Reply on RC2
Davide Zanchettin et al.

We thank Referee #2 for her/his helpful comments on our manuscript. We report in italics relevant comments by the referee.

Concerning the specific discussion points raised by the referee:

I understand the benefits to build a multi-model protocol in which the volcanic forcing is commonly defined, to allow a "good agreement between the different models » and highlight the differences between the models in terms of response to external forcing independently to the implementation of the forcing. Overall, is there any risk that the homogenisation of the climate models encouraged in model inter-comparisons support the building of a unique family of similar models, all of them showing the same uncertainties that would be therefore more difficult to estimate? Should we encourage more contrasted model developments for a better understanding of the processes at play?

This is a very useful comment. The key point of using a consistent forcing across models in terms of aerosol radiative properties is explained in the VolMIP original paper (Zanchettin et al., 2016). In brief, this allows us to focus on how models differ in the climate response, as uncertainties generated by aerosol chemical and microphysical properties are neglected. In fact, in this sense VolMIP is a companion to the SPARC/SSiRC Interactive Stratospheric Aerosol Model Intercomparison project (ISA-MIP, Timmreck et al., 2018)) which covers the uncertainties in the pathway from the eruption source to the volcanic radiative forcing. Specifically, the aim of ISA-MIP is to constrain and to improve global aerosol models by using a range of observations in order to reduce the forcing uncertainties. Both initiatives are supposed to interact in order to progress our understanding of how climate responds to strong volcanic eruptions. Of course, model agreement does not necessarily imply agreement with observations, and we will include a comparison with observed anomalies to illustrate any potential bias in our multi-model ensemble;

Could we expect an impact of the anthropogenic forcings on the climate response to volcanic eruptions? In other words, would we expect different conclusions starting the volc-pinatubo-full experiment from control experiments produced with constant anthropogenic forcings corresponding to those observed at the beginning of the XXIth century, and/or in transient forcing experiments?
Several studies point to the fact that background climate conditions can affect the climate response to volcanic forcing. We consider the volc-pinatubo experiments idealized exactly because the forcing is realistic but the background climate state (from piControl) is different from the actual climate state during the 1991 Pinatubo eruption. In the original manuscript, we even illustrate some details about the mean climate state and variability in piControl simulated by the different models as a possible source of inter-model differences. Therefore, we agree with the referee that conclusions might differ if the background climate state differs and/or in transient conditions, i.e., in presence of additional forcing agents. We will better stress this in the revised manuscript.

P4, L.125-130: How do we deal the fact that the modes of variability typically show different patterns for the different models? Why not having considered an EOF approach, using for each model its mode of variability with its specific pattern?

We dealt with it in the definition phase of the VolMIP protocol, when we choose to use box-based indices over EOF-based ones because our expectation is that total variability is separated differently into principal components in different models, which would add another level of uncertainty. We agree that there is an intrinsic problem in the use of predefined indices based on mathematical constructs (being EOF or box-based indices) rather than physical understanding and will expand discussion in this regard in the revised manuscript.

P5, Characteristics of volc-pinatubo-full, the multi-model ensemble: Would it be possible to give more details about the spectral resolution of the models? Does it differ among the models? The way that the VOLMIP forcing is distributed over the spectral bands could be detailed. More information about the vertical distribution of the forcing in the different models would be also welcomed: is the forcing vertically distributed in a stationary way, using monthly climatologies of the elevation of the atmospheric layers, or is the forcing vertically distributed on-line? I saw that this weakness is discussed at the end of the paper, but why not including more information about these model features in this publication?

We will improve the description of all models in the revised manuscript. Spectral bands differ across models, and EVA produces forcing input data for each model’s specifics.

