We thank the referee for the insightful comments. Below, we repeat the reviewer's remarks in red italics, and add our respective responses in normal text.

> *Reviewer 3*
>
> *The study addresses the impact of long-term intensified GW forcing on SPV. The authors classified SPV into ten groups based on morphology and analyzed the impact of GW forcing on three major orographic GW hotspots (HI, EA, and NA) separately. Using the UA-ICON global circulation model, the authors conducted multi-ensemble simulations to account for the internal variability of the signal. The paper showed that all hotspots exhibited a consistent decrease in the amplitude of planetary wave number 1 and the corresponding change in circulation. Furthermore, the paper showed that these anomalies are due to changes in the mean SPV structure and frequencies of several clusters, to varying extents depending on the hotspot.*
>
> *While Mehrdad et al. address highly relevant scientific questions within the scope of ACP, the paper requires minor revisions to be published. The paper introduces a novel methodology for SPV classification and carefully addresses the impact of GW forcing on different clusters. However, the discussions of statistical significance and the mechanism of the signals are lacking and could be improved.*

We thank the reviewer for their thoughtful evaluation and constructive feedback. We have carefully addressed the reviewer's concerns, which has improved the overall quality and clarity of the manuscript.

> *General comments*
>
> 1. *Are there any differences in the vertical profiles of GW forcing by the hotspot? Perhaps there are no significant differences, but showing the general picture of how the forcings are applied would be helpful. If differences exist, they could suggest potential reasons for different responses. The example figure in Mehrdad et al. (2025a) helps with understanding, but a more general distribution of the forcing by hotspot would be helpful.*

We thank the reviewer for this helpful suggestion. In the revised manuscript, we now provide a concise summary of the zonal-mean vertical distribution of the SSO-induced zonal wind tendencies by hotspot in a new Appendix A (lines 486–506 in the revised manuscript). We also added a one-sentence pointer in Section 2.1 directing readers to Appendix A for this summary of the imposed SSO forcing (lines 98-100).

Figure A1 in the revised manuscript presents the column-averaged (200–1 hPa, stratosphere) maps of the SSO-induced zonal drag during NDJFM, highlighting where the forcing is applied. Figure A2 shows the corresponding vertical structure of the SSO-induced zonal-mean zonal tendencies for the control and for the anomalies in the sensitivity experiments (HI–C, NA–C, EA–C). These diagnostics are adapted from Mehrdad et al. (2025a).

The three hotspots exhibit distinct vertical structures. Over HI and NA, the forcing primarily strengthens the lower-stratospheric (valve-layer) peak, with HI showing a clearer upward extension and NA remaining more confined in height and accompanied by upper-level westerly anomalies. Over EA, the enhancement is broader and more vertically extended, with less emphasis on a distinct lower-stratospheric maximum. These regional differences reflect variations in GW generation by the SSO scheme and in the background flow.

Because the present paper focuses on the SPV response classified by vortex shape, we keep the treatment of vertical structure brief and refer readers to Mehrdad et al. (2025a) for a full analysis. A dedicated sensitivity study of how the vertical placement of the GW drag influences the SPV would be valuable but lies beyond the scope of the current work.

2. *The methodology in the paper is convincing and suggests a novel approach. However, a discussion on why this method was chosen over well-known clustering methods such as EOF, k-clustering, and self-organizing maps would be beneficial. What are the advantages?*

Thank you for the positive assessment and for prompting us to clarify the advantages. We have added this discussion to Section 2.3. Traditional EOF/k-means/SOM frameworks operate on gridded fields and therefore optimize gridpoint (or mode) similarity rather than geometric similarity. Consequently, they may assign geometrically similar vortices to different clusters when the structures are shifted or rotated. Our method constructs the feature space from Fourier descriptors of the boundary, yielding a low-dimensional, geometry-aware representation where geometrically similar vortices cluster together, even with modest displacement/orientation differences. This provides (i) richer morphology than a few scalar diagnostics, (ii) greater robustness to small spatial perturbations than gridpoint-wise measures, and (iii) direct physical interpretability of the features themselves. We then apply hierarchical clustering (Ward linkage) in this descriptor space to obtain stable, morphology-based classes, which we use to quantify geometry-conditioned impacts of GW hotspots. (See Section 2.3, lines 112-133)

3. *Most of the statistical discussion here relies on the consistency of the signals. However, the significance of these signals (e.g., GPH, zonal wind, and frequency) is questionable. For instance, the authors repeatedly emphasize the strong signal in zonal mean zonal wind, yet the magnitude is less than 1 m/s in most figures. Considering the strong and highly variable winds in the stratosphere, can this signal be significant enough to lead to a meaningful change in circulation?*

We thank the reviewer for this insightful comment. We agree that statistical significance tests can provide valuable complementary information. However, in this study, we focus primarily on the consistency of the signals across ensemble members, rather than on the amplitude-based statistical significance. This choice is motivated by the inherently high variability of the stratosphere, where large background fluctuations can obscure physically meaningful but spatially coherent responses.

