Reply to reviewer 1:

Comments of reviewer 1 are provided in *italic blue*, our corresponding replies are in black, and the text to be revised or added in the manuscript is highlighted in red.

Summary: The study analyses the response of volcanic eruptions in a set of climate model simulations, comparing them with proxy-based temperature and circulation reconstructions, over the past millennium. The study distinguishes between tropical and extratropical eruptions in the Northern Hemisphere. The main results are, in my interpretation, that the qualitative response is, first, strongly dependent on whether or not the eruptions have been tropical; it also strongly depends on the model, and on the forcing data set of past volcanic aerosols.

The study is motivated by a series of prior publications that also target the response of the atmospheric circulation to volcanic eruptions, which have found somewhat contradictory results. Whereas it seemed clear that the reconstructions did show a shift towards a positive NAO state and thus an Eurasian warming after eruptions. The modelled response was not that clear. Some studies, e.g., Tejedor et al. and Polvani et al., suggested that there was no response at all, and that the signal from the reconstructions, at least at the surface level, is just random internal variability.

This study offers an alternative interpretation, suggesting that the unclear signal derived from simulations may be due to differences among models and to differences in the volcanic aerosol data sets used to force the models. They find that the more recent forcing data, along with a focus on the most realistic models, allow them to identify a consistent signal, particularly for tropical eruptions.

Recommendation: I found the study interesting, and the manuscript is well-written. It can be somewhat controversial, as it does not agree with those previous studies, but I am happy to recommend it for publication. I have a few suggestions that the authors may want to consider:

Response:

Thank you for the positive evaluation regarding our manuscript. Below, we provide our point-by-point responses.

1) *If my interpretation of the framing of the study is correct, I would suggest stating more clearly in the abstract the conceptual links to those prior analyses, explaining briefly why this study is important. The current abstract sounds correct, but it does not clearly convey the present backdrop and where this study fits.*

Response:

Thank you for bringing it up. We agree that the abstract can be improved. We will revise it as follows:

Abstract. Large tropical (TROP) volcanic eruptions can influence North Atlantic climate by inducing a positive shift of the North Atlantic Oscillation (NAO), typically resulting in winter warming across northern Eurasia. However, these changes remain highly uncertain, as they may coincide with strong internal variability in Northern Hemisphere wintertime climate. In contrast, Northern Hemisphere extratropical (NHET) eruptions are proposed to have opposite impacts, but they have been comparatively less studied, and large uncertainties remain regarding the ability of climate models to capture volcanic responses. This study examines winter North Atlantic climate responses to TROP and NHET eruptions by comparing temperature and atmospheric circulation patterns from last millennium simulations with multiple proxy-based reconstructions. We find distinct differences in NAO-related climate changes in reconstructions, with TROP eruptions followed by a shift towards positive NAO and NHET eruptions associated with a negative NAO. In comparison, modelled responses exhibit a wide spread with strong dependence on the choice of volcanic forcing dataset. Notably, simulations using the latest volcanic forcing data show improved agreement with reconstructions, particularly for TROP eruptions. This model-proxy agreement provides a useful basis for investigating the mechanisms that drive positive NAO responses after TROP eruptions. However, the simulated impacts of NHET eruptions are less consistent and remain unclear. These results highlight the importance of improved volcanic forcing datasets, refined paleoclimate reconstructions, and robust statistical approaches to better constrain uncertainties in assessing volcanic impacts on North Atlantic climate.

***2)** In my opinion, one key figure of the manuscript is Figure 5 (also Figure 6). However, I had to stare at Figure 5 for a long time to fully grasp the message. First, it won't be easy for many readers to see all the details. The purportedly grey dashed lines showing the sigma and 2xsigma bounds can barely be discerned; actually, what the reader sees are other grey dashed lines marking the zero anomalies for both axes. Additionally, the zero grey lines are not visible in all panels, and the reader may wonder if there is a hidden meaning behind this omission. The explanation of the circle colour is also left to the figure inlet in the first panel, and it is not mentioned in the caption. The caption also states that the circle sizes, displaying the magnitude of the eruptions, are normalised to the Samalas eruption. Thus each panel should have one largest red circle of the same size (the Samalas eruption). However, this does not seem to be the case for all panels, e.g., not for MPI-ESM-P CEA or ACCES evolv2k.*

My recommendation is that the authors give further thought to this complex yet important figure.

Response:

Thank you for the advice on Figure 5. We will change the patterns for the zero anomalies on both axes, as well as the $\pm\sigma$ and $\pm2\sigma$ bounds, to improve visual distinction. Also, we will move the legend to the top and add corresponding explanations in the caption. In addition, we will correct the issue where the axis ranges were not properly set, which caused the circle representing the Samalas eruption to be missing in some subplots.

