REVIEWER 3

This study investigates the perfomance of 7 subseasonal-to-seaonal (S2S) forecast models in simulating 3 distint SSW events (2 in the NH and one in the SH). Moreover, it investigates the role of the stratosphere in triggering the Sudden Stratospheric Warmings (SSWs) in free-running simulations and two nudging setups where the stratosphere is either nudged to observations or climatology. The authors find that for some SSWs, the stratosphere plays a major role in modifying the stratospheric wave flux, but results seem to be very event-dependent. The paper is well written and results are presented clearly. The question posed is novel, since previous research has mainly focused on tropospheric drivers of SSWs. The paper is, however, very technical and I have some minor suggestions that migth improve readabilty. I recommend publication after the comments below have been addressed.

Thanks a lot for your comments. Here is our reply to the general and detailed comments in blue:

General comments:

1) Given that the main motiviation of the study is the improvement of predictability of SSWs, I think there is insufficient discussion on if and how the results presented help towards improving S2S forecasts, especially given that the mechanism/role of the stratopshere seems to be very event-dependent

We thank the Reviewer for this suggestion. We have expanded the discussion on the implications of our results for improving SSW predictability at the end of Section 6 (L740-759 of the revised manuscript), with particular emphasis on WN2 events, given the general low skill of models in forecasting this type of SSWs. In addition to the comments regarding WN2 events, we now also highlight that, although the mechanism and role of stratosphere are event—dependent, our results reveal that improved prediction of tropospheric circulation is key for increasing the forecast skill of all types of SSW. This is an important outcome of our analysis as it confirms the findings of Taguchi (2018). However, he could not isolate the potential influence of the stratospheric state on the upper troposphere and therefore, could not demonstrate that tropospheric precursors are even important for WN2 SSWs, where the perturbations are detected simultaneously throughout the whole atmospheric column (Esler and Matthewman, 2011). The SNAPSI experiments allow this isolation.

2) Especially section 5 is very technical and would benefit from a clear summary of the most important pints at the end of the section. It might even be shortened a bit to bring the main points across more clearly.

In the revised version, Section 5 has been rewritten and reduced. Following Reviewer's 1 recommendation, we have simplified the former Figures 12 and 13 (now combined into the new Figure 12) by removing the refractive index. Additionally, the description of this figure has been integrated with that of the previous Figure 11 to provide a more concise summary of the upward wave propagation in the FREE experiment.

Finally, when making some visual modifications in Figure 15 (now Figure 14), we have discovered a minor coding error that affected the Fz10 of WN2 component and consequently, the net EP flux associated with this wave component. Although the main conclusions remain unchanged, the corrected figure offers further evidence supporting the presence of wave

resonance in the stratosphere, thereby strengthening the connection between this subsection and the rest of the section.
Additionally, as suggested by this Reviewer, a very brief summary of the main points of this section has been added at the end of Section 5.

Detailed comments:

1) It is unclear to me how the boxes in Figs. 5-7 are derived. Please add a more detailed explanation of this. I think it would help to add the box also to the ERA5 panel.

The boxes indicate the tropospheric precursors of SSW or at least, regions associated with a weaker vortex state compared to the other states of the model ensemble. For each SSW, we define the areas characterized by substantial anomalies in the multi-ensemble mean of Z500 patterns (difference of Z500 between "weakest u" and "strongest u" ensemble members groups), as shown in Figs. 5-7 in shading. These areas are also located near the antinodes (ridge or trough) of the climatological WN1/WN2 waves in ERA5, suggesting that they could lead to constructive interference of waves if the locations of these antinodes in the models closely matched those in ERA5. Since this definition relies on the distribution of anomalies in the multi-model mean, we have added this multi-model mean to Figs. 5-7. The definition of the areas has been clarified in the text in L379-383 of the revised manuscript, and the corresponding boxes have also been added to the ERA5 panel.

2) Lines 375 ff.: I am not sure whether I understand how the Z500 anomalies are combined. In the text, it says "… *by computing the sum of averaged anomalies for centers with positive anomalies or positive-minus-negative…".* In the latter case, are you substracting the absolute mean value of the negative anomaly?

No. We are subtracting the negative anomaly, i.e. we are adding the absolute value of the negative anomaly, because we are computing the mean of the absolute values of the averaged anomalies for each center. We thank the Reviewer for this comment because we realized that the description of this calculation was incomplete in the original version of the manuscript and we have now clarified it in L381-383 of the revised manuscript.

