

General comments:

I thank the authors for their detailed responses to the reviewers' comments. The revised version shows significant improvement and satisfactorily addresses my previous concerns. I have just a couple of suggestions regarding interpretation of the wave reflection and some terminology. I believe these points should be clearly clarified before publication.

Major Comments:

1. Interpretation of wave reflection

The mechanism linking different tropospheric responses to wave reflection appears to be primarily inferred from negative Fp anomalies. However, reduced upward EP-flux does not necessarily imply actual wave reflection (i.e., downward propagation). A more rigorous diagnosis of reflective behavior (e.g., refractive index or wind shear reversal, Hu & Tung 2002, Harnik 2009) would be required to substantiate the use of terms such as 'wave reflection' or 'wave reflection surface'.

Reference:

Hu, Y., and K. K. Tung, 2002: Interannual and Decadal Variations of Planetary Wave Activity, Stratospheric Cooling, and Northern Hemisphere Annular Mode. *J. Climate*, 15, 1659–1673, [https://doi.org/10.1175/1520-0442\(2002\)015<1659:IADVOP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1659:IADVOP>2.0.CO;2)

Harnik, N. (2009), Observed stratospheric downward reflection and its relation to upward pulses of wave activity, *J. Geophys. Res.*, 114, D08120, doi:10.1029/2008JD010493

Thank you for this important comment. We agree that reduced upward EP-flux anomalies alone do not provide a rigorous diagnosis of wave reflection.

In our analysis, the interpretation is based on absolute EP-flux, not anomalies. Specifically, persistent downward EP-flux is taken as a signature consistent with wave reflection (e.g. Harnik, 2009, Perlwitz and Harnik, 2003; Shaw et al., 2010), while acknowledging that a more rigorous diagnosis would require additional dynamical metrics such as the refractive index (Hu & Tung, 2002).

To reduce the influence of transient variability, we require that downward EP-flux persists over several consecutive days (≥ 5 days) and is preceded by upward EP-flux, thereby supporting its interpretation as reflective behaviour rather than short-lived variability. The hatched regions in the figures indicate areas of persistent downward EP-flux in the ensemble mean and form the basis of this interpretation. We have revised the caption and added text in the manuscript to clarify this (Lines 373–386, also see below).

In addition, we have added the zonal-mean wind contours to the EP-flux panels (Fig. 14), providing

additional context on the flow configuration associated with reflecting conditions. The zero-wind line is a simplified indication of reflective conditions, and in the first two examples (SSW 1 and 2), occurrence of negative EP-fluxes coincides well with winds turning negative. We further examined the refractive index (not shown), which confirms that the background conditions are indeed favourable for wave reflection also in SSW 3. For brevity, we decided not to include this diagnostic, but we now note in the manuscript (Lines 373–386, see below).

To improve clarity and avoid overinterpretation, we have revised the manuscript to better explain the diagnostic basis of our interpretation and clarified the terminology where appropriate.

Lines 373–386:

“To provide insight into the stratosphere-troposphere evolution following an SSW, key terms from the TEM momentum budget and corresponding EP-flux diagnostics for these simulations are shown in Fig. 14. These include the temporal tendency of the background zonal mean flow and its forcing by large-scale waves, as well as anomalies of the WN1 and WN2 EP-flux components. While these $-F_p$ anomalies highlight departures from the climatological state, the interpretation of reflection-like behaviour is based on the presence of persistent downward EP-flux in the absolute field. Specifically, hatched areas indicate regions where the ensemble mean exhibits downward EP-flux ($-F_p < 0$) in absolute terms for at least 5 consecutive days and is preceded by upward EP-flux, thereby distinguishing it from transient variability and indicating wave propagation consistent with wave reflection (e.g. Perlwitz and Harnik, 2003; Shaw et al., 2010; Harnik, 2009). While this diagnostic does not constitute a rigorous identification of wave reflection, it is consistent with established interpretations based on EP-flux behaviour (e.g. Perlwitz and Harnik, 2003; Shaw et al., 2010; Harnik, 2009). Additional analysis of the refractive index further indicates that the background conditions are favourable for reflection (not shown, see e.g. Hu and Tung, 2002; Harnik, 2009). The meridional residual circulation term in the TEM momentum budget (f_{0v}^ , see Eq. 1 in Sec. 2.2)) is not shown, as it effectively balances the combined WN1 and WN2 EP-flux divergence contributions.”*