In our analysis, a signal is considered consistent if at least five out of six ensemble members exhibit anomalies of the same sign as the ensemble mean. This approach highlights robust dynamical responses that persist despite the model's internal variability. Figure 1 (below) illustrates this for the 10 hPa GPH anomalies (adapted from Figure 3 in the manuscript). The first row shows ensemble-mean anomalies, with dotted regions indicating consistent signals. The second row presents the same anomalies, but shaded only where they are statistically significant based on two-sample t-tests (Student's t for equal variances and Welch's t for unequal variances, with $p < 0.05$).

As shown, the spatial structures identified by the consistency criterion are broadly similar to those based on statistical significance, but they are less dependent on the absolute anomaly magnitude. This difference is expected because significance testing primarily reflects the mean amplitude of differences, whereas the consistency approach emphasizes reproducibility across ensemble members and is less sensitive to distributional assumptions. Alternative bootstrapped significance tests, while computationally demanding, are expected to yield similar spatial patterns.

Therefore, for this study, where the focus is on ensemble-mean responses under strong internal variability, the consistency criterion provides a more robust and physically meaningful measure of signal reliability.

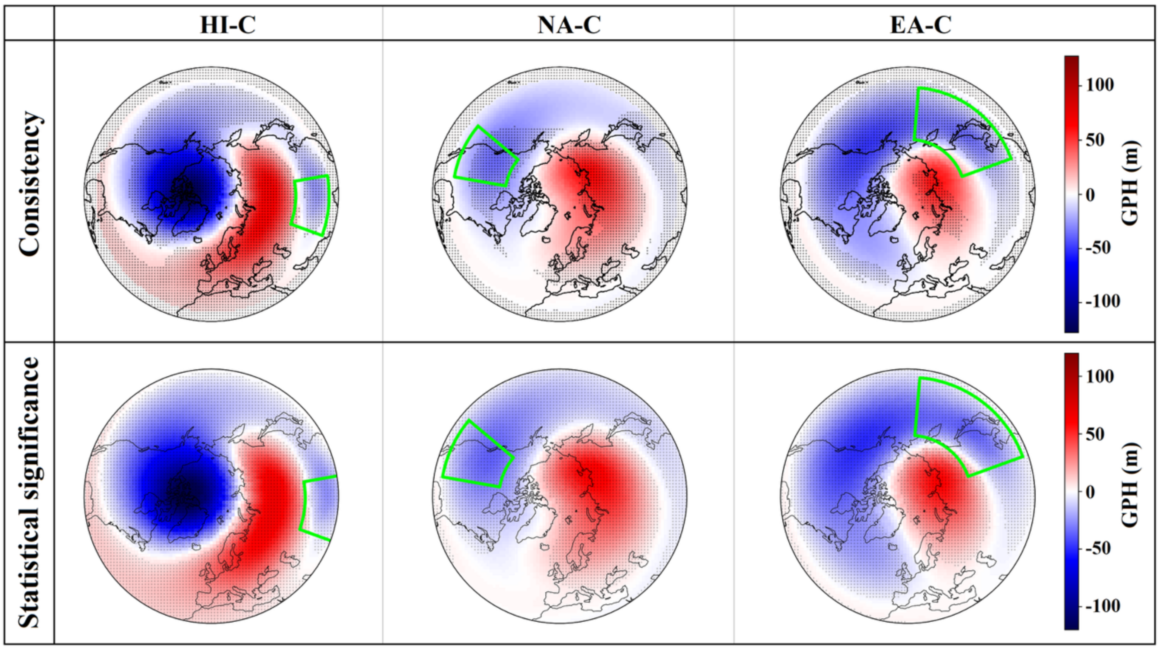

Figure 1: Ensemble-mean 10 hPa GPH anomalies for the Himalaya (HI-C, left column), Northwest America (NA-C, middle column), and East Asia (EA-C, right column) forcing experiments relative to the control run. The first row shows the ensemble-mean anomalies with dotted regions indicating areas of consistent response. The second row shows areas where anomalies are statistically significant based on two-sample t-tests ($p < 0.05$). Green outlines mark the regions of enhanced SSO GW drag.

