reviews. Our comments are given in blue and bold text.

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 1:

Summary: The manuscript focuses on the climate response to the strong volcanic eruptions that occurred in the 6th century. The authors analyze an ensemble of climate simulations with the model MPI-ESM, compare their results with other previous simulations and with temperature reconstructions based on dendroclimatological data. The analysis of model data includes the atmospheric response (temperature, precipitation, wind) and the response of the ocean circulation in the North Atlantic. One important issue is the duration of the response of these variables, as previous simulations and decollimated reconstructions seem to disagree - at lest in some regions.

Recommendation: In my opinion the scope and focus of the article are valuable. There are some knowledge gaps about the response of large volcanic eruptions and the origins of Little Ice Ages in the last millennia, that need to be filled.

However, the manuscript leaves clear room for improvement, and I believe it requires considerable revisions. The structure is sometimes confusing, the language is also not always clear, and the data analysis is not deep enough. This is reflected in the unclear take-home messages of the manuscript. Disagreements between simulations are explained by general model differences. Disagreements between reconstructions and simulations are explained by possible errors in the volcanic forcing or deficiencies of the proxy data, mainly in the tree-ring data, but the manuscript does not include

more specific and solid explanations. Sometimes, the discussion is inconsistent, and some examples of this are given in the list of particular points below.

Thanks for your constructive comments. We have taken your and the three other reviewers' comments into account in the revised manuscript, as further answered above in the general comments and below in detail.

1. ‘ sea level pressure and a decrease in hydrological variables occur’

The text could be here clearer. Does this sentence mean that both precipitation and

evaporation become smaller?

The sentence indeed refers to both precipitation and evaporation. We have rewritten the abstract (see reviewers 2 and 3’s comments) and we have taken your comment into account.

‘...a decrease in both precipitation and evaporation poleward of....’

2. ‘However, most reconstruction data sets go back to about 1200 CE, and the further back in time, the fewer proxy records remain, and the more uncertainties they contain (Masson-Delmotte,2013; Neukom et al., 2019). The sentence is not clearly formulated (perhaps it is not grammatically correct). Consider an alternative formulation like ‘ Further back in time the network of proxy data becomes sparser and uncertainties in each individual temperature record also grow.

We have revised the introduction due to reviewers 2 and 3’s comments and now write the following:

“**Multiproxy reconstructions in the NH during this period consist of mainly tree-ring, marine sediments, and ice core records. Marine sediments have a coarse chronological resolution and would therefore not capture the volcanic signal, and ice core records are confined to the Greenland Ice-sheet (PAGES 2k Consortium, 2017). Tree-ring records have an annual resolution with an absolute dating. Therefore, we use the most recent tree ring ensemble reconstruction for the past2k (Büntgen et al., 2021) to compare to the model simulations in this study.**”

We now use the tree rings ensemble data from Büntgen et al. (2021) as a reference data set, where one finds a detailed discussion about proxy uncertainties. This info has been added.

3. ‘The aim of this study is to investigate whether a multidecadal to centennial cooling may have occurred in the mid 6th to 7th century.’

This sentence states the main objective of the study. However, the reader will not be wiser after reading the manuscript. The disagreement between the proxy records that do indicate a centennial cooling do not agree with the model results, but the study does not include a solid explanation for this disagreement (apart from speculating that perhaps the high altitude of the Alp records may be responsible for the long temperature recovery).

This is perhaps my main concern with the manuscript. It is in general too descriptive and does not go deep enough into explaining those disagreements.

According to the suggestions of reviewers 3 and 4, we now include the new tree-ring reconstruction ensemble from Büntgen et al (2021). Thereupon, the disagreement of proxy records and model data is not the focus of our study anymore but a motivation for our study. We also refer to the discussion in Timmreck et al (2021) about the uncertainties in the model simulations. Therefore, the aim of the study has not changed/ has changed to "The aim of this study is to investigate whether a series of volcanic eruptions induced a multidecadal to centennial cooling in the mid 6th to 7th century.".

