Response to the comments from Referee-2

We sincerely appreciate your thoughtful questions and valuable suggestions, especially the suggestions regarding the structure of the introduction and highlighting the dynamic linkage between SSWs and wave reflections. 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.

Note that there are five main modifications in the revised manuscript: 1) a new structure of the introduction section; 2) a new Table 1 summarizing the analyzed experiments in this study; 3) a new Figure 2 showing the period in days from 6 to 12; 4) a new Figure 6 showing the longitude-pressure cross section of Plumb flux (Fx, Fz); 5) a new Figure 10 showing the zonal wind anomalies at 321 hPa.

General comments:

This study uses an intermediate-complexity GCM (MiMA) with imposed stratospheric momentum torques to investigate how zonally asymmetric forcing affects the tropospheric and surface response following SSWs and wave reflection events. By comparing symmetric forcing with different wave-1 forcings, the authors isolate the causal impact of stratospheric zonal structure. The manuscript is generally well written and the analyses are systematic. However, the current manuscript lacks integration across sections, which weakens the logical flow. In addition, a clearer illustration of the experimental design is needed. Therefore, I recommend a major revision.

Major Comments:

1. Improve integration and narrative flow. While individual sections provide informative analyses, they are currently disconnected. This is reflected in several aspects:

• First, the introduction section heavily focuses on the SSWs, while wave reflection receives less attention, despite both being central in the title and abstract. In my opinion, a more balanced structure is needed, perhaps starting with general features and influence of stratospheric polar vortex variability would better frame the study.

• Second, the analyses of SSW-like events and wave reflections appear as separate topics, although they arise from the same experiments. This suggests that these SSWs and wave reflections are dynamically linked. One implication could be that SSWs are associated with different types of wave reflections depending on the phase of the imposed stratospheric wave forcing, which itself is an important conclusion worth highlighting.

• Third, the surface response results from both the wave reflection and the downward coupling discussed in Section 4.4. However, in the current format, these aspects are presented rather separately.

I highly suggest adjusting the structure to make the story more coherent. For instance, emphasize that multiple mechanisms can together explain the surface response (wave reflection, mass streamfunction). Alternatively, the authors could first introduce and discuss the features and dynamical processes, and then link them to the surface response.

Great suggestions! We focus more attention on wave reflection events in the introduction and highlight the additive impacts on surface responses from both SSWs and concurrent wave reflection events, as explained by mass streamfunction and their dynamical linkage, driven by the imposed forcing in the revised manuscript.

1. Two paragraphs introduce the wave reflection event starting from L53 in the revised manuscript.
2. Highlighting the dynamical linkage between SSWs and wave reflection due to the phasing of imposed forcing in L1-L2, L330-L331, L445-L449, and L467-L472.
3. Highlighting the joint impacts on surface response from the SSW and the wave reflection event in L19-L20, L393-L395, and L467-L472.

2. Clarify and illustrate the experimental design. The description of the model experiments is a bit hard to follow. There are several groups of experiments, the control runs (9 runs times 50 years), the CTRL run (the median gravity wave drag one, 50 years), and the branch experiments (50 years of the CTRL run times different forcings). This information is described across multiple paragraphs and I had to re-read Section 2 several times to fully understand it. To help readers quickly grasp the essentials of the model settings, I suggest adding a schematic diagram or a concise summary table to visually summarize the experimental setup.

Good suggestion! Please see Table 1 here and also in the revised manuscript.

Table 1. List of experiments analyzed in this study. Note that the symmetry, phase-90, and phase-270 ensembles are branched off from the CTRL.

