Our replies to the referees comments are *in italic* below each comment.

REPLIES TO REFEREE # 1:

This paper has a fundamental error, in that it ignores new work that shows that SSTs matter when it comes to winter warming. Coupe and Robock (2021) showed that large ensembles using state of the art models do not produce robust winter warming when run with a coupled ocean model, as current models do not robustly simulate El Niño following large eruptions. But they showed that if SSTs are specified, then every ensemble member simulates winter warming after the three largest recent eruptions of the 20th Century. That's 60 ensemble members, and they all get it.

> *We thank Dr. Robock for his careful reading of our manuscript. We* totally disagree *with his statement that our paper "has a fundamental error," which we rebut below. We thank him for the minor corrections and suggestions.*

Lines 68-69: The authors ask, "if Pinatubo and Krakatau are not large enough, how large does an eruption need to be to cause wintertime surface warming at Northern mid-latitudes?" I think this is the wrong question. The correct question is, "What other factor is missing from all these previous studies that prevents them from simulating what was actually observed after past large volcanic eruptions?" As Coupe and Robock (2021) have shown, it is the confounding influence of El Niño. If there is no El Niño, then erroneous tropospheric forcing destroys the stratospheric circulation forced by the volcanic eruptions. Stratospheric forcing by itself is not enough, and the present paper just reinforces that, and is not a new result.

> *First, we note that in their pioneering study, Robock and Mao (1992)* specifically subtracted out the ENSO *signal to isolate the volcanic response. In fact, most modeling studies have been performed with ENSO-neutral conditions. So we are following a well trodden path, setting aside the ENSO question, to avoid confusing things unnecessarily.*
>
> *Second, we agree with the referee that the question of whether El Niño conditions do or do not alter the Eurasian winter response to volcanic eruptions is potentially interesting. But, the referee will agree, it is a* separate *question from the one we are addressing in this paper. In fact, is is a* secondary *question. Before addressing that question, we need determine whether volcanic eruptions – over a* much broader *range of amplitudes than those analyzed in Coupe and Robock (2021) – are able to produce a surface winter response* without additional factors. *This is the* primary *question.*

The referee will recognize, we hope, that our paper is focused on answering that primary question. Therefore, we deliberately *chose to start all runs in an ENSO-neutral state, to isolate the volcanic from the response to ENSO, as we explain in the paper. We plan to address the secondary question too, and a follow-up paper on that is in preparation.*

In any case, to clarify all this, in the revised manuscript we have added a paragraph to the last section to discussing the potential impact of ENSO.

On line 162, they say, "to avoid unnecessary confusion, therefore, we solely focus here on ENSO-neutral eruptions." Rather, it seems that this also removes addressing an important scientific question.

This is already addressed above.

The author's current results just replicate what Polvani et al. (2019) have already found. I recommend major revisions, which would involve using AMIP runs to see whether the new model they use can also simulate winter warming when surface forcing from inaccurate SSTs is removed.

Our current results do not *"just replicate" those of Polvani et at. (2019): in that paper, only the Pinatubo eruption of 1991 was explored (and the smaller El Chichón 1982 eruption in the supplementary material). In the present manuscript, we explore a very wide range eruption magnitudes, from Pinatubo to Tambora to Samalas, and well beyond. Perhaps the referee meant to say that "the current findings confirm and extend those of Polvani et al (2019)"?*

As for recommending "major revisions": the referee is again focused on the ENSO questions, and is asking us to write a different paper than the one he was asked to review. Doing AMIP runs is of little help in addressing the primary question at hand. Consider this: in a recent paper very similar to ours, Azoulay et al. (JGR, 2021) have addressed the same question as we do here, and they perform no AMIP simulations, nor mention the potential impact of ENSO (since it is a secondary question).

Lines 14-19: These lines in the abstract need to be qualified. They should include that the results depend on the SST specification, that is with no El Niño.

There is really no need for us to qualify the statement in those line. The "potential" impact of El Niño has been ignored by just about every single study until 2021, and the only suggestion for its existence is the Coupe & Robock (2021) paper, which only analyzed at a single model. Furthermore, the claims in that paper not – to date – been replicated or validated by any other study. Such tenuous evidence, therefore, does not warrant a change in the abstract of our paper.

Lines 30-33: This is not correct. First, it does not matter how many papers there are. What matters is how they were done. Second, Zambri and Robock (2016) did find winter warming after large eruptions, and showed the errors of previous studies, such as by Driscoll et al. (2012). Third, the claim that Bittner et al. (2016) supports their claim is wrong. As Zambri and Robock (2016) wrote, "In contrast to the findings of Driscoll et al. [2012], however, Bittner et al. [2016] showed that for the largest eruptions, the CMIP5 ensemble does produce a robust strengthening of the polar vortex." Fourth, the authors have ignored the paper by Zambri et al. (2017), which also showed that past model simulations did produce winter warming.