4. ' the short term (years), as well as the long term (decadal to centennial)

the short term (annual)...

Thanks, corrected.

5. 'sea ice impacts, we also study atmospheric and ocean circulation, hydrology and the ocean-sea ice feedbacks in maintaining the climate signal'

the 'climate signal' is too unspecific. Please, help the reader by being more specific, for instance 'in maintaining the volcanic induced cooling'

Thanks for the suggestion, the sentence has been reformulated as follows:

'...we also study atmospheric and ocean circulation, hydrology and the ocean - sea-ice feedbacks in maintaining the volcanic induced surface cooling.'

6. 'For this study, we ran ten ensemble members for 160 years from 520-680 CE.'

for 160 years, starting in 520 CE. Or alternatively, covering the period 520-680 CE

The sentence has been reformulated as follows:

'For this study, we ran ten ensemble members, covering the period 520-680 CE.'

7. For each ensemble member the atmospheric diffusivity was changed by 1 · 10−5 to simulate slightly different climate states by the year 536 CE, the year of the first large volcanic eruption.

physical units are missing

Thanks, we added the units:

‘For each ensemble member the atmospheric diffusivity was changed by $1 \cdot 10^{-5}$ $m^2 s^{-1}$...’

8. Historical Land Use Data Set for the Holocene (HYDE3.2, Klein Goldewijk, 2016). Considering several options (e.g. linear ramp-up) we decided to simply let the land-cover data be constant for the first 850 years of the past2k runs.

We prescribed a constant land-cover for the first 850 years of…

Thanks for the suggestion. The sentence was reformulated as follows:

‘Considering several options (e.g. linear ramp-up), we prescribed a constant land-cover for the first 850 years of the past2k runs.’

9. 'The tree-ring sites are displayed in Fig. A1. For the model-tree-ring comparison a land mask was applied to the model 2m air temperature analyzing the NH extratropics between 40°N and 75°N.'

Here and in other instances in the manuscript, it is not clear whether the reconstructed NH temperature was just calculated as the to be the simple average of the local temperature reconstructions at the tree-ring sites or whether there was a more sophisticated reconstructions method, for instance by calibrating a statistical model to replicate the NH mean temperature (as in Stoffel et al. 2015). The present manuscript lists in Table 2 just 6 records. Is the NH temperature the average of only these 6 records or is it the temperature reconstructed by Stoffel et al. ? I think that the simple average of these 6 records cannot meaningfully be considered a Northern Hemisphere average.

Thanks for bringing this to our attention. The temperature reconstructions are used directly from Stoffel et al, and therefore have the statistical model included. We changed to using the data from Büntgen et al. (2021), where the post-processed data was publicly available. The statistical methods are therefore as described by Büntgen et al. (2021). This has been clarified in the methods section.

10. 'Towards the end of the simulation period the ensemble shows a larger spread than at the beginning of the simulations, which corresponds to the ocean heat content state being more different between members in the end than at the beginning of the simulations.'

Could this be an indication that the ensemble set-up is not adequate to investigate the main objective of the study, and that the spread of ocean initial conditions is too narrow ?

The ensemble spread being larger at the end of the simulations is mainly apparent in the ocean heat content. For our proposed mechanism the AMOC and the ocean heat transport is more important, which shows a similar range of variability. We cannot rule out that starting from an earlier point in time, or using an ensemble would give a more different ocean state between the ensemble members. However, to carry out an ensemble, we would have needed an ensemble of past2k runs, which was and is not available and not feasible.
We have an included a paragraph on this in the new discussion section (see our response to remark #23 below).