4. *Figure 6 shows that GW forcing seems to explain the zonal wind response over the hotspot and to some extent northward (L230-231). In higher latitudes, however, especially in the HI experiment, EP flux divergence does. What do the authors think led to this difference? L440-441 argues that the delayed adjustment is a potential cause of this inconsistent response, but how can this explain the latitudinal dependence of the balance between GW and resolved waves?*

Thank you for highlighting this subtle but important observation. Different forcing components dominate in different regions, including both resolved wave forcing (via EP flux divergence) and parameterized GW drag, encompassing both SSO and non-orographic sources. Changes in the zonal-mean zonal wind induced by any of these components can, in turn, modify the others through feedbacks on the mean flow and residual circulation.

These components are not distributed uniformly in space. In the zonal-mean frame, SSO GW drag typically exhibits a peak in the mid-latitudes near the valve layer, whereas resolved wave momentum deposition (EP flux divergence) tends to dominate at higher latitudes, where stronger wave propagation and convergence occur within the polar stratosphere. This distinction helps explain why GW forcing governs the response near the hotspot and its immediate northward extension, while EP flux divergence becomes dominant at higher latitudes—particularly in the HI experiment. These interactions are examined in detail in Mehrdad et al. (2025a), and the SSO component is now briefly discussed in the new Appendix A, which provides diagnostics of the vertical and latitudinal distribution of SSO GW drag.

Regarding the former Figure 6 (now Figure 5 in the revised manuscript), the anomalies are averaged over the extended winter and across all ensembles, thus reflecting time-integrated (and potentially delayed) responses that combine the effects of these components. As for the discussion at L440–441 (now Section 4.3 lines 445-446), this specifically concerns the C6 WCVC contribution to the HI zonal wind anomalies (Figure 9 in the revised manuscript, panel C6a) and the corresponding EP flux divergence (Figure 10 in the revised manuscript, panel C6a). We initially expected C6, which contributes most strongly to the zonal-mean zonal wind, to also display the largest EP flux divergence signal. However, the class contribution framework captures only instantaneous covariances between fields and cannot represent the delayed adjustments typical of wave–mean flow interactions. We have clarified this limitation in the revised manuscript.

5. *Overall, the mechanistic discussion is lacking. Why do we see certain anomalies in WCVC? What causes frequency changes? Why do they respond differently depending on the hotspot? While this may be outside the scope of this study, it would be better to express the authors' opinion and leave it as an open question.*

We thank the reviewer for their thoughtful comments. We agree that the Discussion in the preprint was dense, which impeded clear answers to the questions raised. In the revision, we have (i) divided the discussion section into shorter subsections, each focused on a specific aspect while retaining the links among them, and (ii) edited several paragraphs in the discussion to make the central message clearer. The lack of explaining the GW forcing in the paper and purely relying on Mehrdad et al. (2025a) also contributed to obscuring the message in the discussion, which has now been addressed in Appendix A and integrated into the discussion.

Below we address each specific question you raised, with pointers to the relevant parts of the manuscript.

Why do we see certain anomalies in WCVC? What causes frequency changes?

WCVC anomalies arise from forcing-induced internal structural changes within clusters, such as minor shifts in geometry or location that do not reassign days to different clusters, while frequency changes reflect shifts in cluster prevalence that favor certain morphologies (e.g., low-PW1 configurations). This dual response allows the forcing to both modify existing cluster characteristics and alter their occurrence, contributing to/favouring overall anomalies like PW1 amplitude reduction. See Section 2.4 for the methodological framework and Section 4.2 (paragraphs 1–4) for the application to our results; we have modified paragraphs 3 and 4 to enhance clarity on these mechanisms (lines 394-426).

Why do they respond differently depending on the hotspot?

Responses differ due to the hotspot's spatial position relative to the SPV geometry (with a possible role for the vertical distribution of the forcing), which shapes edge mixing, wave-mean flow interactions, and internal variability. For example, HI's southern position deepens the vortex via edge sharpening, while NA/EA's poleward placement allows deeper core intrusions. See Section 4.2 (paragraph 5, lines 427-434) for geometry-forcing interactions and Section 4.3 (paragraph 1, lines 436-446) for experiment-specific examples, including references to the vertical GW drag distribution in the new Appendix A; we have modified these paragraphs to integrate Appendix A more explicitly and clarify the mechanistic links.

Minor comments

L53: The first sentence of this paragraph does not seem related to the rest of it.

We thank the reviewer for this comment. We have modified the beginning of the paragraph to improve clarity and ensure that it is better integrated with the rest of the paragraph (lines 53-58 of the revised manuscript)

L145-146: Is rescaling not necessary for the boundary size and time feature?

We thank the reviewer for this insightful question. Rescaling was not applied to the boundary size or time features because we did not intend to emphasize their role in the clustering process, and their variances are comparable to those of the Fourier descriptors.

Figure 2. What do different colors mean?

