Response to the comments from Referee-2

We sincerely appreciate your thoughtful questions and valuable suggestions, especially the suggestions regarding carefully concluding our results and reducing the comparison with cluster 5 of Kretschmer et al. (2018a). We have carefully addressed each of your comments point by point (highlighted in blue) and made the corresponding revisions in the revised version of the manuscript.

General comments:

I thank the authors for their detailed responses to the comments made by me and the other reviewer. The revised version shows significant improvement and addresses most of my previous concerns. However, a few statements remain imprecise or overly strong, and adjusting them would help ensure that the conclusions accurately reflect what can be supported by the analyses. Please see the detailed comments below.

Major Comments:

1. Some of the statements regarding the causality are too strong. The manuscript frequently uses expressions such as 'causally'. I understand that the main aim is to isolate the influence from the stratospheric forcing, and the authors wish to emphasize this aspect. However, it might be misleading in several contexts. For instance, the last sentence in the abstract, while it is true that the tropospheric response in these experiments originates from the stratospheric forcing, the sentence 'the observed surface response … causally forced by, stratospheric perturbations' (L18-20) is not precise. It is also unclear whether 'observed' refers to the detected model response or observational data. If the latter, the statement is too strong. I suggest rephrasing this as 'can be forced by …' or something similar. In addition, the divergent mass streamfunction can help explain the surface temperature response, but only to a certain extent. The phrase 'causally linked' (L17) might overstate the role of this diagnostic. I recommend softening the tone so that the conclusions more precisely reflect what can be inferred from the present analysis.

We agree that the mass streamfunction by the divergent meridional wind can explain only a part of the surface temperature responses. We also acknowledge that we can't demonstrate causality of the signal from observational data using these experiments. The word "causally" is removed from L17 in the revised manuscript.

However, it is not ambiguous that the surface response following SSWs can be causally forced by stratospheric perturbations, based on nudging an SSW in a model (Hitchcock and Simpson, 2014) or MiMA simulations by imposing zonally symmetric forcing in the stratosphere (White et al., 2020, 2022). In observations, the tropospheric precursors could also impact the surface conditions, and it's impossible to fully isolate the impact from the stratosphere and troposphere in the observations because they are coupled with each other. While in this study, we can fully isolate the tropospheric precursors and only address the role of the stratosphere.

2. The statement regarding the consistence with observation/reanalysis is not accurate. The manuscript currently states that the two phases correspond to clusters 4 and 5 of Kretschmer et al. (2018a). However, while the surface temperature response indeed show similarities, the stratospheric circulation does not align in the same way. In particular, wave reflection only appears in cluster 4, where cluster 5 shows a reduced upward wave propagation, but the raw Fz is still positive over Eurasia, unlike the negative raw Fz in the phase-270 experiments here. I understand that we cannot expect the idealized experiments to reproduce every observational feature. However, the comparison should be presented more carefully to avoid implying equivalence where the mechanisms differ. I suggest refining the relevant statements accordingly. In addition, repeatedly referring to cluster numbers may confuse readers unfamiliar with the cited work; it may be clearer to describe their defining characteristics when first introduced and avoid relying solely on “cluster 4/5” labels thereafter.

We agree that cluster 5 of Kretschmer et al. (2018) is not a pure composite of wave reflection events, and the stratospheric circulation in their clusters 4 and 5 is not fully consistent with phase-90 and phase-270 ensembles in our study. However, the zonal dipole characterized by positive Fz anomalies over North America and negative Fz anomalies over North Eurasia does indicate that some wave reflection events are included in cluster 5. The relevant comparison has been more carefully presented in the revised manuscript in L321-329, together with a simple description of cluster 4/5. Also shown here:

“A similar spatial pattern is also observed in the composite of cluster 4 events in Kretschmer et al. (2018a), which are characterized by raw positive Fz over Eurasia and raw negative Fz over North America. They demonstrated that such events are associated with North American cold spells via reflected upward-propagating waves over eastern Siberia by using causal effect network analysis. In the phase-270 ensemble, downward wave propagation is instead concentrated over North Eurasia, and is immediately followed by pronounced surface cooling over North Eurasia (Figure 4d). Moreover, observations of cluster 5 in Kretschmer et al. (2018a), characterized by anomalously positive Fz over North America and anomalously negative Fz over Eurasia, show a similar surface temperature response to that in the phase-270 ensemble. However, it is important to understand that cluster 5 in Kretschmer et al. (2018a) does not represent a pure composite of reflection events, as it still displays raw positive Fz over Eurasia despite the anomalously negative Fz. ”

3. The alignment of timing between branch ensembles and control runs. While the authors noted that ‘there is no expectation for the timing of the surface responses to match’, the magnitude of response in the branch ensembles after day 13 appears more comparable to those in the control runs after day1. I understand that ‘day0’ represents different reference points, and strict alignment is not required. But align the timelines based on the peak zonal-wind reversal (e.g., day 0 in the control run and day 12 in the branch runs) may make the comparison more straightforward for readers. Alternatively, omitting direct cross-experiment comparisons at fixed lags, or explicitly noting their limitations, would avoid potential confusion.

Good suggestion, thanks! We have redefined day-0 in the CONTROL runs so that the day with peak easterlies is the same in both CONTROL and branch runs. It turns out that the peak easterlies in the branch runs are on day 12 after branching (the last day with imposed forcing), and so we had to redefine day-0 in the control runs accordingly. We have revised all figures in the paper to use this new definition.