11. The results section includes several paragraphs that actually would belong to a (missing) Discussion section. An example is this paragraph:

'Zhong et al. (2011), and Miller et al. (2012) argued that the ocean - sea-ice feedback could play a major role in sustaining a century long cooling after a cluster of four volcanic eruptions in the mid 13th century. In contrast to these studies, we simulate a multi-decadal sea ice response in the mid 6th to 7th century. '

Consider also this reformulation: In contrast to these studies, our simulated cooling is shorter and lasts only for a few decades.

The manuscript was written in a way that the results and discussion were placed in one section. The figures were first described and then discussed in order. We have now split the results and discussion section as suggested.

The sentence given as an example has been reformulated as follows:

‘In contrast to these studies, our simulated cooling is shorter and lasts for several decades.’

12. After the 536/540 CE double event, the ensemble mean of the model simulations does not return to zero sea-ice cover anomalies before 560 CE.

the ice-cover in the ensemble mean takes longer to recover and only reaches the climatological mean value by year xxxx.

Thanks for your suggestion, we have taken your comment into account, and now refer to the 0-1850 CE mean.

13. 'Fig 3 (TRW) and triangles (MXD) in the 2 m air temperature maps. The 2σ (1σ) standard deviations for 2m air temperature SLP, evaporation (and precipitation) are stippled (hatched). '

The ensemble standard deviation sigma (2xsigma) ...are stippled (hatched)

The comment (figure caption) has been taken into account, and has been corrected.

14. ‘there is a land-sea contrast present for evaporation in summer, where the signal is opposite over the ocean.’

In summer, the sign of the evaporation anomalies over ocean and land is opposite.

The sentence (line 195 in the original manuscript) has been reformulated as follows:

‘In summer, the sign of the evaporation anomalies over land and ocean is opposite.’

15. ‘the north side of the climatological high pressure systems reflecting an atmospheric circulation separation at around 45°N.’

this sentence is unclear. Does it mean that the sign of the anomalies is apposite north and south of 45N?

Indeed, it means the sign of the anomaly is opposite north and south of 45N. We have separated the results and discussion sections, and we have taken your comment into account when rewriting the sections.

16. ‘The long term response is shown on the right side of Figure 3.’

on the right side or in the right half of the picture. Better still is to label all the panels and refer to them accordingly.

The comment has been taken into consideration, and the figure has been changed accordingly. In addition, for better readability of the figure, we have separated Figure 3 into two figures (Fig 3 and 4), see also our reply to reviewer 2.

