Reviewer #2 - Report #1

We thank the reviewer for raising these important issues. We are certain that our response to both these issues and comments as well as the ones raised in report #2 has resulted in valuable improvements regarding the reliability of our results.

In addition to the reviewer comments and general changes associated with those comments, we made three changes to the manuscript that was not part of the reports. Firstly, we added an explanation behind the LW/SW difference between cpl and atm-only. That is mentioned in subchapter 3.1. (L368-371):

"We also detect a slight difference in the LW and SW fluxes that arises from differences in high cloud cover between *atm-only* and *cpl*, where *cpl* shows a decrease in high cloud cover in the northern high-latitudes, compared to *atm-only*"

and, in the Discussions (L676-681):

"Interestingly, a slight difference between *atm-only* and *cpl* is detected in both the LW and SW flux that is caused by a strong significant decrease in high cloud cover in the *cpl* simulation (not shown). This cloud cover decrease, especially at mid- to high latitudes, agrees with the increased LW fluxes at higher latitudes in addition to the decrease in SW flux and the associated surface cooling. This raises a question on the role of forced surface processes in these high cloud changes but since it is out of the current scope we leave it open for further studies."

Secondly, we also changed the scale in the Tgrad figure (Fig. 10) to improve clarity.

Finally, we added panels showing the tropospheric zonal wind anomalies at 200 hPa (U200) for cpl and atm-only to figures 8-9 in addition to a figure showing changes in waveguides (Figure 10) (shown as a probability of favorable propagation for Rossby waves). We did this since the nature of the strong tropospheric response detected and the equatorial movement of the subtropical jet should result in changes in wave activity. This we confirm, where we show that upward wave propagation is more favorable in winters 1-2 in addition to a waveguide decrease in the subtropical troposphere that also favors large-scale waves to be redirected towards the pole.

We added the following in the Tropospheric section (L527-555):

"Since upward wave activity depends on wave mean flow interactions, several factors are at play to explain the strong response in *cpl* vs *atm-only.* First, the change in zonal flow is substantially different between the two pairs of experiments, as shown by the U200 anomalies (Fig. 8g-i and Fig. 9e-f). In the first two winters we observe an intense deepening of the Aleutian low in *cpl* (Fig. 8a-b) associated with a large equatorward shift of the subtropical jet over the North Pacific (Fig. 8g-h, also seen in the zonal mean averages of Fig. 4). The change in zonal flow is not as large in *atm-only*, where there is a general decrease of U200 on the poleward side of the subtropical jets, rather than a marked equatorward shift as in *cpl*. This further emphasizes that amplified surface coupling when the ocean is coupled to the atmosphere has a dramatic impact on the amplitude of the tropospheric response. Because the zonal flow acts as a waveguide for large-scale planetary waves, we expect changes in upward wave propagation in the stratosphere. To measure how waveguides change, Fig. 10 shows the probability of favorable propagation conditions for large-scale stationary waves number k=1,2,3, in function of pressure and latitude (see section 2 for the method). Areas of high probability show where large-scale waves preferentially propagate, while low probability regions indicate where wave propagation is prohibited. Generally, the mid-latitude troposphere is favorable for wave propagation, unlike the high-latitudes, consistent with the tendency for stationary waves to propagate upwards and to be deflected towards the equator, in climatology. After injection of the volcanic forcing, both cpl and atm-only exhibit an increase in the probability between 40 and 60°N in the lower stratosphere, but the responses in the troposphere are markedly different. In *atm-only*, wave propagation is inhibited in the free troposphere north of 60°N, for both winters 1 and 2 (Fig. 10d-e), which is consistent with the EP-flux anomalies of Fig. 5. This response is absent from *cpl* during winter 1 (Fig. 10a), and opposite during winter 2 when an increase of favorable conditions for wave propagation is observed (Fig. 10b). This increase in favorable conditions for wave propagation in the troposphere between 60 and 80°N persists during winter 3 in cpl (Fig. 10c), which is a partial explanation for enhanced upward wave propagation in the stratosphere described in Fig. 4f."

We also added to the Abstract, Methods, Discussions and Conclusions accordingly.
Since this is much more significant to our study, we have shortened the eddy-feedback part and put the figure in SI.

Then the reviewer's comments (in black) and the associated response (in red) are listed here below:

The study is noticeably improved. There are a few smaller remaining issues I describe below, but if the editor finds that the authors sufficiently address these I would support publication without needing to see the manuscript again.
We thank the reviewer for these encouraging words.