3) Figure 8: Two models have almost same color (CESM2-CAM6 and NAVGEM). I suggest changing colors for better visibility.

We tested different colors but having so many models makes it difficult to select an additional one. Thus, for NAVGEM we have used a different marker in Figure 8 and dashed lines instead of solid lines in Figures 1-4.

4) Lines 446 ff.: In CNRM, not only an INCREASE in eddy heat flux at 100 hPa (Fig. 9a) is seen in NUDGED in the SSW 2018, but also a DECREASE in HF100 in this model in the SSW 2019 (Fig, 9b). Why? This should be discussed. The same tendencies can be seen in GLOBO and UKMO in the SSW 2019.

The Reviewer raised an important question. A short analysis has been performed to try to answer it. Based on linear wave theory, one might expect an increase in HF100 in NUDGED relative to the CONTROL run for the SSW 2019 as well. Figure 9 of the manuscript shows the total eddy heat flux at 100hPa (HF100) for all wavenumbers. However, when examining the heat flux for individual wavenumbers, it becomes clear that the decrease in HF100 in NUDGED for CNRM-CM61, GLOBO, UKMO-GloSea6 and also CESM2-CAM6 during SSW2019 is mainly due to a reduction in WN1 HF100 (Figure R3.1a). Indeed, for WN1, the negative difference in HF100 between NUDGED and CONTROL runs in these models increases with respect to HF100 for all wavenumbers. Moreover, more models such as KMA-GloSea6, GRIMs, NAVGEM or IFS also show a negative difference in HF100 between NUDGED and CONTROL runs, when just isolating the WN1 wave activity. In contrast, if we look at the WN3 wave activity, we find opposite results (Figure R3.1b), with overall higher values of WN3 HF100 in NUDGED than in the CONTROL run. The latter result can be understood by the very weak PNJ in the few days preceding the wind reversal in ERA5 and thus in the NUDGED runs (see Figure 1b of the manuscript). These weak westerlies are weak enough to allow the propagation of WN3 wave activity. However, in the CONTROL runs the PNJ is stronger, making WN3 wave propagation difficult.

The differences in HF100 between NUDGED and CONTROL runs are only detected in the stratosphere, but not in the upper troposphere (HF300), where both experiments show very similar values of HF100 for all wavenumbers, WN1 and WN3.

Based on these results, we hypothesize that when the WN3 wave propagation in the stratosphere is favored by the stratospheric state, the propagation of WN1 wave activity is reduced. This is what we observe in the time evolution of WN1 HF100 and WN3 HF100 in ERA5 for this SSW too (Fig. 3 of the manuscript). There might then be a competition between WN1 and WN3 wave propagation in the stratosphere. In fact, Smith et al. (1984) also documented examples during the winter 1978/1979 when the amplitude of stratospheric WN1 wave displayed lower values, but WN2 and WN3 waves were enhanced. These authors linked it to the effects of wave-wave interactions. Similarly, Shi et al. (2017) also described interactions between WN3 and WN1 during the SSW 2005. We think that this is a really interesting topic that we should analyze in more detail as a follow-up analysis. Nevertheless, we have included a short comment about this in the L464-471 of the revised manuscript.