References:

Perlwitz, J., and N. Harnik (2003). Observational Evidence of a Stratospheric Influence on the Troposphere by Planetary Wave Reflection. *J. Climate*, 16, 3011–3026, [https://doi.org/10.1175/1520-0442\(2003\)016<3011:OEOASI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3011:OEOASI>2.0.CO;2)

Shaw, T. A., J. Perlwitz, and N. Harnik (2010). Downward Wave Coupling between the Stratosphere and Troposphere: The Importance of Meridional Wave Guiding and Comparison with Zonal-Mean Coupling. *J. Climate*, 23, 6365–6381, <https://doi.org/10.1175/2010JCLI3804.1>

Harnik, N. (2009). Observed stratospheric downward reflection and its relation to upward pulses of wave activity, *J. Geophys. Res.*, 114, D08120, <https://doi.org/10.1029/2008JD010493>

Hu, Y., and K. K. Tung (2002). Interannual and Decadal Variations of Planetary Wave Activity, Stratospheric Cooling, and Northern Hemisphere Annular Mode. *J. Climate*, 15, 1659–1673, [https://doi.org/10.1175/1520-0442\(2002\)015<1659:IADVOP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1659:IADVOP>2.0.CO;2).

2. Interpretation of the ‘initial intensity’

The manuscript concludes that the ‘initial intensity’ of the SSW is less important than the lower-stratospheric anomaly in determining the surface response. Given that the correlation remains relatively strong ($r \approx -0.71$), the statement regarding the limited role of intensity should be more cautious. In addition, the metric used to represent ‘initial intensity’ is a two-week mean of U10, based on absolute values rather than anomalies. Given that it does not consider how abnormal it is relative to climatology (i.e., anomaly), and the fact that it also reflects the persistence of SSW, this metric is not strictly an instantaneous measure of SSW amplitude. The terminology should be made more precise, and the interpretation should better reflect what this metric actually represents.

Thank you for this helpful comment. We did not intend to understate the role of SSW intensity given the relatively strong correlation ($r \approx -0.71$) in the original manuscript. We have revised the text to express more deliberately that both upper-stratosphere (SSW-intensity) and lower-stratospheric anomalies are related to the surface response, while the latter shows a more consistent relationship across events (Lines 310–317):

“The lower stratospheric GPH anomalies in the second week after SSW onset display a particularly strong relationship with the surface anomalies averaged over weeks 3 to 7, with a Pearson correlation coefficient of 0.85 (Fig. 10a), which is statistically significant ($p < 0.01$). In comparison, the relationship between the surface anomalies and post-SSW vortex strength, quantified by the two-week mean zonal-mean zonal wind at 10 hPa and 60° N, is also substantial, with a negative correlation coefficient of -0.71 (Fig. 10b) and statistically significant ($p < 0.01$). This metric reflects the magnitude and persistence of the post-SSW vortex rather than an instantaneous amplitude. While both metrics are clearly related to the surface response, the lower-stratospheric state exhibits a stronger relationship across events.”

In addition, we have clarified the terminology by replacing “initial intensity” with “intensity” throughout the manuscript. The metric used (two-week mean U1060) does not represent an instantaneous amplitude but rather reflects both the magnitude and persistence of the event, and this is now explicitly stated in the text (see Section 2.2 *Definition of metrics*, Lines 165–169):

“The SSW intensity metric U^{10}_{60} is defined as the two-week mean zonal-mean zonal wind at 10hPa and 60°N following the event onset. As such, it reflects both the magnitude and persistence of the wind reversal rather than an instantaneous amplitude. Absolute values are used, as the zero-wind threshold represents a physically meaningful reference for SSW conditions, largely independent of the timing within the season. Using U^{10}_{60} anomalies instead of absolute values yields qualitatively similar results (not shown), indicating that our conclusions are not sensitive to this choice of metric.”

