Reviewer 2:

This article presents results from a new set of decadal forecast models which allows the authors to isolate the effect of the three most recent large volcanic eruptions on climate. The analysis is interesting and very clearly presented, the paper is well written and methodology seems sounds – so I think that this article would be worthy of publication is in this journal.

I have two major concerns.

The first is due to the small sample size (only three eruptions) and the comparison to the observations. While the large ensemble sizes allow for proper verification of model results, it is not clear to me how best to interpret the comparison to observations. It the signal to noise ratio is low (as it presumably is for regional results). Including a volcano could improve skill in general while degrading the fit to certain aspects of climate in a specific case. Equally the models could have no skill while by chance a better fit to observations could occur. I think more discussion regarding this is needed in the context of predictability, so that the conclusions are clear and the results are not over interpreted.

Reply: We agree with the reviewer that while three volcanic eruptions is a small sample size, we are limited by the observational record. Despite the small sample size, previous studies have shown the improvement (and deterioration) of the predictive skill associated with the volcanic forcing based on retrospective forecasts for the last several decades (e.g. Timmreck et al. (2016), Ménégoz et al. (2018), Wu et al. (2023)). With our protocol for three selected initialization dates right before the eruptions, we cannot estimate the impact of the volcanic forcing on the predictive skill, so our approach has been therefore to perform a qualitative comparison of the two alternative forecasts (with and without volcanic forcings) with the observations, to see in which cases adding the volcanic forcing improves the agreement with observations. The analysis of the DCPP-A and DCPP-C hindcast differences motivates such comparison since the post-volcanic cooling is significant and therefore should be detectable in the observations. We agree with the reviewer however that some of these apparent improvements, in particular those related to dynamical variables, might occur by chance, a caveat that we now explicitly mention in the conclusions sections.

The second is regarding novelty. Certainly, all findings do have novelty as they represent results from a new set of model simulations. However, much of this article details changes in aspects of the climate by volcanic eruptions which are already well documented in many other studies - ENSO, NAO, winter warming, AMOC etc as outlined in the introduction. I therefore think that more should be added to put them into the context of previous studies. What is new and what confirms or contradicts existing work?

Reply: We think that in the results section we have included many references to previous studies contextualising the results, but we agree that in the abstract and conclusions section we should have been more precise and state which results support previous studies and which results are new. We have re-written both to improve these aspects and insist on the fact that we are in a predictive mode, which differ from a number of former studies about volcanic eruption impact

Minor points:

Line 127-128. “on” the 30th October in both cases

Reply: Corrected.

Paragraph L130. Would it be possible to include some discussion about the uncertainty in the volcanic forcing. Agung in particular is pre-satellite, so are the uncertainties larger? I note that the reference given only seems to covers the satellite era.

Reply: There is no peer-reviewed publication documenting the CMIP6 volcanic dataset, and therefore no detailed description of such uncertainties. We have nonetheless cited the very short documentation provided by Luo (2018), who prepared this volcanic forcing dataset. Within this documentation, Thomason et al. (2016) is referenced for further details. We do indeed expect the uncertainties in the volcanic forcing of Agung to be larger since before 1979 the forcing dataset was produced by the AER-2D model (Arfeuille et al., 2014). While it is important to consider this uncertainty when making conclusions, it is beyond the scope of this study to quantify it. We have included a comment in the paragraph.

Line 143 – could you give more information regarding the ozone depletion as this isn’t that clear to me.

Reply: Previous studies have shown that changes in stratospheric circulation and temperature in response to volcanic aerosols can affect stratospheric ozone photochemistry, and therefore result in ozone changes (e.g. Stenchicov et al., 2002; and references within). We simply note that these processes which lead to ozone depletion are not simulated in these climate models. Also, note that climate-chemistry models have not yet been used in climate prediction.

L155-156 – were these the infilled version of HadCRUT5?

Reply: Yes, it is the HadCRUT5 analysis gridded data ensemble mean.

Figure 8 – could you specify what period the anomalies are respect to?

Reply: It is stated in the figure caption: ‘The anomalies have been computed with respect to the period 1970-2005.’