Name	Ensemble members	Total period	Forcing
control	9 (8+CTRL)	9 × 50 years	free running (differing in gravity wave drag in yr-1; CTRL with median drag)
symmetry	50	50 × 4 months	$M_S = -15$ m s^{-1} day^{-1}
phase-90	50	50 × 4 months	$k = 1$, $\Theta_A = 4$, $\lambda_0 = 90°$, $M_S = -15$ m s^{-1} day^{-1}
phase-270	50	50 × 4 months	$k = 1$, $\Theta_A = 4$, $\lambda_0 = 270°$, $M_S = -15$ m s^{-1} day^{-1}

Specific Comments:

1. Introduction: The Introduction currently spends many paragraphs on SSWs while only briefly mentioning wave reflections. Since both are emphasized in title and abstract, I suggest the

authors to restructure it to (i) introduce the general polar vortex variability and its tropospheric impact, (ii) introduce SSWs and wave reflection in turn and their linkage, and (iii) identify the gaps that this study aims to address. This will provide a more balanced background and better motivate the focus of this work.

Thank you for the valuable suggestions! We put more attention on the wave reflection events and highlight the additive impacts from both SSWs and concurrent wave reflection events in the introduction and conclusion. Please see them in the revised manuscript.

2. L50: 'cleanly' should be 'clearly'.

Corrected.

3. L142: by 'In all, 48 SSW events are identified across all nine control runs during the JFM period', do you mean 48 SSWs in total over the 9 runs*50 years (i.e., 450 winters)? Please clarify. If so, this implies a very low frequency of SSWs even only JFM is considered, which may warrant brief discussion.

Yes, we obtained 48 SSWs across 450 winters. White et al. (2022) reported a ratio of 0.29 SSWs per year in their control runs by using MiMA. This ratio is still smaller than the observed ~0.67 SSWs per year, while slightly different gravity wave settings lead to an SSW frequency above 0.4 per year.

This discussion has been added to L132 in the revised manuscript.

4. Figure 1: the color bar in the upper panel spans from -120 to 120m/s. Are the zonal wind anomalies really this large in your experiments? Please clarify whether this reflects the imposed forcing amplitude and discuss whether the results are sensitive to the forcing amplitude.

Yes, the zonal wind anomalies reach these large values in our experiments due to the strong symmetric component of forcing, -15 m/s/day. The raw U1060 values are of similar magnitude to some of the stronger observed events, but well within the range of observations (see Figure 3 in our manuscript). White et al (2022) examined the linearity of the surface impact from the strength of the stratospheric symmetric component of forcing and found strong linearity. Please see Figure 4 in White et al. (2022), who examined the responses of zonal-mean zonal winds at 10 hPa and 60°N and polar cap temperatures to different symmetric forcings, which shows that a larger forcing leads to weaker zonal winds and warmer polar cap. For this paper, we choose a forcing that is on the strong end of observations to get a strong surface response.

We have also performed experiments with a wide range of zonally asymmetric scaling factors, though for all but the final configuration the model output wasn't retained, and plots weren't saved for most configurations either. But to give the reviewer a sense that results are insensitive, below we include a figure (Figure R1) showing the Z 500hpa anomalies in week 1 for a scaling factor of 0.8 (the paper uses 4; Figure R2) for the zonally asymmetric component. The anomalies are very linear. A larger scaling factor amplifies regional anomalies, while the spatial pattern of responses is essentially identical in Figures R1 and R2. Note that the anomalies in Figures R1 and R2 are the differences between phase-90/270 and symmetry ensembles.

This discussion has been added to L181 in the revised manuscript.

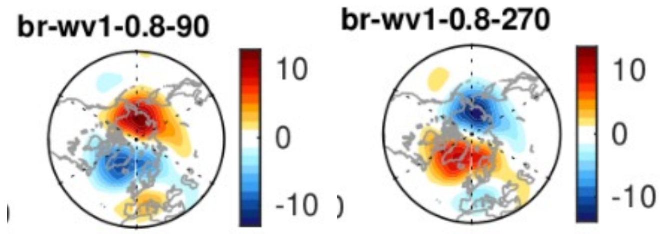

Figure R1. Geopotential height anomalies (m) at 500 hPa in week 1 with scaling factor 0.8 for phase-90 and phase-270 ensembles. Note that the anomalies here are differences between phase-90/270 and symmetry ensembles.