Main comments

The study is now more appropriate in less strongly emphasizing the SSW impacts than before, as there isn't enough data to rigorously settle this here (i.e. too much noise, too few simulations). I pinpoint a few remaining lines below that I find don't sufficiently convey this context to the reader. I encourage the authors to be very upfront about the issues here, which could help guide future research on these dynamical effects.

Some of the key lines are worded insufficiently clearly or confusing. I point out a number of these in the abstract and a few in the text below. There are also a couple plot issues (two Fig. 6s, tiny font size in one of those).
Thank you, we have now changed the font size and the figure numbers.

Specific comments

line 16: "beyond the duration of the radiative forcing" is not representative of this study, as most of the assessed years are when there is at least some volcanic aerosol in the stratosphere. Saying "for up to several years after the eruptions" or similarly rewording would better convey the time frame this study focuses on.
We agree and now L16 reads:
"for up to several years following such events"

Lines 17-18: Comparing long-lasting effects from tropical eruptions to dynamical effects in high-latitude eruptions gives the impression that atmospheric dynamics is causing eruption effects to outlast the volcanic aerosol presence. Instead it is the slow ocean timescale for sea surface temperatures to re-equilibrate. I recommend rewording to avoid this confusion.
This is well spotted, and now reads (L16-18):
"Whereas the mechanisms responsible for the prolonged response to volcanic surface cooling have been extensively investigated for tropical eruptions, less is known about the dynamical response to high-latitude eruptions."

line 17: "long-lasting response" here is confusing, as this is easy to conflate with the "long-lasting" duration of the eruptions themselves mentioned in the title, which is a quite different feature.
We agree, where this has been addressed in the above comment.

Line 28-9: I find confusing the line "The identification of a deterministic response such as increased SSWs following HL volcanic eruptions calls for increased attention". I think it would be better to say what the ensemble here "suggests" impacts on SSWs and clarify that more simulations than used here are needed to rigorously establish this.
We are certain that our edit to the Abstract addresses this comment, since this sentence has been removed.

Lines 31-2: I find this line problematic: "sensitivity of such events to eruption magnitude needs to be evaluated in terms of a possible source of increased seasonal predictability of

NH regional climates". The results of this study show too much noise to expect noticeable benefit, contradicting this "need". Also, the part "sensitivity [...] to eruption magnitudes" is odd because a once-per-century or rarer event would be a very minor "source of [...] seasonal predictability", while more common eruptions will have little signal. I would omit this line or make a different sentence stating that/how the results further the understanding of regional climates that follow such an eruption.
Similar to the above, this sentence has been removed.

Lines 163-168: I find the description of Lake and Eldgja as "much larger" than the 14 Tg case examined here misleading, as global climate responses are primarily caused by the stratospheric aerosols, while the much larger tropospheric aerosol amounts not simulated here will have far less impact per mass, as they fall out of the atmosphere much more quickly.
We agree, "much larger" has now been removed from the sentence where it serves as a natural comparison to our idealized experiment

Lines 362: I cannot believe this tiny p-value, as the very varied number of SSWs in the control case show these results lack high confidence. There is not enough data to seriously show results from this test. I would omit this and let the plotted results speak for themselves.
That is true, the p-value has been removed from the sentence.

Lines 366: I don't believe that atmospheric dynamical response from one winter is "an important precursor to the significant increase in SSWs" a year later. The atmosphere does not have this kind of memory. If there is a connection between the two winters, it would be because sea surface temperatures maintain a similar post-eruption pattern, which in turn could cause similar atmospheric circulation patterns.
We agree and thank the reviewer for identifying this misstatement. This now reads (L458-461):
"This upward wave flux and weaker winds continue into the third winter, where winter 2 likely acts as a precursor, allowing for SSWs to develop more frequently as detected in the T50 warming that is now confined over the polar cap (Fig. 4c and Fig. 5c, respectively)."

Lines 386-7: Please word this important sentence more accurately and preferably walk the reader through it. The results here "suggest" an effect on SSWs, while a considerably larger ensemble would be needed to properly establish this.
We agree, where we feel that our additions related to RW1 comments like the SSW analysis for atm-only (that is also related to an earlier and much relevant comment from RW2) and the analysis on the potential impact on ensemble size on the signal to noise clarifies this part and adds more confidence (Fig. S7 and Fig. 7 respectively). This is described in L411-438:
"To better understand the cpl SSWs response, we also did an SSW analysis on *atm-only* (Fig. S7) where 50-75% less SSWs were detected in the perturbed simulation compared to the

unperturbed one. Such a response should not be unexpected during the forced polar vortex strengthening as detected in *atm-only* (see Fig. 5). Furthermore, only single SSWs per winter were detected in all 20 ensemble members of the perturbed simulation while two (one) ensemble member(s) detected double SSWs per winter 1 (winter 2) in the unperturbed simulation. Although these results do show strong evidence of an increase in the number of SSWs in the *cpl* simulation, internal variability is large and the frequency of SSWs fluctuates substantially between the three winters in the unperturbed simulation.

