REVIEWER 2

General comments:

This study analyzed how the stratospheric condition influences the upward wave flux, and thus the occurrence of SSWs, based on a set of S2S model experiments in SNAPSI. The S2S forecasts are still struggling to predict SSW with long lead times. Thus, the current results provide helpful insights into understanding the role of stratosphere, which could potentially help improve the forecast skill in the future. The manuscript is very clean and well-written. However, since there are various diagnostic metrics and a huge amount of information, a major revision is needed to help with the presentation and delivery of the main conclusions.

Thanks a lot for your comments. In the revised version of the manuscript, we have worked to make the figures more informative and simplify, when possible, the discussion. In this sense, a large part of Section 5 has been rewritten and a better link has been established between model spread and model mean of the presented variables. Here is our reply to the major and minor comments in blue:

Major Comments:

1. Balance in presenting the model mean and model spread. Currently, most analyses are based on the model mean, while the model spread is not fully presented. For example, in Section 3, while Section 3.1 shows the mean skill of models in predicting zonal wind, Section 3.3 focuses on the difference in Z500 between 'weaker u' and 'strong u' groups. This 'inconsistency' poses a mismatch in the obtained information. It would be helpful to link the spread in Z500 with the spread in zonal wind. Thus, the model spread in predicting the zonal wind can be added in Fig. 1, and the model mean of Z500 should be included in Figs. 5-7. Although the fraction of members forecasting an SSW is shown in Table 3, this cannot accurately reflect the spread. For the latter, including the model mean will help in interpreting Figs. 5-7 as discussed in L360-365.

We thank the Reviewer for their valuable suggestion. In the revised version, we have included both types of information: the ensemble spread in zonal wind predictions and the ensemble mean of Z500 anomalies prior to SSWs.

Including the ensemble spread in u60_10 for each day and each model in Figure 1 is not straightforward as we have eight models plus two more ERA5 reference lines. Adding shadings to represent the ensemble spread would likely cause substantial overlap among them. However, we agree with Reviewers 1 and 2 on the necessity of illustrating the model spread for at least a relevant time period such as the dates around the occurrence of each SSW in ERA5 (12^{th}-16^{th} February 2018, 2^{nd}-6^{th} January 2019 and 18^{th}-22^{nd} September, same as those used to identify the "strongest u" and "weakest u" members). To do so, we have included three additional panels in Figure 1 (Fig.1d-f) showing the ensemble distribution of u60_10 for those dates. A brief discussion about the ensemble spread in each case has been added in the L251-260 of the revised manuscript that complements the analysis of the models forecast skill of the three SSWs.

As for the ensemble mean of the Z500 pattern, we have now included this mean for each model in contours in Fig. 5-7. This facilitates the interpretation of the spread in the results, as it allows direct comparison between the mean values and the differences between the two groups of ensemble members ("weakest u" and "strongest u"). A corresponding discussion has been added accordingly.

Finally, we have connected the spread in Z500 with that in u60_10 by including, in each plot for Fig 5-7, the differences in the averaged u60_10 between "weakest u" and "strongest u" groups below each model name. This enables us to rapidly assess whether weak values in the Z500 difference patterns are due to low spread in both the stratosphere and troposphere, as it is the case for GLOBO during SSW2018, or whether the similarity is only restricted to the troposphere suggesting other processes may contribute to the deceleration of the vortex, as appears to be the case for most models during SSW2019 SH.

2. Discussion on the stratospheric state. While I understand that the main focus of this study is to provide a general understanding of the role of the stratospheric state, it would be worthwhile to add more discussion on what might constitute the stratospheric state. For instance, for the 2018SSW, the poleward shift of the PNJ is missing in the models, which is one possible candidate. In addition, the manuscript did not discuss about the QBO (although very briefly in L686), which is also included in the stratospheric condition and reflected by the experiment design. Although at such a short timescale the bias in the QBO may not be evident, the three SSWs occurred under different QBO states (e.g., Butler et al. 2020; Shen et al. 2020), and it would be worthwhile to reflect this and discuss the potential implications. In addition, as briefly discussed in Section 3.1, the intensity of the polar vortex relative to climatology also differs, which could also serve as preconditioning. The related discussion should be added to provide general implications.

