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

General comments:

This study focuses on the variability in the surface response following sudden stratospheric warming (SSW), using the ensemble re-forecasts. The authors identify that the lower stratospheric anomalies can be an effective predictor for the surface response, and attribute the mechanism to the height of the reflective surface. The experiments are nicely designed and the results have the potential to deepen understanding of under which condition an SSW may have strong influence on the surface. However, while the authors highlight the aim of 'quantify' the variable surface response, the analyses are mostly based on ensemble means, so the advantage of using ensemble forecasts is largely masked. Moreover, several strong conclusions are not sufficiently supported by quantitative evidence. In addition, the initial condition setup in the model experiments poses non-negligible limitations and requires further discussion. Hence, I would recommend a major revision, to make the most use of the ensemble forecasts and to provide stronger evidence for both the role of lower-stratospheric anomalies as a predictor and the proposed mechanism. Please see my detailed comments below.

Major Comments:

1. Deterministic vs probabilistic analyses. The analyses are mostly based on the ensemble mean. While I understand the authors wanted to rule out the influence of the tropospheric initial condition, the advantage of using ensemble forecasts has been masked. Especially given that the authors highlight the aim of 'quantifying' the variability in the abstract and introduction, this mismatch reduces the significance of this work. More importantly, the key conclusion regarding the role of the lower stratospheric anomalies is based on 18 samples (the ensemble means), which limits the statistical robustness of the results. The proposed mechanism involving the height of reflective surface is based on three samples, which is insufficient for a general conclusion. I suggest the authors make the full use of the large sample size, to better reflect both the uncertainty and the robustness of the results. For example, aside from the scatter plots (e.g., Fig. 10), I recommend the authors stratify the ensemble members for all the selected SSWs based on the lower stratospheric response, and show the temporal evolution of the surface response to check if there is a systematic shift in the distribution of surface response. In this way, it can not only provide stronger evidence to support the role of lower stratospheric condition, but also present a quantitative result on how likely the response would differ (i.e., the probability question arise in the manuscript). Please refer to my specific comments.

2. The interpretation on the role of external variability. The authors highlight that the role of external variability, including ENSO and QBO, has been excluded given that the initial atmospheric condition is the same and the SST has been set to a climatological state. However, the initial conditions from October 2020 possess a specific QBO state, which is then inherent in the model setting for all members. As previous studies have highlighted the role of the QBO influencing the surface response to weak polar vortex events (e.g., Ma et al. 2024), I believe a discussion is needed to better reflect this limitation and contextualize the results here.

Reference:

Ma, J., Chen, W., Yang, R. et al. Downward propagation of the weak stratospheric polar vortex events: the role of the surface arctic oscillation and the quasi-biennial oscillation. Clim Dyn, 62, 4117–4131 (2024). https://doi.org/10.1007/s00382-024-07121-5

3. Manuscript formatting and readability. The manuscript requires a thorough proofreading and correction for formatting and citation errors, which currently detract from the paper's professionalism and readability. For instance, there are several inconsistencies in citation style and grammar, e.g., L363, (Maycock and Hitchcock, 2015, see e.g.,) should be (see e.g., Maycock and Hitchcock, 2015).

Specific Comments:

1. L20: This statement on the evolution timescale is ambiguous. The authors seem to be referring to the difference in ‘memory’ or ‘predictability timescales’, not just ‘evolution timescales’, as tropospheric blocking can also persist for weeks.

2. L45-50: Please include the related reference.

3. Section 2: I suggest the authors move the description of the reanalysis dataset to Section 2.1 (i.e., before introducing the model setup). Otherwise, it is unclear what baseline climatology has been used in the model simulation.

4. L76: As mentioned in my major comment 2, the initial atmospheric condition already contains a specific QBO phase. Thus, some discussion should be included, perhaps following L89 where the authors currently highlight that the influence of QBO has been excluded.

5. L96: Does day d_0 indicate the calendar day? If so, please clarify this.

6. L120: ‘Butler et al. (2015)’ should be ‘(Butler et al., 2015)’

7. L127: Are the results sensitive to the choice of time periods and the threshold? By ‘Sensitivity to the choice of time periods in further discussed in Section 5 (see Fig. 9)’, I was expecting the authors to show a sensitivity test on these parameters there, but Fig. 9 actually

shows the correlation between the zonal mean circulation with the weeks 3-7 response. A more objective way to define the threshold and period could be to separate the SSWs based on the surface response, and then use the features of the strong surface response group to define the parameters.