> ***First.*** *It* does *matter "how many papers there are": it is called* the weight of the evidence. *It is a valuable in science as in a court of law. Our readers would want to know how many recent papers have claimed that Pinatubo and similar eruptions cause a winter warming, and how many papers have claimed the contrary.*
>
> ***Second.*** *We agree that Zambri & Robock (2016) claimed to have detected a winter warming in their paper. This is why we have carefully phrased our sentence and say "the vast majority of the later studies" and not "all studies."*
>
> ***Third.*** *We suspect the referee might be confusing Bittner et al. (2016), which is not cited in that paragraph, with Bittner (2015), which we cite. In the latter, which is Bittner's PhD thesis, it is clearly shown that, even with 100 members, no winter warming is found in their model. That crucial fact, somehow, was left unmentioned in Bittner et al. (2016), which only discussed the strengthening of the polar vortex, while failing to inform the reader that no surface warming was found in their model.*
>
> ***Four.*** *Again, we are aware of the Zambri et al (2017) claiming a post-eruption winter warming. However, we are not, in this section, writing an exhaustive review. We are simply citing a few representative studies to illustrate the current state of affairs. For the sake of completeness, in the revised version we have added a footnote citing Zambri & Robock (2016), as that paper pertains to the post-1850 eruptions which are discussed in that paragraph.*

Lines 245-250: Please explain how the NAM index was calculated and how it can explain more than 75% of the variance. On what time scales? Are you talking about daily, monthly, seasonal, or what?

> *Our computation of the NAM is explained in great detail in Section 2.6. It is based on monthly mean model output of the zonal mean zonal wind. The method is standard.*

There are multiple cases of excess verbiage, e.g., "notice," "we remind the reader," "we draw the reader's attention to," "we note that." These should all be deleted. Every sentence should be important or it should not be in the paper. Does this imply that other sentences should not be noticed? Such writing style should be avoided in scientific articles.

> *We hope the referee will do us the courtesy of not interfering with our writing style. That* "such writing style should be avoided in scientific articles" *is the personal opinion of the referee, which we respect. We merely ask for the same respect in return.*

Lines 410-412: "early claims of robust Eurasian winter warming for eruptions such as Pinatubo – and even smaller ones, such as the 1982 El Chichón or the 1962 Agung eruptions – simply cannot be reproduced with current-generation climate models: these have consistently failed to show any warming for such historical eruptions" is simply wrong. You have to qualify this claim by conditioning it on the models only being forced from the stratosphere. In AMIP runs, current models have done an excellent job.

> *The cited sentence is* not *wrong, and we do* not *need to qualify the claim. One more time, the referee wishes us to cite the Coupe & Robock (2021) study, where a AMIP runs from one single model were analyzed. We see no reason to cite that paper here again, as we have already cited in Section 2.3, and we cite it again in the last section.*

Figs. 1, A1: What are the x- and y-axes for each panel? 0 0 0? 0 20 10? 0 45 0? Give the variable and the units for each axis, and mark them with numbers that make sense.

Fig. 2, A2: There are no units given for temperature or wind.

Fig. 2, A2: What are the x-axes? 0 45 0? What does this mean?

Fig. 2, A2: What are the y-axes? They are missing the variable and the units.

Fig. 3a: What are the x- and y-axes? Give the variable and the units for each axis.

Fig. 3, all panels: What are the y-axes? They are missing the variable and the units.

Figs. 5, 6, 7, 8: What are the units for temperature?

> *We are sorry for the confusion caused by these missing items. They were all* accidentally removed by the ACP typesetting process, *which is beyond our control. We did not realize this had happened, otherwise we would have alerted the editor for corrections.*

Please also address the 14 comments in the attached annotated manuscript.

> *We have addressed all those comments. We appreciate the referee's attention to detail.*

REPLIES TO REFEREE # 2:

1. General comments:

This ms illustrates the importance of internal variability for winter warming following volcanic eruptions, and that only for very large eruptions or an average of a very large number of eruptions the signal is expected to be significant. It is based on a largish ensemble of climate model simulations with somewhat idealized conditions. The ms describes a well conducted study, and the figures illustrate nicely how a highly significant stratospheric warming leads to significant zonal wind enhancement which, however, only sometimes reaches the surface and has to compete with strong variability there. It is a very valuable addition to the literature. I recommend a few suggestions for discussion/consideration.