Reference: Butler AH, Lawrence ZD, Lee SH, Lillo SP, Long CS. Differences between the 2018 and 2019 stratospheric polar vortex split events. Q J R Meteorol Soc. 2020; 146: 3503–3521. https://doi.org/10.1002/qj.3858

Shen, X., Wang, L., & Osprey, S. (2020). Tropospheric forcing of the 2019 Antarctic sudden stratospheric warming. Geophysical Research Letters, 47, e2020GL089343. https://doi.org/10.1029/2020GL089343

In the "Summary and discussion" section, we have emphasized what we define as stratospheric state, referring to both the polar vortex and the Quasi-Biennial Oscillation (QBO) and described it for the three analyzed SSWs in L690-712 of the revised manuscript.

Additionally, we have included brief references to the stratospheric state throughout the manuscript, particularly in the detailed analysis of SSW2018. In this case, as also suggested by the Reviewer, we have extended the figures showing the EP flux and wind distribution up to 10S (new Figure 12). This extension allows us not only to see more clearly the PNJ structure but also show the tropical winds, where an easterly phase of the QBO in the middle stratosphere is observed. A corresponding comment has been added in the figure description (L558-560 of the revised manuscript), emphasizing the role of equatorial easterly winds as a barrier to equatorward wave propagation and, consequently, contributing to the confinement of waves within the extratropics.

Nevertheless, we prefer to keep the discussion of the QBO influence on the predictability of SSW concise, as another SNAPSI working group is specifically dedicated to addressing this topic in depth.

3. Visualization of the plots. There are quite a few plots which are all informative. However, for some plots it is difficult to identify the regions/features being discussed in the main text. I suggest the authors adjust the plots to make the information more straightforward, which will help readers grasp the key information more quickly. Please see the detailed comments in the Specific Comments.

Following Reviewer's recommendation, most of the figures have been modified to make them more informative. These are the modifications:

- Figure 1: three new panels with the ensemble distributions of u60_10 averaged for the period 0/+4 days after the SSW in each model have been included.
- Figure 2-4: cyan solid lines representing NAVGEM results have been converted to dashed lines to easily distinguish them from those of CESM2-CAM6
- Figure 5-7: the ensemble mean of anomalous Z500 for the days preceding the SSW for each model has been added in contours. This helps in interpreting the difference in Z500 between the "weakest u" and the "strongest u" members. Finally, we have also added the multimodel mean of all these difference plots.
- Figure 8: the marker representing NAVGEM has been changed to distinguish it from CESM2-CAM6. Three more panels have been added to analyze the linkage between tropospheric circulation, upper tropospheric wave activity (HF300) and the lower-stratospheric wave activity (HF100). To do so, we show the pairwise correlations of the predominant WN HF100 and HF300, and predominant WN HF300 and the Z500 anomalies corresponding to the tropospheric precursors.
- Figure 11: dashed vertical lines of different colors have been added to delimit the periods of interest discussed in the text and a magenta solid vertical line has been included to indicate the central date of each SSW.
- Figure 12 and 13 have been merged to favor the comparison of zonal mean u and EP flux distribution in the two periods of interest (3^{rd}-5^{th} Feb and 6^{th}-8^{th} Feb). The plots have been extended to 10ºS to see tropical winds configuration. Additionally, the refractive index has been removed from these figures to simplify the discussion.
- Figure 13 (old figure 14): a magenta solid vertical line has been added to indicate the central date of each SSW.
- Figure 14 (old figure 15) has been corrected as we found a small bug in WN2 Fz10 that also affected the net EP flux. The color of the median value line in each boxplot has also been changed to black to make it more visible. Further, the ERA5 reference line is now shown in magenta to distinguish it from the median value line and to ensure consistency with the boxplot formatting used in the other figures.

<u>Specific Comments:</u>

1. L210: Suggest explaining that the negative eddy heat flux indicates upward propagation in the SH. Despite the figure caption of Fig. 4, there is no explanation in the main text. Please add this.

We have added the following sentence in L212 of the revised manuscript: "HF is defined as positive for upward wave activity flux in both hemispheres"

2. L227: The classification of 'weakest u' and 'strongest u' is a bit counterintuitive, as the 'weaker' group corresponds to a better forecast, whereas the 'stronger' group corresponds to a worse forecast. Would it be better to define them as something like 'SSW-like' and 'no SSW', which is more straightforward?