Regarding the use of absolute values, we chose this definition because the zero-wind threshold is physically meaningful for characterizing SSW conditions, largely independent of the time of occurrence within the winter season. Nevertheless, in addition to the analysis based on absolute U1060 values, we tested the sensitivity of our results using U1060 anomalies instead of absolute values. We find that the relationship with the surface GPH response and the overall conclusions remain qualitatively similar. To illustrate this, we include an additional scatter plot of U1060 anomalies versus surface GPH (see Fig. 1 below) in this response.

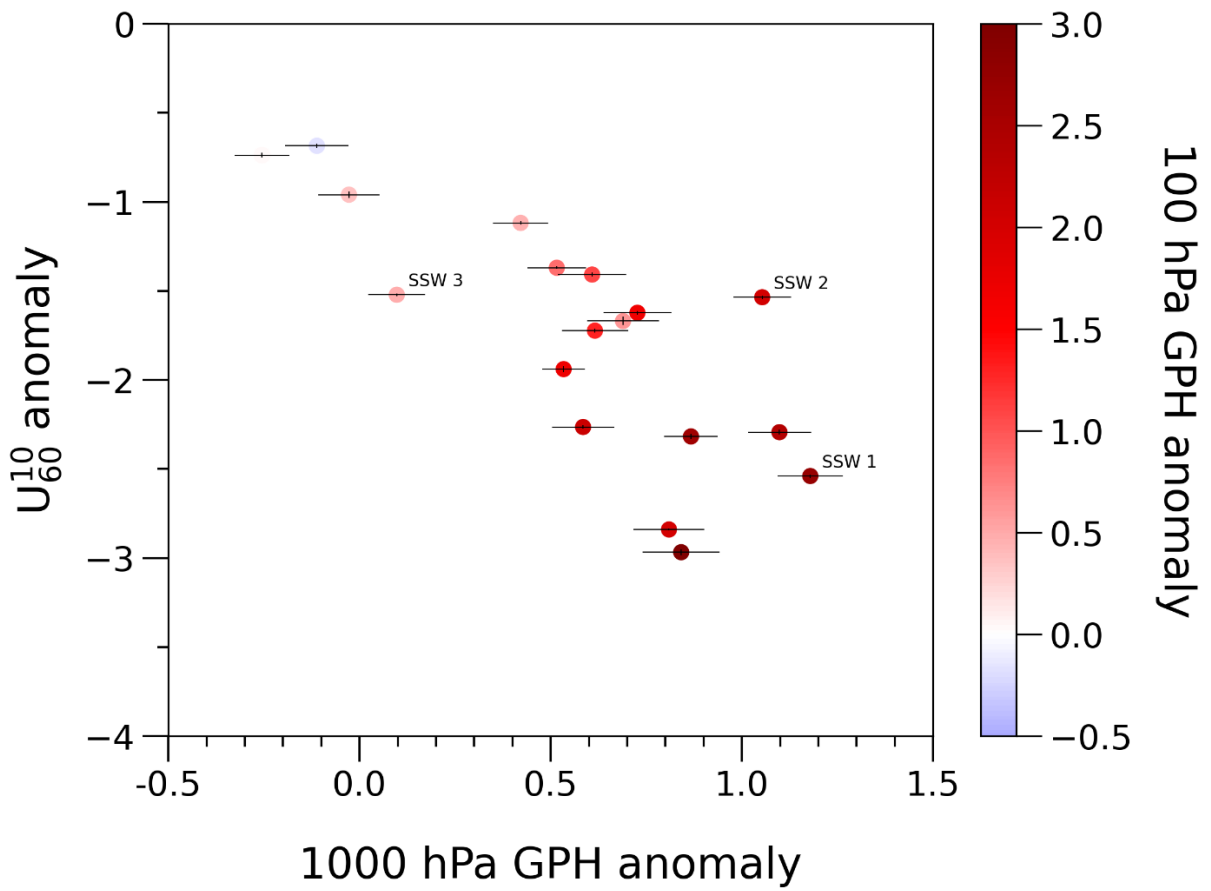


Fig. 1: Scatter plot of zonal-mean zonal wind anomalies at 10 hPa and 60°N (averaged over the first two weeks) versus 1000 hPa GPH anomalies (averaged over weeks 3-7), with color indicating 100 hPa GPH anomalies (averaged over week 2). Error bars depict the standard error of the ensemble mean. The correlation coefficient between 1000 hPa GPH anomalies and zonal-mean zonal wind anomalies at 10 hPa and 60°N is -0.77. Marker

Specific comments

1. L133-140. As the spin-off setup ‘guarantees that each ensemble member captures the chosen SSW identically.’ I wonder whether this is because the tuning factor results in very small perturbations in the stratosphere, or whether the perturbation was only applied to the troposphere. A brief clarification would be helpful.

The tuning factor controls the amplitude of the applied perturbations and is uniform throughout the entire atmosphere. Therefore, it does not explain why the SSW evolution appears nearly identical across ensemble members. This behaviour is related to the inherent predictability of the stratosphere rather than to the tuning factor itself:

The spin-off setup “guarantees that each ensemble member captures the chosen SSW identically” primarily due to the fact that the ensemble initialization is performed on the day of the SSW, meaning that the dynamical event has already occurred. In addition, the stratosphere has relatively high intrinsic predictability, with typical predictability limits of about two weeks (see e.g., Garny 2025; Rupp et al. 2023). As a result, the initial (≥ 2 weeks) stratospheric evolution remains very similar across ensemble members.