Since this is indicative of a low signal-to-noise ratio and uncertainties in the response of polar vortex variability, we tested the potential impact of ensemble size on the signal-to-noise ratio for two key diagnostics of our mechanism, namely U10 and SSW. We express the signal-to-noise ratio as the uncertainty related to the expected (i.e., ensemble mean) response, calculated as standard error of the mean of post-eruption paired anomalies for the first three post-eruption winters. The standard errors converge toward the value obtained for the full 20-member ensemble for all winters and both variables (Figure 7), with a common tendency of growing uncertainty in the expected response with the ensemble size. The mean curves consistently level off for ensemble sizes larger than 10, suggesting that the full ensemble estimate is representative of uncertainty in the expected response for larger ensemble sizes. Otherwise, winter 3 produces larger uncertainty than winters 1 and 2, suggesting a less constrained forced response. This is especially evident for estimates of standard error of SSW anomalies and ensemble sizes larger than 15, where winters 1 and 2 closely superpose on each other while winter 3 does not overlap with winters 1 and 2 within the 5-95th percentile range. Overall, the flattening of the ensemble-mean expectations on the standard error, and the large values diagnosed in winter 3, suggest that the winter 3 response features an intrinsically lower signal-to-noise ratio. In fact, since the value of full-ensemble mean SSW paired anomaly in winter 3 (+1.05 events) is similar to the associated expected standard error (+1.02 events) we conclude that even a much larger ensemble would not provide more certainty in the signal detected. ”

Discussions and Conclusions have also been changed accordingly.

Fig. 6: There are now two Fig. 6's. Please shift the number of the second Fig. 6 and all later plots, and mentions of these in the text, to correct this.
We have corrected this mistake.

Fig. 6 (second): The font size in all the sub-plots is far too small.
This has been updated and should be more clear now.

Line 660: "A significant increase" in SSWs should just be "an increase", as there is not enough data to reliably say this is statistically significant. We would need to know what the distribution of SSW counts is across many unperturbed years, and the 3 very different counts in each control year in Fig. 6d show that the current evaluation does not have enough.
Correct, we have now changed accordingly.

Reviewer #1 - Report #2

We thank the reviewer for these important comments. We are certain that our response has resulted in valuable improvements regarding the reliability of our results.
In addition to the reviewer comments and changes associated with those comments, we made three changes to the manuscript that was not part of the report comments.
Firstly, we added an explanation behind the LW/SW difference between cpl and atm-only. That is mentioned in subchapter 3.1. (L368-271):
"We also detect a slight difference in the LW and SW fluxes that arises from differences in high cloud cover between *atm-only* and *cpl*, where *cpl* shows a decrease in high cloud cover in the northern high-latitudes, compared to *atm-only* (not shown)."
and, in the Discussions (L676-681):
"Interestingly, a slight difference between *atm-only* and *cpl* is detected in both the LW and SW flux that is caused by a strong significant decrease in high cloud cover in the *cpl* simulation (not shown). This cloud cover decrease, especially at mid- to high latitudes, agrees with the increased LW fluxes at higher latitudes in addition to the decrease in SW flux and the associated surface cooling. This raises a question on the role of forced surface processes in these high cloud changes but since it is out of the current scope we leave it open for further studies."

Secondly, we also changed the scale in the Tgrad figure (Fig. 10) to improve clarity.