We understand the Reviewer's concern. However, the suggested terminology ("SSW-like" and "no SSW") may be misleading, as in many cases, the "weakest u" members are still far from predicting an SSW, particularly, for the SSW2018. This can be better appreciated in the new panels d)-f) of Figure 1, which show the ensemble spread of u60_10 over the same periods used to identify the "weakest u" and "strongest u" ensemble members. For this reason, we prefer to retain the terms "weakest u" and "strongest u", but we have clarified their definition in the revised manuscript (L221-227).

3. L260, Figure 1: Suggest also showing the spread in predicted [u] as in my major comment 1.

Done. See the rest of our comments in the reply to the major comment 1.

4. L278: Suggest briefly stating that WN2 is mainly responsible for the 2018 SSW, otherwise it is a bit abrupt to focus on WN2 directly.

We added a clarification in L274-275 of the revised manuscript.

5. Figures 2-4: Since there are lots of lines, I suggest bolding the model being discussed in the main text for visualisation.

We agree with the Reviewer that many lines appear in the same figure. However, we believe that bolding the models discussed in the text could be misleading for the reader as bolding is typically associated with highlighting statistically significant results. Moreover, in many cases we refer to the general behavior across all models, so emphasizing only a subset of them may not be ideal and could hide the results of non-bolded models. To improve the interpretation, we have instead included the corresponding color line in the main text, when discussing results of specific models in L275-297 of the revised manuscript.

6. Figures 5-7:

• Suggest adding the variance of forecasted [u] among the members in the subtitles after the model's name. This could help provide a straightforward comparison and understanding of the linkage between the spread in tropospheric circulation and stratospheric response.

Since we have already shown the ensemble distribution of forecasted u60_10 in new panels d)-f) of Fig.1, we have added the difference in the mean of u60_10 between the "weakest u" and the "strongest u" members groups below the subtitles.

• Suggest also adding the box to indicate the region in ERA5, as the map is not very visual and thus takes time to identify the region of focus.

The boxes have been added.

• The climatological PWs are from ERA5. While I understand this is what one can do, it is also possible that the model bias in the climatological PWs can influence the interpretation of linear interference. Perhaps it would be better to include a brief discussion.

To avoid confusion, we have removed the climatological PWs from model plots. We have also added a warning when relating linear wave interference to the position of the Z500 anomalies in models with respect to the antinodes of climatological PWs in ERA5 in L368-369.

7. L348: Please clarify what 'this center of action' refers to.

Clarified

8. L380-410: Suggest checking the HF300 as well. As stratospheric wave forcing does not entirely originate from tropospheric forcing as discussed later on and shown in Yessimbet et al. (2022). It would be helpful to establish a linkage between the tropospheric circulation, the tropospheric wave forcing (HF300), and the stratospheric wave forcing (HF100).

Reference:

Yessimbet, K., Shepherd, T. G., Ossó, A. C., & Steiner, A. K. (2022). Pathways of influence between Northern Hemisphere blocking and stratospheric polar vortex variability. Geophysical Research Letters, 49, e2022GL100895. hRps://doi.org/10.1029/2022GL100895

We thank the Reviewer for this comment. We have examined the linkage between tropospheric circulation, the upper tropospheric wave forcing (predominant WN HF300) and the stratospheric wave forcing (predominant WN HF100). To do so, we compared the correlation of HF300-HF100 with the correlation of Z500-HF300 for each event and model. The values are now shown in the new panels d-f of Figure 8.

We indeed find interesting results that support our conclusions derived for the NH events. For instance, during SSW2019, there is a strong coupling between the tropospheric precursors and both HF300 and HF100, with very similar correlation values across models. This indicates that the troposphere played a dominant role in triggering this event. For SSW2018, although the correlations between Z500 and WN2 HF300 and between WN2 HF100 and WN2 HF300 are also high, they exhibit a wider spread than in SSW2019, particularly in the latter relationship. Together with the low correlation between WN2 HF100 and Z500, this suggests that the stratospheric wave activity during SSW2018 might be modulated by additional sources beyond the troposphere, at least in some of the models.