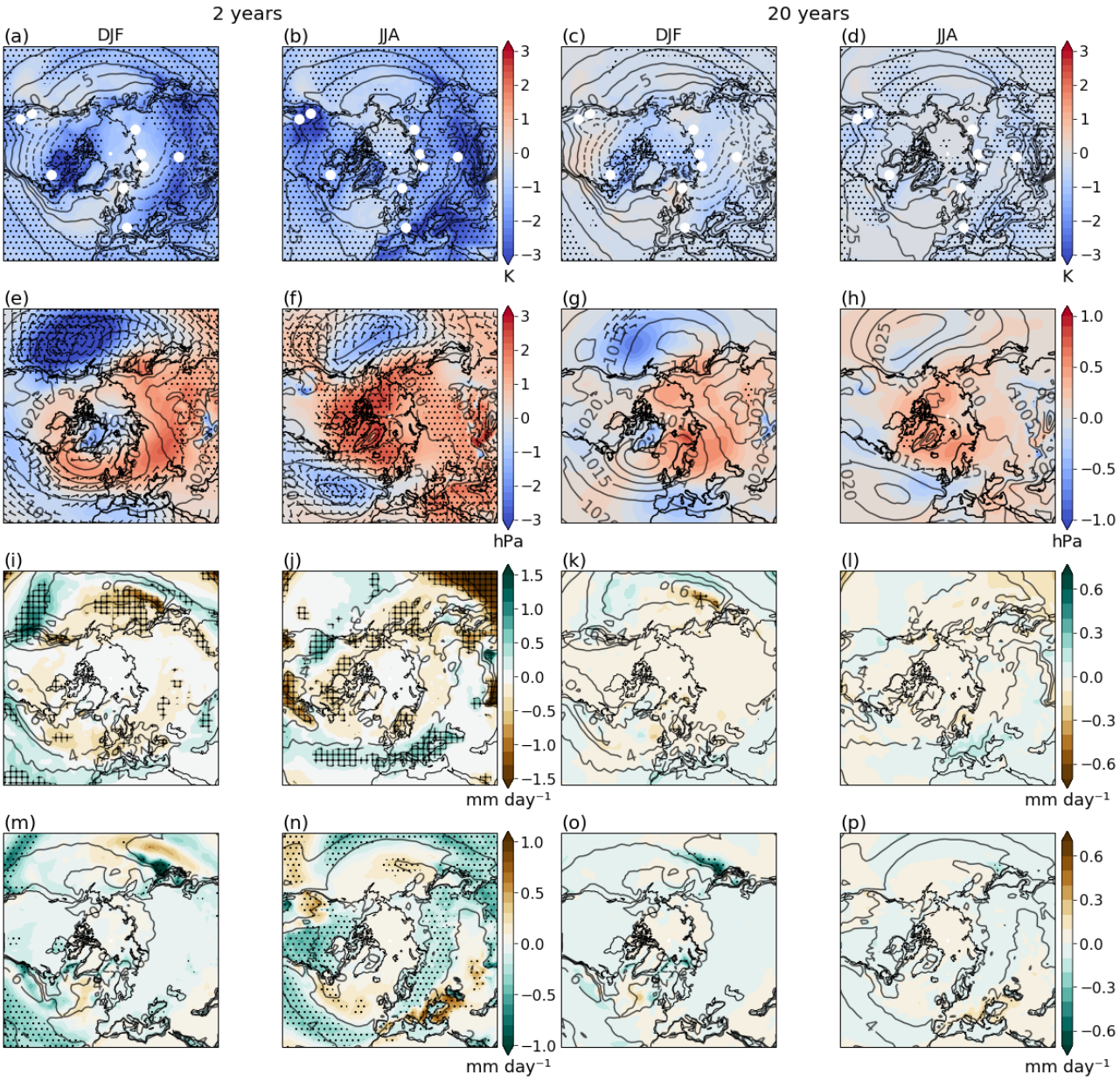


Fig. 3 NH maps of boreal winter (DJF) and summer (JJA) 2m air temperature (a-d), sea level pressure and 10 m wind (e-h), precipitation(i-l), and evaporation (m-p) for 2 years and 20 years after the eruptions, poleward of 30◦N. The maps represent the ensemble mean of the 2and 20 year mean after the four major eruptions in the study period. The 2 years after the eruption are year 1 and year 2 after the eruption forDJF, and the year of the eruption and one year after for JJA, respectively. All variables are given as anomalies wrt 0-1850 CE. The 0-1850CE climatology is given as contours and the tree-ring locations are represented by white dots in the 2 m air temperature maps. The ensemble standard deviation 1σ(2σ) for 2m air temperature, SLP, evaporation and precipitation are hatched (stippled). Note the different scales for the2 year and 20 year maps and that wind anomalies are shown only for 0.5 and 1.0 m/s intervals.

