Reviewer #1 (report #2)

The manuscript is greatly improved after adding analysis on the comparison between cpl and atm-only, the impact of ensemble size and the tropospheric response. Although it's hard to verify the reliability of three SSWs in one winter, the results do show a clear difference of winter 3 compared to winter 1 and 2. And the mechanisms explored in this study should add extra value to the research topic. The language can still be improved with a proof reading. The study is worth publication after addressing a few remaining issues as commented below.
We thank the reviewer for their detailed and much relevant revision during this process that has contributed to the accuracy of this work. Here below are our response to the latest reviewer's comments in red. Line numbers refer to the track changes version of this manuscript.

L29: an potential increase? Since it's hard to verify the three SSWs in one winter.
Indeed, we have altered the sentence accordingly.

L57: "play a deciding role" really?
Since we further explain this statement in that paragraph we are confident of this wording, although we are aware that the polar vortex is not the only player as we also discuss in the text.

L187-189: Confusing sentence, better to rephrase. Is it the SSWs play a role in the atmospheric circulation response?
We agree that is can cause confusion and have changed lines 130-132 according to the following:
"This is the focal point of this study where we investigate for the first time the role of wave-mean flow interactions and SSWs in the climate response to a HL volcanic eruption."

Figure 2: It's a bit confusing to show darker color for larger anomaly for SW, but darker color for smaller anomaly for LW. Better to reverse one to keep it consistent.
Ok, we have changed back to the original version.

Figure 1 and 2: "aerosol mass" is confusing, as kg/kg is for mass mixing ratio, not mass. Please double check what's exactly used in the model, what is shown in figure 1 and 2 and use the correct text. Please also check texts written in the manuscript. Figure 1 why starts from month 5 instead of from 0?
We thank the reviewer for this well spotted comment. Indeed it is important to define between these two parameters in more details than we did. Kg/kg represents dry volcanic aerosol mass mixing ratio while kg/m2 represents average aerosol column mass. We have edited the text accordingly (L180, 183, 202) and changed figures appropriately.
Regarding Figure 1 starting from month 5, we mainly want to draw the attention towards the different profiles used for scaling the final WACCM input. Also, unlike Figure 2, Figure 1 does not show a model output. Also, since the values in months 1-4 are zero it would make the figure less clear starting the x-axis in month 1 (january).

Figure 2 and 3: If cpl runs for 4 years and atm-only runs only for 3 years, then it's clear and consistent to show four years for both and leave the 4th year blank for the atm-only. Especially, the aerosol distribution and U10 variations in the 3rd post-eruption winter in Fig. 3 should also be shown.

We do agree and have edited Fig. 3 accordingly. In addition, we took the advantge to increase the number of contours to show more details. In stead of leaving the 4th year black, we underline its different time axis in the figure caption.

We also added the following sentence (L531-533):

"The SSW development is also evident in both U10 and U50 and T50 timeseries (Fig. 3c and Figure S3a-b respectively), where peak T50 warming occurs late in winter 3."

We have also edited the following sentence to explain further the volcanic forcing used for winter 3 (L231-232):

"...,where the January and February forcing of winter 3 are defined to be a continuation of the December value of year 2"

We did not want to complicate part 2.3 but of course it is important to mention the forcing in winter 3 in more details although we do not see the need for an extended Figure 2.

L259-261: For a clear comparison, what is the total loading of SO_2 in Tg in your simulations?

We have added the following to the sentence in L294-295 for clarity:

"and the total aerosol mass of 14.04 Tg being largely confined north of 45° N (Fig. 2a-b)."

L262: the radiative forcing is SW, LW or you mean the idealized volcanic forcing?

Indeed, this now reads idealized (L295).

L427-438: Better rephrase this. Figure 7 supports that winter 3 is a less constrained response compared to winter 1 and 2, but does not necessarily mean a much larger ensemble would not provide more certainty. The results only show that the standard error will decrease (increase) when ensemble of more (less) members compared to the ensemble of 20. But How's it when create ensemble of members more than 20. And analysis in Figure 7 is based on these same ensemble members, and but with different new members results may be different?

We agree where this comment is also related to the comment from reviewer #2 and we refer to our response regarding a more detailed response. We have moved this Figure into supplementary and shorten the text that now reads (L462-467):

"We explored the impact of the ensemble size for the ensemble spread of two key diagnostics of our mechanism, namely U10 and SSW, calculated as the standard deviation of post-eruption paired anomalies for the first three post-eruption winters (Supplementary Fig. S8). Winter 3 produces larger spread than winters 1 and 2, indicative of a least constrained forced response, which is especially evident for ensemble sizes larger than 15. Therefore, only much larger ensembles may provide signals not encompassing the value of zero within uncertainty."