As further suggested in the following comment 3) on marking the Pinatubo eruption in Figure 5, we will include the temperature anomalies calculated from the 20th Century Reanalysis version 3 (20CRv3), marked with a green star in all subplots, to serve as a reference for the Pinatubo eruption. Because the Pinatubo eruption is not covered in the PMIP last millennium simulations (the main focus of this manuscript) but in the historical simulations (which have already been discussed in detail in other model comparison studies), we think that employing reanalysis data could provide a more robust reference here. This can also reveal the model biases mentioned in previous studies (e.g., Driscoll et al., 2012) that CMIP models tend to overestimate the tropical cooling following tropical eruptions.

Our revised version of Figure 5 is shown below.

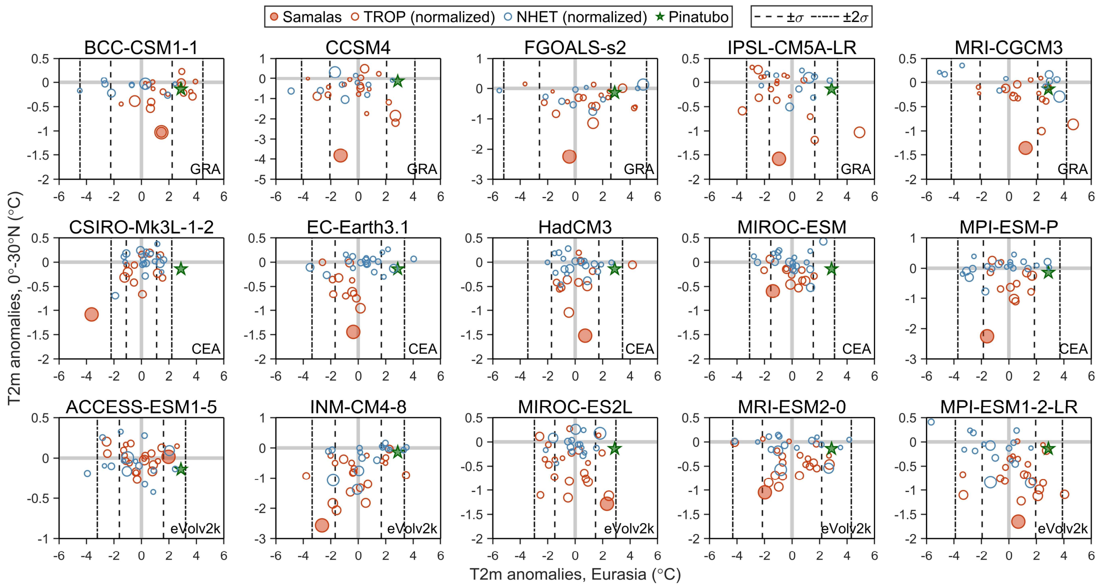

Figure 5. Near-surface air temperature (T2m, unit: °C) anomalies during the first winter after each eruption in PMIP last millennium simulations. Each subplot shows the T2m anomalies over 0°–30°N and northern Eurasia (55°N–70°N, 10°E–120°E) relative to the mean of the five years preceding the eruption. Tropical (TROP) eruptions are shown as red circles, and Northern Hemisphere extratropical (NHET) eruptions as blue circles. The sizes of the circles represent the magnitudes of the eruptions which are normalized to the Samalas eruption (the filled red circle in each subplot). The filled green star denotes the Pinatubo eruption based on the 20th Century Reanalysis version 3 (20CRv3). Dashed lines indicate the ±σ and ±2σ ranges of T2m anomalies over northern Eurasia during the last millennium.

3) *One of the previous hypotheses is that there is no NAO response or Eurasian temperature response. This study retorts that the choice of model and forcing data is, or can be, important, and that the signal may have been smeared out by different model responses and differences in the forcing data. One way to contribute to clarifying this question is to look at a one-model simulation ensemble. If the MPI-ESM-P model, according to the authors, is one of the more realistic models in this regard, would it be possible to examine the ensembles of historical simulations with this model for the Pinatubo eruption? What is the spread in that ensemble? Marking the Pinatubo eruption (perhaps with green colour) in the panels of Figure 5 would also help.*

Response:

Thank you for this suggestion. Our results indicate that the MPI-ESM1-2-LR model, with the latest PMIP4 volcanic forcing, demonstrates overall good performance. This model has 50 ensemble members from its CMIP6 historical experiment (1840–2014), which allows us to examine whether it can reproduce winter Eurasian warming after the Pinatubo eruption and to assess the spread across ensemble members. We also examine the 1883 Krakatau eruption, as previous studies (e.g., Bittner et al., 2016; Zambri and Robock, 2016) have suggested that CMIP5 models are able to capture winter warming when only consider these two largest tropical eruptions after 1850. To further support our findings, we will add a new figure (Figure 8, shown below) in Section 3.4 together with the following text:

Overall, the MPI-ESM1-2-LR model demonstrates skill in capturing volcanic impacts on the winter North Atlantic climate in its last millennium simulation with the PMIP4 volcanic forcing, particularly after TROP eruptions. To further assess its performance, we analyse 50 ensemble members from its historical simulation (1850–2014) with CMIP6 forcing to examine post-volcanic SLP and T2m changes. This evaluation focuses on the two largest TROP eruptions since 1850 (the 1991 Pinatubo and 1883 Krakatau), as previous studies have shown that CMIP5 models can reproduce a strengthened Northern Hemisphere polar vortex and associated winter warming when only considering these two events (Bittner et al.,2016; Zambri and Robock, 2016). The ensemble means of the MPI-ESM1-2-LR historical runs exhibit a clear positive NAO pattern and Eurasian warming after both eruptions (Fig. 8). Nearly all members simulate post-volcanic tropical cooling signal. However, due to the strong natural variability of the Northern Hemisphere winter climate, as previously shown in Figure 3, the T2m anomalies over northern Eurasia display a large spread across the 50 ensemble members. Despite this wide spread, more members show Eurasian warming than cooling, and 17 members after Pinatubo and 13 members after Krakatau exhibit a warming signal that exceeds one standard deviation of their temperature variability. These results further indicate that, although a large spread exists, the MPI-ESM1-2-LR model performs well in reproducing post-volcanic wintertime climate responses.

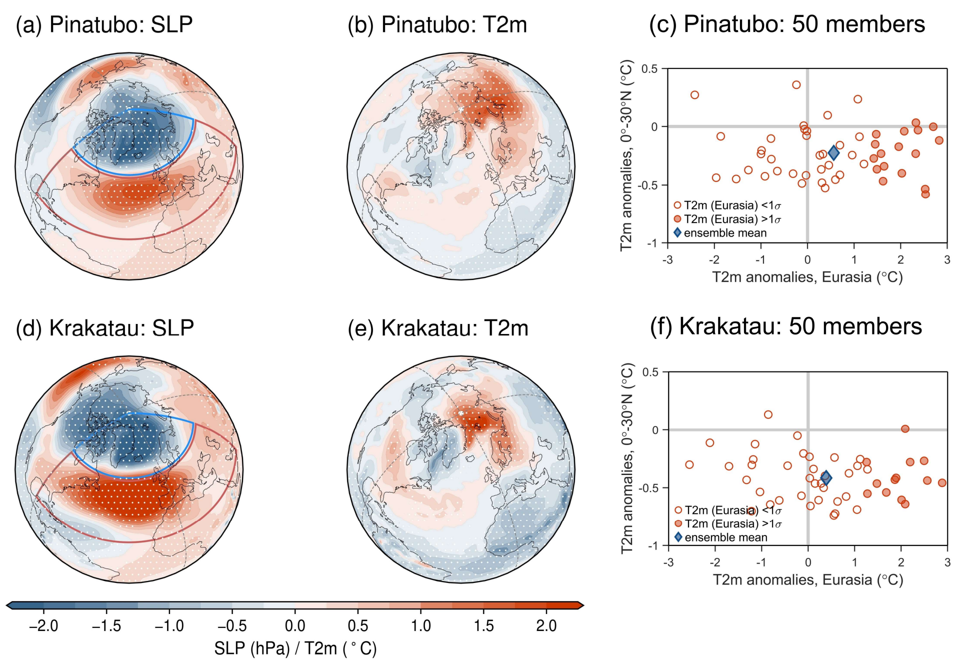

Figure 8. (a) Sea level pressure (SLP) and (b) near-surface air temperature (T2m) anomalies in the 50-member ensemble mean of the MPI-ESM1-2-LR historical simulation after the 1991 Pinatubo eruption. White dots indicate regions where more than 30 of 50 members agree on the sign of the anomalies. (c) T2m anomalies over 0°–30°N and northern Eurasia (55°N–70°N, 10°E–120°E) relative to the mean of the five years preceding the Pinatubo eruption, with red circles denoting individual members and the blue diamond denoting the ensemble mean. T2m anomalies over northern Eurasia exceeding one standard deviation (σ) are filled with color. Panels (d–f) are the same as (a–c) but for the 1883 Krakatau eruption.

***4)** The resolution of the figures needs to be improved. Many of them are multiple panels of relatively small size. A finer resolution would help the reader*

<u>Response:</u>

Thank you for pointing out this issue. We will replace all figures with higher-resolution versions.