In the case of the SSW2019 SH, the results show that while the correlation between Z500 and HF100 is relatively high, the correlations involving intermediate variables, i.e. HF300-Z500 and HF300-HF100, are generally lower than in the NH events. A more detailed analysis would be required to fully understand this behavior, but it lies beyond the scope of the present study.

A detailed discussion of this result is included now in L406-430 of the revised manuscript.

9. Figures 8 and S1. I like Fig. S1, which is very informative. It not only confirms the linear relationships among the three variables but also indicates the spread among the model members. I suggest the authors move it to the main manuscript, perhaps merging it with Fig. 8 as they are related, and add more discussion on it. For instance, we can see that for SSW2018, the scatters are densely located in the upper level, indicating that the forecast [u] is overall quite strong and related to the weak HF (Fig. S1a). Moreover, for the HF and Z500, the scatters are located in the lower panel (Fig. S1d). Despite the weaker Z500-HF linkage, the Z500 is also quite diverse. For comparison, it would be better to use the same range for x and y axes across different cases, also add the corresponding ERA5 variables. This will also help to interpret the conclusion from Fig. 8.

We agree with the Reviewer that Fig. S1 is informative. However, following the revisions made in the manuscript, most of the information highlighted by the Reviewer, such as the ensemble spread in forecasted u60_10 and HF100 values for each model and SSW, is now presented in the boxplots of Figures 1 and 9, respectively, and discussed in the corresponding sections. Moreover, Figure 8 has been expanded in response to specific comment 8 of this Reviewer, so adding these scatter plots to the main text would require an extra figure. Given that the number of figures in the manuscript is already high, we have decided to keep Fig. S1 in the Supplementary Material. We also prefer not to fix the scale of x and y axes to show the linear relationships among the three variables for all events. Otherwise, in some events, we could not see in detail the distribution of the values for the ensemble members. As suggested, the ERA5 variables have been added in the new version of the figure.

10. L432-437. Please briefly explain how the quantitative changes are computed.

We have just quantified the fraction (in %) of the multimodel mean of HF100 in the CONTROL run represented by the mean in the NUDGED experiment. A brief indication has been added, in L456-457 of the revised manuscript, the first time this is mentioned.

11. L440: It is interesting that for 2019SSW SH, although the models have relatively good performance in capturing the tropospheric wave forcing (i.e., for the multiEnsM, the observation is within the 1.5IQR), they fail to capture the observed stratospheric wave forcing. This seems to imply that the stratospheric wave forcing does not completely come from the tropospheric wave forcing. Whereas in the other two events, the skill for HF300 and HF100 is more similar.

Thank you for the comment. We agree that this is an interesting result. However, since this section focuses specifically on the comparison between NUDGED and CONTROL experiment results, we prefer not to include additional discussion here in order to maintain clarity and readability. We have noted this point, however, in the updated discussion of Figure 8, which now includes information on HF300 as well (L406-430 of the revised manuscript).

12. L477: Suggest adding ‘for the majority of models’ for clarity.

Added

13. L480: Suggest changing ‘model simulates very similar anomalies to ERA5’ to ‘the ERA5 value lies within the IQR ..’

Changed

14. Figure 11: Suggest marking the period of interest.

We have added a solid vertical magenta line marking the central date of the SSW in ERA5, as well as four dashed lines (two yellow and two green) indicating the two periods of interest discussed in the text (3rd-5th February and 6th-8th February).

15. L537-538: Suggest changing ‘relatively weak wave activity that is simulated by reanalysis and models’ to ‘ …wave activity seen in the reanalysis and simulated …’

The whole section has been modified, so this sentence has been removed.

16. Figures 12-13: Suggest adding the box to indicate the regions for EP flux budget and changing the y-axis labels to hPa for consistency with the main text. In addition, suggest extending the latitudes to 10S to show the QBO structure, related to my major comment 2.

We have modified these figures following the Reviewers’ recommendations. In particular, the refractive index has been removed and Figures 12 and 13 have been merged and extended up to 10ºS. The corresponding description has been modified accordingly. The y-label of the new Figure 12 has also been changed to hPa. The box has been added in Fig. S5 as it is the one showing the exact period used for the EP flux budget.

17. Figure 14: Please make sure the vertical magenta line mentioned in the figure caption is visible.

Thanks for the reminder. The vertical magenta line has been made visible in the figure.