Initial conditions: Why not sampling the QBO on a similar way as the other modes? QBO impact on climate response to volcanic forcing has been evidence in Thomas et al., 2009. (Thomas, M. A., Giorgetta, M. A., Timmreck, C., Graf, H.-F., and Stenchikov, G.: Simulation of the climate impact of Mt. Pinatubo eruption using ECHAM5 – Part 2: Sensitivity to the phase of the QBO and ENSO, Atmos. Chem. Phys., 9, 3001–3009, https://doi.org/10.5194/acp-9-3001-2009, 2009). At lower frequency, why not considering different states of the AMV that might affect also the response of the modes of variability (Ménégoz, M., Cassou, C., Swingedouw, D., Ruprich-Robert, Y., Bretonnière, P.A. and Doblas-Reyes, F., 2018. Role of the Atlantic Multidecadal Variability in modulating the climate response to a Pinatubo-like volcanic eruption. Climate Dynamics, 51(5), pp.1863-1883). This point is discussed at the end of the article. Nevertheless, we do not know the reasons for which these modes have not been considered in the first edition of VOLMIP

We agree that the state of QBO is a potential influencing factor on the climate response to volcanic eruptions. Since not all models spontaneously generate a QBO, we decided to not include it as a requirement for sampling in the final protocol. The original VolMIP paper states that “volcanic radiative sampling of an eastern phase of the Quasi-Biennial Oscillation (QBO), as observed after the 1991 Pinatubo eruption, is preferred for those models that spontaneously generate such mode of stratospheric variability.” We will report this more explicitly in the revised manuscript. Concerning the AMV, this is certainly of
interest. The problem is how the increase in the number of variables used for the sampling affects the ensemble size. We need a balance, and for the short-term scales that are focus of volc-pinatubo experiments we decided to opt for the NAO as a descriptor of the North Atlantic state. For the volc-long VolMIP simulations with the focus on the multiannual to decadal scale climate response to volcanic forcing we use the AMOC as reference index for sampling initial conditions. We will elaborate further on this in the revised manuscript, including a perspective on the implications of the coupling between AMV and NAO/AMOC suggested in the literature.

P10: the ENSO differences among the models based on a temperature average over the Niño 3.4 area only might be affected by the ENSO specific position in each model. The ENSO signature in models is often shifted Southward/Northward Eastward/Westward as compared to the observations, and it differs clearly from one model to another one. This could be discussed in the article. The same issue can be highlighted for the NAO signature, and this might be a much more important issue considering the typical spatial biases of the NAO pattern in the current generation of AOGCMs.

We will elaborate further on the ENSO index, also following a comment by Referee #1. As highlighted in a point response above, inter-model differences and biases with respect to observations make it difficult to identify optimal indices that are expected to capture specific dynamics. We will discuss this better in the revised manuscript. At least for the NAO a recent paper using the same index definition employed in VolMIP suggests a marked consistency across CMIP6 models (Cusinato et al., 2021).

P13, L. 402: "dynamical responses may be masked by broad tropical radiative cooling effects " -> So why not considering a relative ENSO index (Nino3.4 tas minus tropical tas) as done in several publications (e.g. Khodri, M., Izumo, T., Vialard, J., Janicot, S., Cassou, C., Lengaigne, M., Mignot, J., Gastineau, G., Guilyardi, E., Lebas, N. and Robock, A., 2017. Tropical explosive volcanic eruptions can trigger El Niño by cooling tropical Africa. Nature communications, 8(1), pp.1-13.). This is discussed in the end of the article, but why not including directly such a "RENSO index" in the article?

We will use relative SSTs in addition to absolute SSTs to calculate the Nino3.4 index in the revised manuscript. We would like to remark here that the choice of using absolute SSTs was based on the original VolMIP protocol using it. We strongly believe that ENSO deserves a follow-up specific study beyond the initial results considered here, that are mostly focused on assessing the effectiveness of the VolMIP protocol.

P13-14: feedbacks: more explanations about the LW and SW ratios would be welcomed, to allow a better understanding of the sign of the feedbacks (negative versus positive) as well as the processes that are suggested in this Section. It is delicate in particular to understand whether the LW and SW changes are related to aerosol changes or to changes in the atmospheric temperature.

We will improve the discussion about LW and SW ratios in the revised manuscript. This analysis is and will remain nonetheless only preliminary for a study using also the volc-pinatubo-strat/surf experiments to fully disentangle the feedbacks involved in the response.

We will account for all technical corrections requested by the referee.

References