8. L164: ‘Butler et al. (2017)’ should be ‘(Butler et al., 2017)’.

9. Figure 1: The shading in this plot cannot be seen clearly, especially for the EGE. I suggest modifying the plot for visualization. In addition, I suggest also including the histogram for the ERA5 for a more straightforward comparison between the EGE and reanalysis.

10. L174: ‘decreased AO’ is not precise, do you mean the ‘weakening of a positive AO’?

11. L175: Suggest using ‘SSW’ consistently, instead of referring to these events as ‘weak vortex events’ and ‘stratospheric event’ (L184).

12. Figure 2: How is the climatology defined in these plots, and what is day 0 for the climatology? In addition, the lower panel presents the raw AO index, where an initial positive AO is seen that should be related to the initial tropospheric condition. This might also explain the very noisy evolution of the AO index. Since normally we would expect a negative AO anomaly following an SSW, it might be more straightforward to show the AO anomaly instead. Also, it might be more informative to use shading to present the spread.

13. Figure 3: Suggest showing the plot for the non-LS group as well.

14. Figure 4: How is the climatology defined? Please clarify this.

15. L225: It’s a bit awkward to say ‘reduction of the zonal circulation’, perhaps ‘weakening’ is clearer.

16. Figure 6: Why are there very positive zonal winds? It is surprising to see the westerly wind during an SSW, is it sensitive to the duration (e.g., two weeks average has been presented here)?

17. L254-255: Are the results sensitive to the selected period (i.e., weeks 3-7)? According to Fig. 9, the correlation coefficient reduces dramatically after around day 30 in the troposphere. This means that the surface response in weeks 6-7 differs from the weeks before. This challenges the selection of using the weeks 3-7 mean.

18. Figure 10: As only the ensemble mean has been shown, it limits the sample size and the robustness of the conclusion. Suggest showing the individual member to better reflect the uncertainty. In addition, Fig. 10b show the linkage between the raw U10 and the surface response. Given the strong seasonal cycle in the stratosphere, the weaker warming (i.e., more positive raw U10) does not necessarily reflect a weaker departure from the climatology. How about the linkage between the anomalous U10 and surface response?

19. L261-264: This question naturally leads the readers to think about the mechanism at this stage. But in the following section, the regional response is shown instead of the mechanism. This narrative flow is a bit disconnected. Also in the abstract and the summary, the mechanism follows the zonal mean circulation response, which is further followed by the regional response. Will this flow work better?

20. L296-297: This is because of they both reflect the intensity change in the stratospheric polar vortex. This statement is confusing. It seems to be stating a known fact from the literature, not a new finding.

21. Section 7: As mentioned in my major comment 1, the analyses in this section are based on three cases, which cannot really represent the general mechanism. I agree that these case study can provide some hints, but a more systematic/objective analysis is needed to further support the hypothesis in the role of wave reflection and height of reflective surface. One possible approach is to calculate the height of the reflective surface, and stratify the members (for all the 18 SSWs) based on the height of the reflective surface, then check the evolution of lower stratospheric anomaly and surface response. If the distribution of these two groups can be distinguished, then it would provide stronger evidence.

22. Figure 14 caption: 'where the ensemble mean EP-flux is positive'. Should it be the ensemble mean vertical component of EP flux is negative?

23. L363. Revise the citation format.

24. L376-377: Figure 9 only presents the linkage between the zonal mean circulation and the surface response, so it cannot really support the strong statement that the initial tropospheric circulation patterns play a less critical role. In fact, in previous study (Ma et al. 2024), the preceding tropospheric circulation is shown be an important factor in influencing the surface response.

Reference:

Ma, J., Chen, W., Yang, R. et al. Downward propagation of the weak stratospheric polar vortex events: the role of the surface arctic oscillation and the quasi-biennial oscillation. Clim Dyn, 62, 4117–4131 (2024). https://doi.org/10.1007/s00382-024-07121-5

25. L391-393: Please add the related reference.

26. L395: A related plot would be helpful. And why do the simulations still indicate an increased chance of cold spells over Scandinavia even though there is not more frequent Scandinavian blocking?

27. L402: What does Fig. 13a refer to? There is only one panel in Fig. 13.

28. L431-432: The correlation coefficient does not really have dramatic difference (r=0.85 vs. r=-0.71). Given the very limited sample size (N=18), it is difficult to draw this strong conclusion.