We have revised the manuscript to state this more clearly (Lines 141–147):

“This approach of generating spin-off ensembles guarantees that each ensemble member captures the chosen SSW identically, primarily because the ensemble is initialised on the day of the SSW (such that the dynamical event has already occurred), and because the stratosphere exhibits relatively high intrinsic predictability on timescales of up to about two weeks (see e.g., Garny, 2025; Rupp et al. 2023). As a result, the initial (≥ 2 weeks) stratospheric evolution remains very similar across ensemble members. This setup furthermore provides the basis for an improved statistical characterisation of the possible tropospheric evolutions following a stratospheric event, and crucially, enables a direct comparison of conditional tropospheric response distributions across different SSWs.”

References:

Garny, H. (2025). Intrinsic predictability from the troposphere to the mesosphere/lower thermosphere (MLT). *Journal of Geophysical Research: Atmospheres*, 130, <https://doi.org/10.1029/2025JD043363>

Rupp, P., Spaeth, J., Garny, H., & Birner, T. (2023). Enhanced polar vortex predictability following sudden stratospheric warming events. *Geophysical Research Letters*, 50, <https://doi.org/10.1029/2023GL104057>

2. L183-185. According to Eq(1), a positive $\nabla \cdot F$ should lead to a positive zonal wind tendency, thus the acceleration. Please double check.

Thank you for pointing this out. We agree that, consistent with Eq. (1), positive $\nabla \cdot F$ implies a positive zonal-wind tendency, whereas EP-flux convergence corresponds to negative $\nabla \cdot F$ and thus to a westward (decelerating) force on the mean flow. We have corrected the text accordingly (Lines 192–194):

“Physically, EP-flux convergence corresponds to a westward (decelerating) force on the mean zonal wind due to breaking and dissipation of planetary waves. In Eq. 1, this corresponds to negative $\nabla \cdot F$, which contributes to decelerating the stratospheric westerly jet.”

3. L217. Should it be Fig. 2a?

Yes, thank you. This has now been corrected.

4. L230. The spread in the tropospheric GPH anomaly in SSW and non-SSW group are exactly the same. Does this imply that the variance is primarily driven by tropospheric processes?