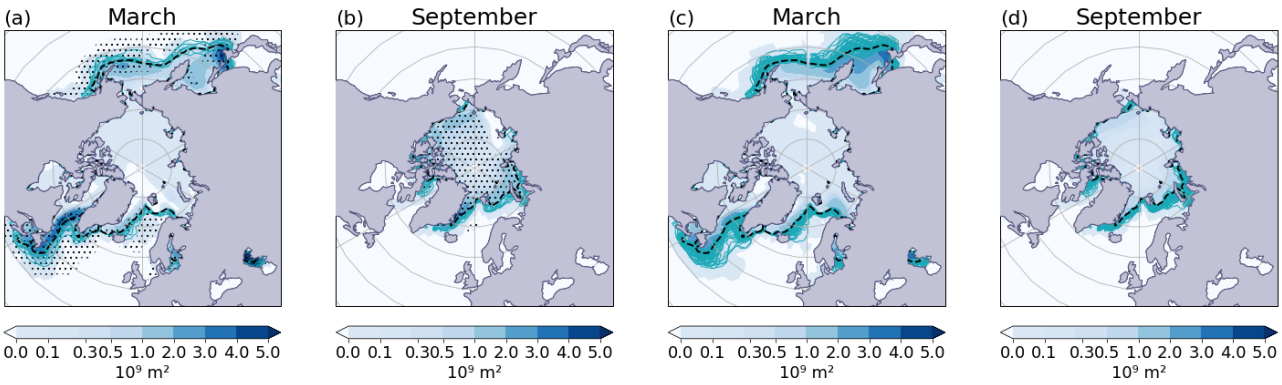


Fig 4 The March and September mean sea-ice area anomaly in contours and the sea-ice extent represented by the teal lines for 2 years after the eruptions (a) and (b) and 20 years after the eruptions (c) and (d). All variables are given as anomalies wrt 0-1850 CE. The blackhashed line is the mean sea-ice extent in the control run and the 2σsignificant sea-ice area anomalies are stippled.

17. 'The increase in precipitation over the Mediterranean in boreal summer in the model simulations in this study are related to the shifting of the inter tropical convergence zone (ITCZ) into the Southern Hemisphere (SH) after the eruptions (not shown here), as well as a weakening of the high and low SLP over the North Atlantic (Figure 3b). After a large volcanic eruption, the ITCZ shifts away from the cooler hemisphere, in this case the NH (Schneider et al., 2009). '

I guess that during the boreal summer , the Northern Hemisphere is the warmer hemisphere, not the cooler hemisphere. Perhaps the authors mean that the negative temperature anomalies are stronger in the NH?.

Thanks, this has been corrected.

18. ‘appeared to be opposite, with a drying over Southern Europe and a wettening over Northern Europe. They accounted this to the models not capturing the winter NAO well and therefore simulating a different response.’

The reference to Iles and Hegerl to discuss precipitation response in boreal summer over Europe is misplaced. Iles and Hegerl refer to the winter precipitation response, so it is not correct to state that Iles and Hegerl agree with the results obtained here for summer precipitation. It would also strange to claim that a wrong simulation of the winter NAO response can explain the wrong sign of summer precipitation anomalies in Europe.

Thank you for bringing this to our attention. The section has been rewritten with an appropriate reference.

‘Iles et al. (2013) used HadCM3 and concluded the mean response to 18 eruptions during 1442-1992 to be a wettening over the Mediterranean and a drying of Northern Europe for up to 4 years after the eruptions during the summer season.’

19. 'The significance was calculated from the 1200 year control run by taking the 2 to 20 year means of the 1200 years, and then taking 4 random time steps from that time series for the 4 large eruptions. This was done 1000 times for each variable, and the standard deviation was then calculated from those new random time series. 1 time or 2 times the standard deviation (1 and 2σ) were then used to calculate the significance.'

I cannot understand how the standard deviations were calculated. This paragraph seems to me rather unclear

Thanks for your suggestion. The paragraph has been rephrased as follows:

‘The significance was calculated from the 1200 year control run. To account for the different fast and slow response times, the 2 and 20 year means have been taken from the 1200 year control run. As the mean response is for 4 large eruptions, 4 random time steps were taken from the mean control time series to account for this. To get a significant sample, this was done 1000 times for each variable, and from these 1000 samples a new time series was created. The standard deviation was then calculated from this random time series. 1 or 2 sigma were calculated for whether the volcanic signal is significant.’