We thank the reviewer for this observation regarding the dendrogram colors. To improve clarity, we have added a brief explanation to the figure caption in the revised manuscript, noting that the colors represent different cluster branches and their hierarchical relationships.

L219-221: divergence is reduced -> convergence is reduced?

We thank the reviewer for this comment. In this context, the phrase "divergence is reduced" is used intentionally (lines 244-246 in the revised manuscript). As described in the sentence, this refers to a positive EP flux divergence anomaly, meaning that the magnitude of divergence (typically negative, indicating strong resolved wave drag) becomes smaller—i.e., less negative—relative to the control run. This corresponds to weaker wave drag, as noted in the explanatory clause ("which indicates weaker wave drag"). Alternatively, this could also be described as an increase in convergence, but we have retained the original phrasing for consistency with standard usage in the literature.

Figure 7. The GPH differences in C7-C10 are difficult to discern. Therefore, it is difficult to determine if the boundary aligns with the GPH. Is it possible to show the discrepancy more clearly?

We thank the reviewer for this valid point regarding the visibility of GPH differences in clusters C7–C10 of the former Figure 7 (now Figure 6 in the revised manuscript). We acknowledge that the weaker GPH gradients in these split-vortex clusters can make alignment with the boundaries harder to discern. After considering options, such as using cluster-specific colorbars, we decided to retain the current uniform colorbar, as varying scales would complicate inter-cluster comparisons and make the figure busier overall.

L259-262: The first sentence states that C3 has a relatively circular vortex, but a later sentence says that C3 is strongly deformed. This is confusing and needs clarification.

We thank the reviewer for highlighting this. To resolve this, we have revised the paragraph (now located in Appendix D, lines 530-532 in the revised manuscript) to clearly distinguish C6 as circular and to emphasize that C3's deformation is mainly associated with PW1.

L349: How is SSW linked to C4 and C8? There is no earlier discussion about this. It would be better to either remove the statement or show the relationship in the supplementary material.

We thank the reviewer for this helpful comment. The SSW events were introduced in the Introduction (lines 30-33) and now explicitly highlighted in Section 3.2 of the revised manuscript (lines 266-267, 287-288) to clarify that they refer to the shape-based identification of displaced and split vortex configurations.

L383-392: First, it is argued that positive GPH anomalies north of the hotspot lead to the weakening and mixing of the SPV edge. However, for the HI experiment, a similar signal leads to the sharpening of the SPV. This needs to be clarified, as a similar signal led to a different response, and the two arguments conflict with each other.

We thank the reviewer for highlighting this point. To clarify, we have revised the paragraph (lines 392-393 in the revised manuscript) to emphasize that the apparent difference arises from the equatorward position of the positive anomaly band in the HI experiment—located south of the vortex edge over northern Eurasia—which leads to a sharpening rather than a weakening of the SPV. The pattern of a positive anomaly in a higher-latitude band relative to the forcing and a negative anomaly within the forcing region is consistent across all experiments; however, the relative position of these anomalies with respect to the SPV determines the nature of the vortex response (see also lines 427-434 of the revised manuscript).

L405-407: Same here. The SSW linkage to C4 and C8 must be shown explicitly first.

We thank the reviewer for this comment. As noted in our response to the previous comment, we have added a brief explanation in Section 3.2 of the revised manuscript.

L426-428: Could you clarify what it means to "follow mean circulation"?

We thank the reviewer for this comment. To address this, we have added a concise parenthetical explanation in the revised sentence (lines 430-431), clarifying that "following the mean circulation" refers to the geostrophic wind direction along the mean GPH isolines, as noted in the preceding sentence (lines 428-430).

L428-429: Again, the strengthening of the SPV edge conflicts with the sentence before. Please clarify this.

We thank the reviewer for this comment. As noted in our previous response, the effect of the forcing on the SPV depends on its latitudinal position relative to the vortex edge. Although a positive GPH anomaly in a higher-latitude band relative to the forcing is consistently found across experiments, its influence on the SPV differs depending on whether the anomaly lies equatorward or poleward of the vortex edge.

L442: C9 -> C8

Th We thank the reviewer for this observation. This has been corrected in the revised manuscript (line 447).

L444-446: Positive anomalies seem to exist at different longitudes.

We thank the reviewer for this observation. To address the comment, we have revised the sentence (lines 449-450 in the revised manuscript) to clarify that these anomalies occur in a higher-latitude band north of the forced region.

References

Mehrdad, S., Marjani, S., Handorf, D., and Jacobi, C.: Non-zonal gravity wave forcing of the Northern Hemisphere winter circulation and effects on middle atmosphere dynamics, Weather Clim. Dynam., 6, 1491–1514, https://doi.org/10.5194/wcd-6-1491-2025, 2025a.