We are grateful to the referee for the kind words.

a) the ensemble explicitly focuses on ENSO neutral conditions in eruption years. i find this a bit surprising and constricting - and it would be interesting to have seen if results vary between ENSO states. Alternatively, it should be flagged in abstract that this study refers to ENSO neutral start dates

We agree that a clarification would be helpful, so we have added a paragraph new to the discussion section on the question of ENSO, to make it clear why we are only focusing on the ENSO-neutral case here.

b) even if the mean change is rather subtle, this could affect the tails of the distribution - would a strong winter warming be more likely with than without a preceding eruption? (the event attribution question)? Based on figure 8, the ensemble size is probably too small for a robust answer... although it would be really interesting to know.

Nice idea but, as the referee avows, our ensemble size is just too small to quantify this.

c) i would have liked to see a bit more discussion of the observed response to Pinatubo, for example, where it sits compared to the model simulations. If it was made clear that it is within the range of what the model simulates that would have made the possibility less likely that the observations behave differently from the model. This possibility has been raised, for example, also with the recent Scaife et al results on predictability of the NAO where the simulated signal is much smaller than observed.

The observed post-Pinatubo warming does fall within the range of model simulations. We believe this is a well known fact. For instance, we showed this clearly in an earlier *paper, which was entitled:* Northern Hemisphere continental winter warming following the 1991 Mt. Pinatubo eruption: Reconciling models and observations *(Polvani et al., 2019), to emphasize the models well capture the observed post-eruption trends. In all honesty, we do not believe that more discussion is needed on this point.*

2. Specific comments:

Abstract last sentence: this isn't very clear – think you are discussing a single simulated response, and that only in rare cases will it emerge from internal variability? (this is what would be interesting to know - how likely is it going to emerge beyond 1σ i expect very similar to the expectation of this occurring by chance? in any case, please phrase more clearly waht the last sentence refers to

Thanks for alerting us that the sentence was not clear. We have rephrased it.

L 24: this is about global surface temperature?

Yes. This is clear from the context.

L 30: a reader may like to know what is meant by methodological issues

We have added a footnote with one example, listing the multiple issues in Robock and Mao (1992). We think this suffices. It would take several pages to describe the issues in all the other previous papers (and there are many); this is not the place to list them.

L 85: it would be helpful here to define what region the surface temperature of eurasia refers to

Agreed. We have added this detail to that sentence.

L 190: Deser 2012 is one of my all-time favourite papers, yet averaging across ensemble members to arrive at fingerprints of forced change has been done a lot longer than that (e.g see discussion Tett et al nature 1999)

We have great respect for Dr Tett, but nobody adopted "large ensembles" until Clara Deser's seminal work. Therefore, we would rather leave the citation as is.

Paragraph starting line 235: would it be worth mentioning where significant wind anomalies reach the surface? to me this is an interesting question

Over the North Atlantic and mostly in the zonal wind direction (i.e. affecting the NAO and the NAM). See the figure below (left panel). While this is well known, we have added a note in the revised manuscript to clarify this. Thank you for the suggestion.

L 310: it would be helpful to have some idea what makes EVA unrealistic even in main text

This is a complex question, having to do with detailed choices of aerosol properties in EVA, and how those choices differ from the ones used to construction the CMIP6 prescribed aerosols for Pinatubo. It is not clear whether EVA aerosols are actually unrealistic; more likely, the choices made for the un-documented CMIP6 volcanic aerosols are simply different from those in EVA. Since no peer-reviewed paper exists for the CMIP6 volcanic aerosols (amazingly enough!), it would be quite difficult to answer the question without a lot of detailed analysis. Hence, such discussion would be well-beyond the scope of this paper. The reassuring fact is that the EVA aerosols are in the public domain, and are being used by others (e.g. the Azoulay et al paper cited above), so modeling studies can be directly compared.

Figure 5: shading - does it refer to lack of significance for the averaged response right?

Indeed. We have clarified this in the caption. Thank you for noting that.

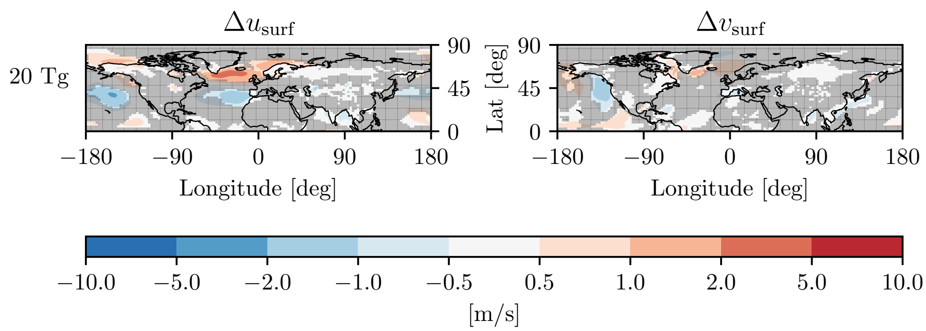

Figure 1: The response in zonal and meridional surface winds to a 20 Tg(S) volcanic eruption.