Finally, we added panels showing the tropospheric zonal wind anomalies at 200 hPa (U200) for cpl and atm-only to figures 8-9 in addition to a figure showing changes in waveguides (Figure 10) (shown as a probability of favorable propagation for Rossby waves). We did this since the nature of the strong tropospheric response detected and the equatorial movement of the subtropical jet should result in changes in wave activity. This we confirm, where we show that upward wave propagation is more favorable in winters 1-2 in addition to a waveguide decrease in the subtropical troposphere that also favors large-scale waves to be redirected towards the pole.
We added the following in the Tropospheric section (L527-555):
"Since upward wave activity depends on wave mean flow interactions, several factors are at play to explain the strong response in *cpl* vs *atm-only*. First, the change in zonal flow is substantially different between the two pairs of experiments, as shown by the U200 anomalies (Fig. 8g-i and Fig. 9e-f). In the first two winters we observe an intense deepening of the Aleutian low in *cpl* (Fig. 8a-b) associated with a large equatorward shift of the subtropical jet over the North Pacific (Fig. 8g-h, also seen in the zonal mean averages of Fig. 4). The change in zonal flow is not as large in *atm-only*, where there is a general decrease of U200 on the poleward side of the subtropical jets, rather than a marked equatorward shift as in *cpl*. This further emphasizes that amplified surface coupling when the ocean is coupled

to the atmosphere has a dramatic impact on the amplitude of the tropospheric response. Because the zonal flow acts as a waveguide for large-scale planetary waves, we expect changes in upward wave propagation in the stratosphere. To measure how waveguides change, Fig. 10 shows the probability of favorable propagation conditions for large-scale stationary waves, averaged for zonal wave numbers k=1,2,3 and meridional wave numbers l=1,2,3, in function of pressure and latitude (see section 2 for more detail). Areas of high probability show where large-scale waves preferentially propagate, while low probability regions indicate where wave propagation is prohibited. Generally, the mid-latitude troposphere is more favorable for wave propagation than the high-latitudes and the stratosphere, consistent with the tendency for stationary waves to propagate upwards and to be deflected towards the equator, in climatology. After injection of the volcanic forcing, both *cpl* and *atm-only* exhibit an increase in the probability for wave propagation between 40 and 60°N in the lower stratosphere during winter 1 and 2, but the responses in the troposphere are markedly different. In *atm-only*, wave propagation is inhibited in the free troposphere north of 60°N, for both winters 1 and 2 (Fig. 10d-e), which is consistent with the EP-flux anomalies of Fig. 5. This response is absent from *cpl* during winter 1 (Fig. 10a), and opposite during winter 2 when an increase of favorable conditions for wave propagation is observed (Fig. 10b). We also see that the waveguide has greatly reduced in the subtropical troposphere in *cpl* winters 1-2 that favor large-scale waves to be redirected towards the pole. This increase in favorable conditions for wave propagation in the troposphere between 60 and 80°N persists during winter 3 in *cpl* (Fig. 10c), which is a partial explanation for enhanced upward wave propagation in the stratosphere described in Fig. 4f."

We also added to the Abstract, Methods, Discussions and Conclusions accordingly.
Since this is much more significant to our study, we have shortened the eddy-feedback part and put the figure in SI.

Then the reviewer's comments (in black) and the associated response (in red) are listed here below:

Some major questions and concerns raised in previous comments still exist in the revised manuscript. For example, the emphasize on the "long-lasting" emission and limited ensemble members to investigate polar vortex and SSWs change, and 2 to 3 SSWs in one single winter simulated by the model.

We have now removed “long-lasting” from the title and replaced it with “large” that should decrease the emphasis on our experiment being “long-lasting”. We also feel that the paragraph here below (L168-176) clarifies this:
“We thus obtain aerosol optical properties for an idealized, long-lasting high-latitude NH eruption. In this experiment we assume stratospheric injection only, although similar eruptions in the natural world would likely inject part of the total aerosol mass within the troposphere during the eruption. Past NH eruptions like Eldgjá and Laki had an atmospheric SO_2 loading of 219Tg and 122Tg respectively that was carried aloft with the eruptive column up into the upper troposphere with portions of the aerosols reaching the lower stratosphere during the eruptions (Thordarson et al., 2001). Hence our experiment can also be considered as a 6-month stratospheric aerosol injection that is analogous to similar although smaller eruptions (as compared to Laki) without the tropospheric aerosols.”
Also, in the main text where we discuss our results, we now refer to an enhancement of stratospheric aerosols that should also contribute to this clarification.