Figure 8: i) DJF3 not DJF2. The figure caption should be rephrased for better clarity.
Thanks, done.

L557-558 and Figure S5: It's confusing. From Figure 1 and 2, it looks winter 1 match Jan01, winter 2 match Jan02, right? Then winter 3 should match Jan03? But Figure S5 only goes to Jan02, then one cannot see which is evident in the SST.
Indeed the reviewer is correct, it is better to show winter 3, we have now updated supplementary Figure S5.

L626: "that we further confirm to occurr" better rephrase.
Thanks, this now reads (L763-764):
"…,agreeing with our results in winter 3 where an increase in the occurrence of SSWs is detected."

L648-654: Can you add a (supplement) figure to show the zonal wind and temperature timeseries of members with three SSWs in winter 3. With only the count in figure 6 and S3, it's hard to know how the SSWs look like. It's hard to know if it's realistic or not, then better to show the evolution in the paper, then if any future study shows three SSWs in one winter, this opens a chance for future comparisons.
This is related to the comment above, where we do agree and have updated both Figure 3 and Figure S3 accordingly, so now the U10, U50 and T50 time series show this evolution. We note however that the temperature pattern at 50hPa (see Figure S3) is spatially different when compared to the zonal wind, where the warming is mostly confined in part of the NH. In short, the asymmetric nature of the warming pattern causes a dampening of the response when averaged over the entire NH. Therefore the T50 timeseries are based on 70-90°N and 0-200°E compared to 70-80°N and 0-360° used for U50.

References: doi links should be added for all the references?
Ok, done (no doi found for two papers).

Reviewer #2, Report #1

I felt this study was nearly ready for publishing, although now the authors have added a problematic figure that seems to convey confidence in a large SSW effect despite too few simulations to address the extremely noisy occurrence of SSWs. If this analysis cannot be thoroughly defended, I request that the new figure and added statements professing confidence in the SSW results please be removed, in which case my review would return to being minor concerns.
We thank the reviewer for their detailed and much relevant revision during this process that has contributed to the accuracy of this work.Here below are our response to the latest reviewer's comments in red. Line numbers refer to the track changes version of this manuscript.

- The new Fig. 7 appears to be giving a misleading argument that 20 ensemble members is sufficient to validate a strong SSW signal when this is not being demonstrated. The possible iterations of <20 ensemble members converge as ensemble size increases toward the full 20-member ensemble, but not because the uncertainty is small. By n=10, nearly any two random 10-member combinations of the full 20-member ensemble will share at least a few common members, and hence convergence toward the mean of the full ensemble is expected. We really need more ensemble members to be confident of an SSW impact in Winter 3. From the noisy data in Fig. 6, which shows very difference mean SSW numbers among unforced seasons that are statistically equal, I expect a different set of 20 simulated eruption realizations could result in an entirely different conclusions, and this new analysis does nothing to change that impression. The information that matters is already presented in Fig. 6, so I don't see this Fig. 7 being anything but a distraction from the more interesting features of this study.
We thank the reviewer for noticing this. There were indeed mistakes in the caption of the figure, which illustrates changes in uncertainty of the ensemble response (standard deviation), not changes in the ensemble mean, which inevitably led to confusion regarding our interpretation. In fact, the key points we made in the main text are in line with the reviewer's interpretation: we observe an increase in uncertainty with ensemble size and especially for winter 3, which implies that, as we wrote, "winter 3 produces larger uncertainty than winters 1 and 2, suggesting a least constrained forced response" and "winter 3 features an intrinsically lower signal-to-noise ratio". The mistake in the caption was corrected, and we are confident that our statements stand. We also agree that Fig. 7 adds little information with respect to what is shown in Fig. 6, so we decided to move it to the supplement. We have also reduced the description of the results in the main text, focusing on the main implications of the analysis that now reads (L462-467):
"We explored the impact of the ensemble size for the ensemble spread of two key diagnostics of our mechanism, namely U10 and SSW, calculated as the standard deviation of post-eruption paired anomalies for the first three post-eruption winters (Supplementary Fig. S8). Winter 3 produces larger spread than winters 1 and 2, indicative of a least constrained forced response, which is especially evident for ensemble sizes larger than 15. Therefore, only much larger ensembles may provide signals not encompassing the value of zero within uncertainty."

We also rephrased the sentence in line 466-467 as follows: " Therefore, only much larger ensembles may provide signals not encompassing the value of zero within uncertainty "

- Also on Fig. 7, it seems odd that the mean among all permutations should be anything but constant as a function of sample size. Every set of permutations would be expected to have each ensemble member represented an equal number of times, and hence equal the mean of the full ensemble. So the statement that the “mean curves consistently level off for ensemble sizes” is entirely mysterious to me, and seems to make no proof of the 20-member ensemble being representative of a potentially larger ensemble.