Yes, the identical spread in the tropospheric GPH anomaly (± 0.97) for both the SSW and NoSSW groups implies that much of the variance is driven by internal tropospheric variability. While the stratosphere can influence the large-scale state and tends to push the system toward particular states, the troposphere itself remains highly variable.

At the same time, this does not imply that the stratosphere has no influence on tropospheric predictability. Previous work has shown that the state of the stratospheric polar vortex can modulate tropospheric forecast uncertainty with regionally varying impacts (Spaeth et al., 2024). In our analysis, the large spread likely also reflects that we consider multiple individual SSW events, which differ in strength and exhibit varying tropospheric responses. As a result, the overall variance in the tropospheric signal remains large in both groups.

Reference:

Spaeth, J., Rupp, P., Garny, H. et al. (2024). Stratospheric impact on subseasonal forecast uncertainty in the northern extratropics. *Commun Earth Environ* 5, 126, <https://doi.org/10.1038/s43247-024-01292-z>

5. L252. According to the figure caption and following analysis, should the 'on the day of the selected event' be 'first two weeks'?

Yes, thank you. This has now been corrected to (Lines 262–264):

"For each spin-off, Fig. 5 shows the stratospheric conditions post-SSW in terms of the U_{60}^{10} index averaged over the first two weeks of the selected event."

6. L257. This is related to my major comment 2. While the authors introduced the events in the upper left and right as the 'weak warming events', giving that it is the 2weeks' average U_{10} , it does not necessarily indicate a weak warming event. For example, it can be a strong but rapidly recovered SSW. It might be better to describe it more accurately.

Yes, in theory this is possible. However, in our case, we verified that all events in the upper left and right quadrants are weak, with only one event briefly reaching strong conditions in the first few days before recovering. To reflect this more accurately, we revise the text to (Line 268):

"The upper left and right quadrants capture brief or weak warming events that are followed by negative or positive (i.e. weak to strong) tropospheric signals, respectively."

7. L301. Similarly, is it accurate to use the term 'initial vortex strength'? In addition, 0.71 still seems to be a high correlation. How about the statistical significance?

Thank you for your comment. We have revised the wording to refer to post-SSW vortex strength and clarified its definition (see Major Comment 2). We also note that the correlation ($r = -0.71$) is statistically significant ($p < 0.005$), and have added this information to the manuscript in Lines 310–317:

"The lower stratospheric GPH anomalies in the second week after SSW onset display a particularly

strong relationship with the surface anomalies averaged over weeks 3 to 7, with a Pearson correlation coefficient of 0.85 (Fig. 10a), which is statistically significant ($p < 0.01$). In comparison, the relationship between the surface anomalies and post-SSW vortex strength, quantified by the two-week mean zonal-mean zonal wind at 10hPa and 60N, is also substantial, with a negative correlation coefficient of -0.71 (Fig. 10b) and statistically significant ($p < 0.01$). This metric reflects the magnitude and persistence of the post-SSW vortex rather than an instantaneous amplitude. While both metrics are clearly related to the surface response, the lower-stratospheric state remains more consistent and exhibits a stronger relationship across events.”

8. L324. How were the 18 SSW events selected? In Figs. 11 and 12, the LS and no-LS groups appear evenly split (9 vs 9). However, the selection criteria for the 18 events are not fully explained. Please clarify this.

The equal split between the LS and no-LS groups (9 vs 9) is coincidental and was not imposed as a selection criterion.

The 18 SSW events were selected to provide a representative sample spanning the range of characteristics relevant to this study. Specifically, events were chosen to span different winter months (December–February), a range of SSW strengths, and varying magnitudes of lower-stratospheric (LS) and surface signals. This ensured that the sample includes events with different combinations of strong and weak stratospheric and surface responses (e.g., strong SSWs with strong or weak LS signals, events with strong LS but weak surface signals, and events with weak LS but pronounced surface signals).

SSWs occurring in March were excluded to avoid including final warming events.