The following paragraph on our recently added SSW analysis for atm-only has been added to the Results section 3.2.3 in addition to the addition of the associated Figure in the Supplementaries (L411-419):
“To better understand the cpl SSWs response, we also did an SSW analysis on *atm-only* (Fig. S7) where 50-75% less SSWs were detected in the perturbed simulation compared to the unperturbed one. Such a response should not be unexpected during the forced polar vortex strengthening as detected in *atm-only* (see Fig. 5). Furthermore, only single SSWs per winter were detected in all 20 ensemble members of the perturbed simulation while two (one) ensemble member(s) detected double SSWs per winter 1 (winter 2) in the unperturbed simulation. Although these results do show strong evidence of an increase in the number of SSWs in the *cpl* simulation, internal variability is large and the frequency of SSWs fluctuates substantially between the three winters in the unperturbed simulation. ”

We also added the following to the Discussions (L642-654):
“We also see that the large decrease in SSWs in the perturbed simulation of *atm-only* (when compared to unperturbed) is consistent with the detected polar vortex strengthening. This further supports the significance of the signal we detect in *cpl* winter 3 compared to the background noise. In addition, all winters examined, in both *cpl* and *atm-only*, showed that there is up to 15% chance of getting more than 1 SSWs per winter in all ensemble members. This is not far from Ineson et al. (2022) who identified a double event once every 9 years in a

66-year ERA5 record. The exception is *cpl* winter 3 that is also the only winter that has 3 SSWs, with the average SSW occurrence also being the only winter above 1 (1.17) while all other winters span between 0.15-0.85 per winter. A similar NH high-latitude eruption has not taken place during the observational period, so we have no comparison. Also, to the best of our knowledge, a similar high-latitude sulfur injection study has not been performed before. Therefore, it is difficult to say at this stage if such a response is realistic or not, but in general more than two SSWs per winter can be considered exceptional yet plausible, as is also the case for our idealized eruption."

Additionally, we added analysis to test the impact of the ensemble size on the signal-to-noise ratio. This is depicted in Figure 7 and the following text was added (L420-438):
"Since this is indicative of a low signal-to-noise ratio and uncertainties in the response of polar vortex variability, we tested the potential impact of ensemble size on the signal-to-noise ratio for two key diagnostics of our mechanism, namely U10 and SSW. We express the signal-to-noise ratio as the uncertainty related to the expected (i.e., ensemble mean) response, calculated as standard error of the mean of post-eruption paired anomalies for the first three post-eruption winters. The standard errors converge toward the value obtained for the full 20-member ensemble for all winters and both variables (Figure 7), with a common tendency of growing uncertainty in the expected response with the ensemble size. The mean curves consistently level off for ensemble sizes larger than 10, suggesting that the full ensemble estimate is representative of uncertainty in the expected response for larger ensemble sizes. Otherwise, winter 3 produces larger uncertainty than winters 1 and 2, suggesting a less constrained forced response. This is especially evident for estimates of standard error of SSW anomalies and ensemble sizes larger than 15, where winters 1 and 2 closely superpose on each other while winter 3 does not overlap with winters 1 and 2 within the 5-95th percentile range. Overall, the flattening of the ensemble-mean expectations on the standard error, and the large values diagnosed in winter 3, suggest that the winter 3 response features an intrinsically lower signal-to-noise ratio. In fact, since the value of full-ensemble mean SSW paired anomaly in winter 3 (+1.05 events) is similar to the

associated expected standard error (+1.02 events) we conclude that even a much larger ensemble would not provide more certainty in the signal detected."

According to the above, we also added to Discussions (L640-642):
"Furthermore, the SSW analysis for *atm-only* and the ensemble size test (Fig. 7) both show strong evidence of a robust signal for winter 3 despite the noisy polar vortex and the limited ensemble size."

Besides, some long sentences can be separated into shorter sentences to increase the clarity.
We thank the reviewer for this important comment. We have now revised the manuscript with this in mind where we have shortened most of the sentences the reviewer refers to.

Except for the unsolved major concerns, below are some minor comments:
Title: "long-lasting". As commented by reviewers, this can be misleading and considering that there were eruptions that is much longer than 6 months as shown in this paper: Gabriel, I., Plunkett, G., Abbott, P.M. et al. Decadal-to-centennial increases of volcanic aerosols from Iceland challenge the concept of a Medieval Quiet Period. Commun Earth Environ 5, 194 (2024). https://doi.org/10.1038/s43247-024-01350-6
We thank the reviewer for this comparison. However we do feel that this paper is not directly comparable to ours since it is about the Icelandic Active Volcanic period where several volcanic systems contributed to the aerosol loading - most of them being tropospheric like the Hrafnkatla eruption that had one stratospheric event. With that being said, we agree that only using "long-lasting" can be confusing, where our experiment simulates long-lasting stratospheric eruption/aerosol injection. We have thus made changes to the text accordingly in addition to removing "long-lasting" from the title.

Line 159-160: October 1 should be November 1 if starts from May 1 and lasts for 6 months.
Correct, this has been fixed.