18. Figure 15: Suggest changing the color of the median value line for visualization. In addition, while currently we can see the difference in Fz and Fy, we cannot find their relative contribution to the net EP flux convergence. According to the previous analysis (Figs. 12, 13), the difference in the horizontal wave forcing is evident, however, it does not appear to play an important role according to this plot. This poses a gap between these two sections. I suggest the authors add a scatter plot of Fz100 vs Fy for each member, similar to Fig. S1. This might help in understanding their relative roles.

We thank the Reviewer for this valuable suggestion. The change in the color of the median value line (also implemented) allowed us to identify a minor bug in the code affecting the WN2 Fz10 and consequently, the associated net EP flux. Although the main conclusions remain unchanged, the corrected figure provides an additional indication of wave resonance in the stratosphere. Specifically, we now see that both Fy and Fz10 increase in the NUDGED runs relative to CONTROL and even FREE, whereas the change in Fz100 across the three experiments is small in most models. This key point has been highlighted in the discussion of the figure.

The updated results are also consistent with the new Figures 12 and 13. The enhancement of Fy associated with the weaker vortex begins on 6th February in ERA5 and was already evident in new Figure 12f. The increase in Fz in the middle stratosphere in the NUDGED runs starts around 8th-9th February, as shown in the new Figure 13. This cannot be detected in Figure 12f because it only covers 6th-8th February. To provide a complete picture, we have added a new supplementary figure (Fig. S5) showing the EP flux and zonal wind distribution averaged over 6th-14th February, the same period used in new Figure 14 (previously Figure 15). The corresponding discussion has been updated to emphasize consistency with the earlier results.

To sum up, we conclude that the main contribution to the strong convergence of net EP flux before SSW2018 arises from the lower stratospheric upward-propagating wave activity and this term is not strongly affected by the prescribed stratospheric state. Nevertheless, we find indications that the zonally symmetric stratospheric state modulates the wave activity at higher levels, amplifying it in agreement with wave resonance phenomena, but only in high-top models.

Finally, following the Reviewer's suggestion, we include the scatter plots of Fz100 vs Fy for the three experiments (Fig. R2.1). While the CONTROL and particularly, the FREE, experiments display an approximately linear relationship, the NUDGED runs deviate from linearity and show a much larger spread. This behavior may reflect the presence of non-linear processes such as wave resonance growth, as discussed above. We have opted not to include this figure (Fig. R2.1) in the manuscript because the number of figures is already high, and the corrected Figure 14 has reduced the gap between the current section and the previous one.

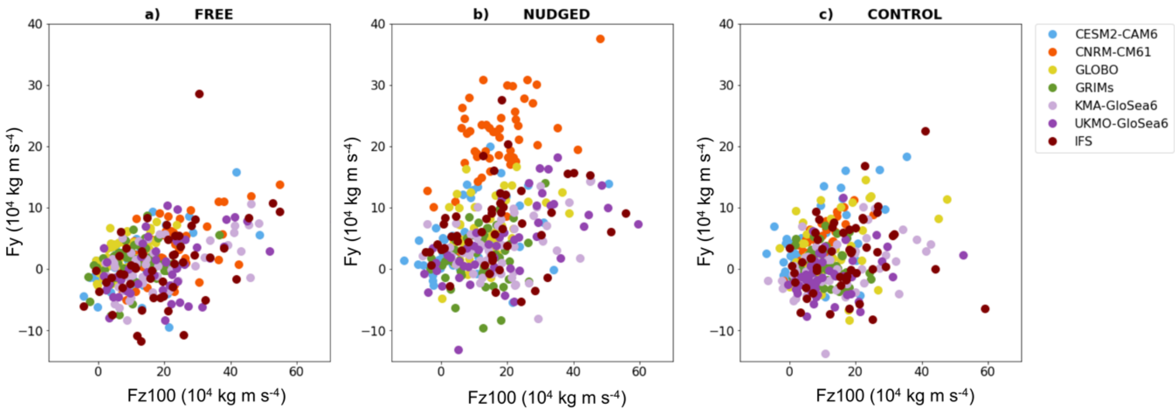

Figure R2.1. *Scatter plot of the integrated F_z at 100 hPa (F_{100}) vs F_y at 55°N averaged during 6th-14th February 2018 for a) FREE, b) NUDGED and c) CONTROL experiments.*