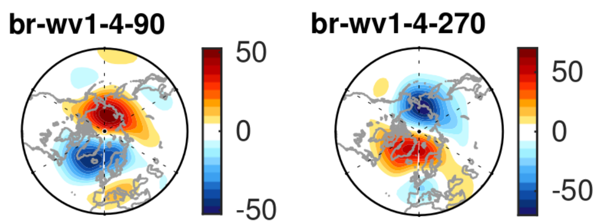

Figure R2. Geopotential height anomalies (m) at 500 hPa in week 1 with scaling factor 4 for phase-90 and phase-270 ensembles. Note that the anomalies here are differences between phase-90/270 and symmetry ensembles.

5. Figure 1 and Figure 2 focus on different period, why is that? Also, should the unit in Figure 2 should be gpm?

There are no big differences between days 1-5 and days 6-12, except for the stronger anomalies in days 6-12. We replace Figure 2 with the geopotential height patterns during days 6 -12 in the revised manuscript to keep it consistent with Figure 1.

Yes, it is better to use the gpm as the unit of geopotential height, although 1 gpm is approximately 1 m. However, we can not replot Figure 3 to modify the unit of geopotential height now because the data disk is offline from our cluster. To keep in consistency with other figures, we continue using "m" as the unit of geopotential height in the revised manuscript. We will modify this issue once the data disk is back.

6. Section 4.2. Figure 4 suggests that the peak surface response amplitude is similar across experiments, but the control run response occurs earlier. Please discuss the possible reason. I also suggest the authors including a time-height evolution of the NAM index or zonal wind

anomaly to illustrate the zonal-mean downward propagation, which might be helpful to understand the different surface response.

In control runs, day 0 indicates the reversal of zonal-mean zonal wind at 10 hPa and 60°N (Figure 3a), transitioning from westerly to easterly. Therefore, days 1-5 are exactly the end of the forcing stage of SSWs in control runs, while days 1-5 are the initial forcing stage in branch ensembles. Consequently, the control run response occurs earlier in days 1-5. We also clarify this in L252 and in L281 in the revised manuscript.

The time-height evolution of the NAM index or zonal wind anomaly can provide only the information when the stratospheric anomalies propagate downward into the lower troposphere without showing the information of regional anomalies, while Figures 4 and 5 show that significant surface anomalies emerge in days 13-19. To better understand the surface temperature response, we suggest linking t2m anomalies (Figure 4) with the mass streamfunction anomalies in the lower troposphere (Figure 11).

We add the geopotential height anomalies at 10 hPa in the new Figure 2. Tropospheric anomalies lie directly beneath the strongest anomalies in the mid-stratosphere but show little westward shift. We believe this could help better understand the downward coupling.

7. Figures 4 and 5. I suggest the authors including the tropospheric circulation to briefly compare if the circulation differs among different experiments. Even a brief illustration would help clarify the connection between stratospheric forcing and surface patterns.

We add the zonal wind anomalies at 321 hPa in the new Figure 10. Please check the relevant discussion starting from L357 in the revised manuscript.

8. L303-304: In the symmetry ensemble, the Fz is not centrally located above the pole, but shows downward propagation over NA in Day 1-12. This might be related to the climatological wave propagation.

Yes, Messori et al. (2022) reported a climatological downward wave propagation over Canada in DJFM periods from 1979 to 2021 based on ERA5 reanalysis (see their Figure 1). This discussion has been included in L317 in the revised manuscript.

References

Messori, G., Kretschmer, M., Lee, S. H., and Wendt, V. (2022): Stratospheric downward wave reflection events modulate North American weather regimes and cold spells, Weather Clim. Dynam., 3, 1215–1236, 2022. https://doi.org/10.5194/wcd-3-1215-2022

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