Line 163-164: atmospheric loading? Better specify SO2 or SO4, that's very different. Are the loading numbers correct based on the reference? In a new paper Hutchison et al., 2024, it writes "The Eldgjá eruption… estimated to have released 220 Tg of SO2 into the atmosphere (with ∼185 Tg of this reaching upper tropospheric and lower stratospheric altitudes, Thordarson et al., 2001)." 120Tg is a huge difference as written in this manuscript. Hutchison, W., Gabriel, I., Plunkett, G., Burke, A., Sugden, P., Innes, H., et al. (2024). High-resolution ice-core analyses identify the Eldgjá eruption and a cluster of Icelandic and trans-continental tephras between 936 and 943 CE. Journal of Geophysical Research: Atmospheres, 129, e2023JD040142. https://doi.org/10.1029/ 2023JD040142
We thank the reviewer for correcting this misstatement, the eruption order had been mixed. It now writes (L163-164):
"Past NH eruptions like Eldgjá and Laki had an atmospheric SO_2 loading of 219Tg and 122Tg respectively,"

Figure 1: still not clear. Previous reviewers' comments gave suggestions which should be helpful but was not adopted.
We agree and now this figure has been updated where each forcing has its own axis. We also took the opportunity to update the figure since the forcing curves ended on 1. Dec but not 31. Dec as they should have. Hopefully this adds more clarity.

L195-197 and figure 1-3: Figure 1 looks correct with three post-eruption winters. But when looking at figure 2 and figure 3, there is only two winters after the eruption? The explanation to subsequent figures should be connected to figure 2 and 3, then they need to be double checked and clarified.
This is true, Figure 1 shows the 36 month long volcanic forcing profile used in our simulations. However, as can be seen in Fig. 1 the aerosols have decreased substantially in the 4th year and since the memory in the atmospheric component is very small, the atmosphere-only simulation only ran for 3 years - thus explaining why the third winter is absent in atm-only. This is mention here in lines 190-192:
"The atmosphere-only experiments were run over three full years, which provides two full winters after the onset of the eruption. We found that there was no need to extend the simulations further given the duration of the forcing and short memory of the atmosphere."
In the coupled simulation we use the same forcing and allow it to run for a longer time since there the memory comes into play.

L223-224 what temperature gradient is this? Needs to be clarified. Different from stratospheric temperature "gradient" in L40?
This would be Kelvin, this has been added in the text.

L234: decrease in the LW? It's confusing. Better change negative value to positive value of different direction to SW and adjust the description in the manuscript.
We agree, this is confusing. We have changed the LW colorbar so it now shows positive values and the text has changed accordingly.

L244-245: radiative forcing confined to NH extratropical summers?
We thank the reviewer for spotting this error, this now reads (L262-263):
"Overall, the radiative forcing is largely bounded by the NH extratropics..."

Figure 6: It would be nice to show the evolution of the SSWs as suggested by reviewer 2, as it's still not clear how you counted the numbers and doubtful about 2 to 3 SSWs in one winter.
We feel this has been addressed in the first comment in addition to showing significant U10 winds in Fig. 3c, where the zonal wind decrease becomes clear for *cpl*. Where in addition to showing SSWs for atm-only, we also show that although 2 SSWs per winter are not common - they do occur in reanalysis datasets spanning 66 years (in total 9 events) in addition to observations (see references in main text). It is the 3 SSWs per winter that is unprecedented - and since it only occurs in winter 3 it is more likely that the forced response is responsible - especially when comparing the SSW statistics between cpl and

atm-only. If we remove the two cpl ensemble members that show 3 SSWs to treat as outliers, the SSW increase is still significant (not shown).

Figure 7: line labels not clear.
True indeed, this should be better now.

L1089-1096 reference format needs to be double checked, these two references, one has the publish year after the authors while the other one is at the end. Also exist in several other references.
This has now been checked and edited in accordance with the journal reference guidelines.

L1098-1100 “eaat6025. doi:10.1126/sciadv.aat6025. PMID: 30050990; PMCID: PMC6059732.” ?
Thank you for noticing, this reference has been updated.

L1105-1106 The citation you used in the discussion part, better check the published paper:

Zhuo, Z., Fuglestvedt, H. F., Toohey, M., and Krüger, K.: Initial atmospheric conditions control transport of volcanic volatiles, forcing and impacts, Atmos. Chem. Phys., 24, 6233–6249, https://doi.org/10.5194/acp-24-6233-2024, 2024.
Thanks, done.