This has been clarified in the methods section in L171-174. Also shown here:

“Day 12 in control runs indicates the day with peak easterlies following the onset of SSWs, whereas day 12 in branch ensembles corresponds to January 12 of every year (the last day with imposed forcing), accompanied by peak easterlies following the onset of forced SSWs too.”

Specific Comments:

1. L17. Should ‘downward propagation events’ refer instead to ‘wave reflection events’?

The wave reflection could be affected by the imposed momentum torque during the forcing stage (days 1-12), while wave reflection freely responds to the zonal asymmetry of the polar vortex when the forcing is switched off (day 13 and onward). To better summarize the two periods, “downward propagation events” is used.

2. L61-63 and L408-410. Previous studies have shown that this type of stratospheric anomaly is linked to preceding tropospheric circulation (e.g., Shen et al. 2023; Tan and Bao 2020) and that similar stratospheric disturbances can lead to distinct surface response depending on the tropospheric processes involved (e.g., Shen et al. 2025). Adding a brief discussion where relevant can be helpful to strengthen the motivation for isolating the role of stratosphere.

Cited in suitable places.

3. L152. Change ‘present’ to ‘represent’.

Corrected.

4. L203. Should be ‘McIntyre’ and ‘Edmon et al.’.

They should be “Andrews and Mcintyre, 1976” and “Edmon Jr et al., 1980” as in the manuscript. See the following links in Google Scholar.

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Planetary+waves+in+horizontal+and+vertical+shear%3A+The+generalized+Eliassen-Palm+relation+and+the+mean+zonal+acceleration&btnG=

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Eliassen-Palm+cross+sections+for+the+troposphere&btnG=

5. L231. It is more accurate to state ‘averaged over days 6 to 12’. The same applies to other similar descriptions.

Corrected.

6. Figure 3c and d. The tropospheric polar-cap height anomaly peaks almost simultaneously with the stratospheric anomaly. Could the authors clarify why this occurs?

In shorter timescales (e.g., 6-hourly), the tropospheric response lags behind the stratospheric anomalies.

7. L273-275. For the reasons described in major comment #2, I suggest reducing the emphasis on direct comparison with Kretschmer et al. (2018a), particularly for phase-270, which does not closely resemble cluster 5 beyond the surface temperature pattern.

This comparison has been removed from the revised manuscript.

8. L283-284. Here the comparison uses days 1-5, but earlier the authors note that the timings are not expected to match. As mentioned in major comment #3, aligning the timing or avoiding such direct comparisons may reduce confusion.

Your suggestion in major comment 3 was a good one - thanks! All figures have been modified to align the timing.

9. L310-312. Please specify the longitude range of the region discussed for easier interpretation.

The longitude range has been clarified in L308-312. Also shown here:

"The shading in Figure 6 shows the raw geopotential height eddy area-averaged from 55°N to 80°N, and there is a clear eastward tilt from 180°E to 300°E in the phase-90 ensemble (Figure 6c) and especially in the phase-270 ensemble (Figure 6d) with downward directed arrows from 90°E to 270°E, which also indicates the presence of wave reflection events over there (Cohen et al., 2022)."

10. L321-326. Phase-90 shares characteristics with cluster 4, but phase-270 does not resemble cluster5. Revising this statement for accuracy would be beneficial.

Good suggestion. This part has been modified in L321-329. Also shown here:

"A similar spatial pattern is also observed in the composite of cluster 4 events in Kretschmer et al. (2018a), which are characterized by raw positive Fz over Eurasia and raw negative Fz over North America. They demonstrated that such events are associated with North American cold spells via reflected upwardpropagating waves over eastern Siberia by using causal effect network analysis. In the phase-270 ensemble, downward wave propagation is instead concentrated over North Eurasia, and is immediately followed by pronounced surface cooling over North Eurasia (Figure 4d). Moreover, observations of cluster 5 in Kretschmer et al. (2018a), characterized by anomalously positive Fz over North America and anomalously negative Fz over Eurasia, show a similar surface temperature response to that in the phase-270 ensemble. However, it is important to understand that cluster 5 in Kretschmer et al. (2018a) does not represent a pure composite of reflection events, as it still displays raw positive Fz over Eurasia despite the anomalously negative Fz. ”

11. L445. Cluster 5 in Kretschmer 2018a does not show a wave reflection. This should be corrected.

Corrected in L448-449. Also shown here:

“However, it is concentrated over North Eurasia in the phase-270 ensemble (Figure 7d), consistent with the anomalous Fz pattern in cluster 5 observed by Kretschmer et al. (2018a).”

References

Hitchcock, P., and I. R. Simpson, 2014: The Downward Influence of Stratospheric Sudden Warmings. J. Atmos. Sci., 71, 3856–3876, https://doi.org/10.1175/JAS-D-14-0012.1

White, I. P., C. I. Garfinkel, E. P. Gerber, M. Jucker, P. Hitchcock, and J. Rao, 2020: The Generic Nature of the Tropospheric Response to Sudden Stratospheric Warmings. J. Climate, 33, 5589–5610, https://doi.org/10.1175/JCLI-D-19-0697.1

White, I. P., C. I. Garfinkel, and P. Hitchcock, 2022: On the Tropospheric Response to Transient Stratospheric Momentum Torques. J. Atmos. Sci., 79, 2041-2058, https://doi.org/10.1175/JAS-D-21-0237.1