9. Figure 12. The figure caption seems uncompleted.

Thank you for pointing this out. The caption has now been corrected and the missing part has been added.

10. L351-353. The introduction about the use of EP flux was explained again in later paragraphs, which is sufficient. So you might want to delete here to avoid repeat.

Thank you for pointing this out. To avoid repetition, we have removed the introductory sentence on the use of EP-flux diagnostics in this section, as this information is described in more detail later in the text.

11. L367 and onwards. Here the analysis is based on the EP flux anomalies. As discussed in my major comment 1, the negative F_p anomalies do not indicate wave reflection. Please check carefully with the appropriate metric (i.e., refractive index, refractive surface).

Thank you for this comment. We agree that negative F_p anomalies do not demonstrate wave reflection. As clarified in our response to Major Comment 1, our interpretation is based on the occurrence of persistent downward EP-flux in absolute values, not on anomalies.

To make this clearer, we have revised the text accordingly, added the zonal-mean zonal wind contours to the EP-flux panels, and clarified that the hatching marks regions of persistent downward EP-flux in the ensemble mean. We have also checked the refractive index, which supports that the background conditions are consistent with reflection-favourable wave propagation.

12. L449-451. The wording here suggests a causal link from zonal-mean flow changes to blocking frequency. However, earlier (L340-341) the causality was described as uncertain. Please ensure consistency in the interpretation.

Thank you for pointing this out. To avoid implying causality, we have revised the text to emphasise a statistical co-occurrence rather than a causal sequence, see Lines 466–469:

“Combining the above results, the robust and near-linear relationship identified here implies a consistent sequence of linked signals: (i) an anomalous lower-stratospheric state modulates the zonal-mean tropospheric flow and favours a pronounced zonal-mean surface response, together with (ii) an increased likelihood and intensity of regional atmospheric blocking.”

13. Please ensure that all supplementary figures are explicitly cited in the main text.

Thank you for pointing this out. All supplementary figures are now explicitly cited in the main text.

EGUSPHERE-2025-4164

Author Responses: Second Review

Peer Review – Loeffel et al. 'Case-to-Case Variability in the Tropospheric Response to Sudden Stratospheric Warmings Revealed by Ensemble Re-Forecasts'

Second review on “Quantifying the tropospheric response to individual sudden stratospheric warmings revealed by an ensemble simulation strategy”

The authors have well revised their paper with all of my comments considered. I have only one comment:

The authors say they produce 120 ensemble members for the forecast, which is a little misleading. Actually, there are only 40 members for each initialization of the three (1, 2, 3 Oct 2020, respectively). Please clarify.

Further, the authors only picked up 18 SSWs, which is much fewer than observation, 0.6 per year. Please add some discussion if you simulate more SSWs but you only picked 18.

Thank you for your comment. In the manuscript (Lines 104–105), we state:

“We constructed the EGE as a set of three time-lagged 40-member ensembles initialised with realistic atmospheric conditions taken from Oct. 1st, 2nd and 3rd 2020 and run through to May 31st 2021”

To avoid misleading phrases, we have clarified this in the revised manuscript. While the ensemble consists of three separate initialisations (3 × 40 members), this distinction is not critical for the purposes of our analysis. After approximately one month, the simulations lose memory of their initial conditions, such that all members can be treated as part of a single, statistically consistent ensemble. Combining the three sets is therefore primarily a technical choice to increase the sample size. We have added the following sentence for clarity (Lines 107–109):

“After the first month, the simulations no longer retain sensitivity to their specific initialisation dates, allowing all ensemble members to be analysed as a single ensemble; the use of three initialisation dates is therefore primarily a technical approach to increase ensemble size.”

Regarding the number of SSWs, we clarify that a total of 57 SSWs are simulated in the full ensemble (see manuscript, Line 96: “We identify 57 SSWs occurring in the EGE”).

From these, we restrict our analysis to events occurring in December, January, and February, and select 18 events for detailed investigation. This subset was chosen to span a range of event characteristics (e.g. vortex strength, lower-stratospheric response, and surface signal).