This comment follows our mistake in the captioning and labelling of the figure (corrected in the revised version). We confirm that, of course, the analysis for the ensemble mean leads to a constant value across all ensemble sizes. Figure 1 here below (ensemble mean analysis) provides proof of this.

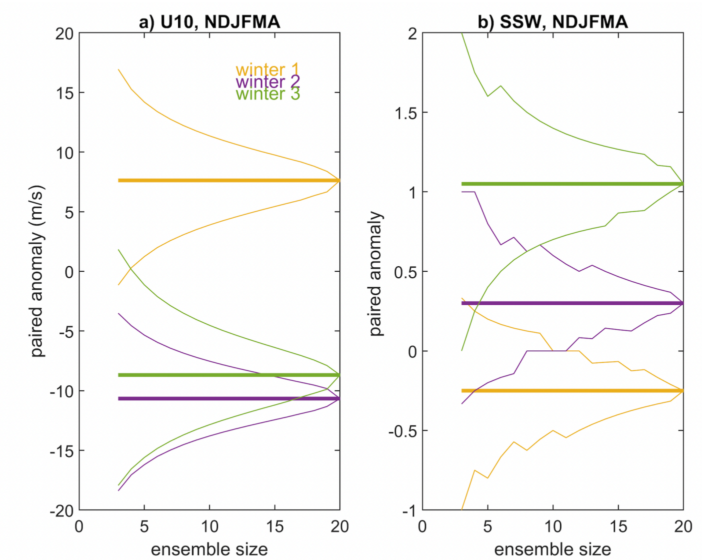

Figure 1: The same as Figure S8 but for the ensemble mean value

- The text in Lines 416-438 that accompanies Fig. 7 makes strong statements that I find disagreeable, i.e. “these results do show strong evidence of an increase in the number of SSWs” and “the full ensemble estimate is representative of uncertainty in the expected response for larger ensemble sizes”. I do not see sufficient evidence for either conclusion. I also find the wording of this section generally confusing.

line 416-417 (now 459): we have changed " show strong evidence of" with "suggest"
line 418-438: this was deleted.

- The phrase “paired anomalies” is used throughout the text but is not defined anywhere.

This we already defined in lines 246-248:
“Model output is analyzed by computing paired anomalies, defined as deviations of each volcanic simulation from the corresponding control simulation (Zanchettin et al., 2022) (volcanic minus control).”

- Of the eruption effects, Lines 28-29 of the Abstract states that "This causes a weakening of the polar vortex and an increase in the occurrence of sudden stratospheric warming events, although with a small signal-to-noise ratio". This still sounds like there is a cause-and-effect link that hasn't been convincingly demonstrated. I'd prefer if this line instead says that simulations with eruptions are found to have an unusually high frequency of SSWs and this warrants further exploration. While I'm only addressing in this review that the signal-to-noise issue has not been resolved, there are other the issues, i.e. that this is a single-model study and SSWs are notoriously difficult to reliably simulate in climate models, so I stress not to place too much confidence in the SSW results.

We agree and now this sentence reads (L29-30):

"We detect unusually high frequency of SSWs in the idealized forcing simulation using interactive ocean that calls for further exploration."

- Line 459: "winter 2 likely acts as a precursor" to "potentially acts" or "seems to act", please.

Indeed, "potentially" has been added to (now) L479.

- Some aspects of the title that the other reviewer and I both disliked have now been removed, and I thank the authors for addressing these concerns. Looking at the now quite short title, I feel it could be improved by specifying that this is a global modeling study, i.e. adding to the title's end something like "in a global climate model" or "in CESM1-WACCM4"

Yes that would be relevant to make such additions, now the title reads:

"Stratospheric circulation response to large Northern high-latitude volcanic eruptions in a global climate model"

- I do not understand Lines 99-100: "In the definition of a framework to study the climatic effects of a high-latitude enhancement of the stratospheric sulfate aerosol layer, Icelandic volcanism provides for an ideal test bed" could just be "Icelandic volcanism is an important case study for high-latitude eruption impacts on climate".

We see the reviewer's point and have removed this part. Now this reads (L116-117):

"Icelandic volcanism has played a role in shaping past NH climate variability and will continue doing so."

- Some wording suggestions to spruce up the abstract: First, the opening sentence is quite wordy, which makes it hard to read. I would simplify "The temporary enhancement of the stratospheric aerosol layer after major explosive volcanic eruptions" to just "Stratospheric aerosols from major explosive volcanic eruptions" or similar. Second, Line 21's instance of "with an interactive ocean, or with prescribed [...]" to "with an interactive ocean and with prescribed [...]".

We thank the reviewer for their detailed comments, we the first sentence now reads:

"Stratospheric aerosols after major explosive volcanic eruptions can trigger climate anomalies for up to several years following such events."