20. 'In boreal winter, there is a see-saw pattern visible in the 2 year SLP response with an increased low pressure over Greenland and a decreased low pressure over Northern Europe, corresponding to the seesaw winter temperature pattern between Greenland and Scandinavia, as described by Van Loon and Rogers (1978). The changes in boreal winter reflects a positive Arctic....'

This whole paragraph is rather unclear. I think there is an error in the first sentence (increased low pressure in both Greenland and Northern Europe?) , but in general it is difficult to follow. It contains a mixture of own results and previous results, and it is difficult to disentangle which is which. After reading the paragraph, it is unclear whether the model does produce a NAO response or not.

I would recommend to first describe the new results, and then briefly compare them with previous results. In a discussion section this comparison can be then deeper and more detailed.

The first sentence does not contain an error, in DJF we see a negative SLP anomaly over Greenland and a positive SLP anomaly over Europe.

The results and discussion have now been separated as recommended, and thus this section has been rewritten.

21. ‘The summer cooling over the continents can have a serious effect on the vegetation and society summer can lead to crop failure and famine in areas that are close to the temperature limit for grow....

This paragraph on the impacts of low temperatures is misplaced here. It is not related with the previous or the following paragraphs. This could go to a Discussion section or be deleted without any loss.

We have moved this paragraph to the end of the conclusion, as an outlook for future work / our next paper.

22. From Fig. 4b and c can be seen that the subpolar gyre (south of’it can be seen...

We are not sure what the reviewer refers to here. It seems that some parts of your comment have been lost. Nevertheless, we have rephrased part of the section “Ocean sea ice response” and hope that this has helped to clarify the discussion on the roles of AMOC and gyre circulation

23. I found the following two paragraphs to be inconsistent with the ensemble set-up:

‘The studies from Zhong et al. (2011) about the onset of the LIA also concluded the response to be depended on the initial state of the North Atlantic, as only 2 out of 4 simulations (one warm and one normal NA state) lead to a cooling long enough to resemble the LIA. Compared to their study, our NA state is relatively warm, but it is hard to compare as a different model and set up were used.’

‘be the initial state of the ocean when the volcanic eruptions occur. This is less likely, as 10 ensemble members were run, which showed a range of variability in the same range as the 0-1850 CE variability, where the response to the volcanic eruptions’

The first paragraph states that the initial ocean state is important for a successful simulation of the LIA. The second paragraph indicates that the ocean initial conditions in the ensemble are wide enough separated. However, the initial conditions were prepared by just perturbing one single parameter in the atmosphere submodel, and else where the manuscript states that the spread of the ensemble at the end of the simulations is clearly larger than at the beginning, indicating that the initial ocean states are not that much separated. This requires a deeper discussion.

We rewrote this part and included more discussion on the initial state of the ocean variables.

‘It was part of our experiment design to start the ensemble simulation from an ocean state representing 6th century conditions and create the model’s ensemble spread by perturbing the atmosphere. While this was done to include a proper representation of the forcing history in the previous decades, it does not allow us to investigate the effects of different ocean initial conditions (for this, an ensemble of past2k simulation would have been needed, which was not available and not feasible). It is therefore possible that the ensemble spread is underestimated. Another choice would have been to start the model integration from different states of the AMOC (Pohlmann et al., 2004). An inspection of the AMOC time series, however, (Fig. A4 d) reveals that the AMOC variations do diverge quite rapidly after the start of the experiment and show a similar range at the time of the first major eruption compared to the end of the experiment. The only quantity where we see an increase of ensemble spread throughout the experiment is the global ocean heat content.’

24. ‘For the model - tree-ring comparison, the model temperature anomalies were taken for grid cells corresponding to the latitude/longitude range for the tree-ring locations. In Fig. 5 the comparison for the NH, the Alps, Altai and Northern Scandinavia are shown.'

Again, it is unclear how the reconstructed NH temperature has been calculated.

The explanation of the reconstructed temperatures from tree-rings is addressed in comment 9 above.

25. ‘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 and the timing of the peak cooling after the four large volcanic eruptions agree very well. ‘

I guess the authors mean that the reconstructed NH temperature fall within the model ensemble spread ?

The comment has been taken into account and the text has been revised accordingly:

‘The reconstructed NH temperatures fall within the spread of the model simulations...’

26. ‘Figure 5b shows the model - tree-ring comparison for Northern Scandinavia (NScan). Just as for the NH, the variability of the model simulations fall within the variability of the tree-ring temperatures. ‘

see previous comment

The individual tree-ring sites have been taken out of the main manuscript and are shifted to the appendix, where we have changed the text as in the comment above.

27. ‘agree very well both in timing and in signal. This could be because the tree-ring data for Fennoscandinavia consists of MXD data, so there is less time lag and smoothing in the signal (Esper et al., 2015). More deviation is visible for the ensemble mean peak cooling for the 574 and 626 CE eruptions.’

This is an example of, in my opinion, cherry-picking results. It seems that for some of the eruptions the MXD data agree better with the model because the wood density proxies better represent the true temperature. However, this is not the case for the other two eruptions. Why?

We now use the Büntgen et al. 2021 tree-ring reconstruction ensemble for our model - tree-ring analysis, see our general reply above. Because of this, the figure has changed and we now focus on the Northern Hemisphere ensemble tree ring compilation in the main manuscript.

28. ‘because the volcanic forcing in the model is overestimating the cooling in the mid-latitudes. ‘

because the prescribed volcanic forcing is too strong

This sentence was part of the tree-ring section that has been rewritten and the sentence is now removed.

29. ‘The concept of a LALIA period was raised by Büntgen et al. (2016), based on tree-ring data. There is a good agreement between the tree-ring temperatures and the model temperatures after normalization. This was done with regard to the time'

In my understanding, there is no normalization involved here. The records have just been re-aligned to a common mean, but they have not been re-scaled to a common standard deviation.

Thanks, agreed. We have rephrased the ‘normalization’ and do not emphasize it.

In addition, we have decided to use the 0-1850 CE mean to re-align the data so that the same reference period is used throughout the entire manuscript.

30. ‘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. Our model resolution is too coarse to fully capture the topography’

This is speculation. It needs to be more strongly supported.

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 (Sigl et al. 2015) agree with the

tree-ring records from the Alps (Büntgen et al. 2017), and so the spatial variability of the LALIA is something to be investigated in future studies. This has been added to the discussion section.

31. This comment may be more a matter of taste but I find that the Summary and Conclusion section is rather repetitive of what has been just exposed in the Results sections. The summary, if the authors wish to keep one, can be considerably shortened - there is no need to repeat all results in detail again. On the other hand, the real conclusions, starting in line 443, could be more sharply written e.g. that none of the simulations of the ensemble reproduce a century-long cooling, but that this could be due to a too narrow choice of ocean initial conditions, that the most important feedbacks mechanisms for multidecadal cooling involve sea-ice cover. and that the findings here do not agree with some of the palaeoclimatological data, but do agree with other dendro data .

As suggested, the summary and conclusion has been shortened with fewer details, and the section has been rewritten.

References

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.

Iles, C.E., Hegerl, G.C., Schurer, A.P. and Zhang, X., 2013. The effect of volcanic eruptions on global precipitation. *Journal of Geophysical Research: Atmospheres*, *118*(16), pp.8770-8786.

Pohlmann, H., Botzet, M., Latif, M., Roesch, A., Wild, M. and Tschuck, P., 2004. Estimating the decadal predictability of a coupled AOGCM. *Journal of Climate*, *17*(22), pp.4463-4472.

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.

Timmreck, C., Toohey, M., Zanchettin, D., Brönnimann, S., Lundstad, E. and Wilson, R., 2021. The unidentified eruption of 1809: a climatic cold case. *Climate of the Past*, *17*(4), pp.1455-1482.