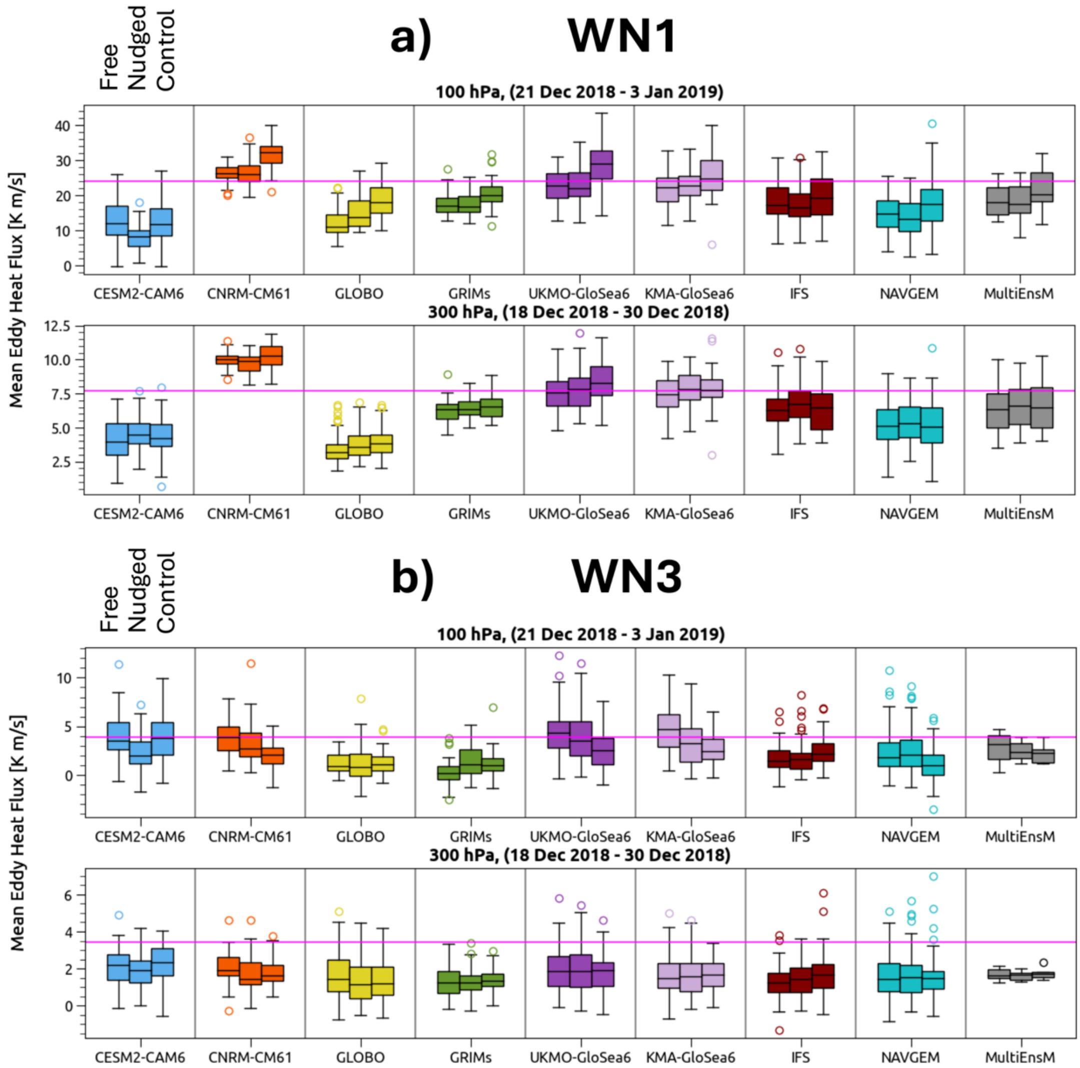

Figure R3. 1 *Boxplots showing the ensemble distributions of (a) WN1 and (b) WN3 eddy heat flux (HF) at 100 and 300hPa and 45°-75° N for the period with the strongest value of HF preceding the SSW2019 for all models and multimodel mean (MultiEnsM), and the different experiments (FREE, NUDGED, and CONTROL runs). The interquartile range (IQR) is represented by the size of the box and the horizontal black line corresponds to the median value. Whiskers extend from the box to a distance of 1.5 times the IQR. Outliers (colored circles) are defined as points with values greater than 1.5 times the IQR from the ends of the box. ERA5 values are represented by horizontal magenta lines.*

5) Figure 10: Although the multi-model mean shows no difference between nudged and control, in many models there seem to be significant changes, but they do not agree on the sign (especially in the SSW2019 SH). I think it would be worth investigating/discussing this in more detail, as 5 out of 7 models show clear differences between nudged and control in Z500 impact for the SSW2019 SH.

We thank the Reviewer for pointing this out. In the original version of the manuscript, the differences between NUDGED and CONTROL runs in Z500 impact for the SSW2019 SH had not been highlighted because they are only significant in three out of the five models. Please note that we define significant differences when the interquartile range of the two distributions do not overlap.

However, although not always significant, all models except for NAVGEM tend to show a higher amplitude of the tropospheric precursors in NUDGED than in CONTROL runs. The main contribution to this difference comes from the Amundsen Sea Low (ASL) that tends to weaken more in the NUDGED experiment, as revealed by the composite maps of Z500 anomalies prior to the SSW (Fig. S4). Since a weak vortex projects on a weakening of the ASL (Turner et al. 2012) and the vortex is already weak well before the SSW in the NUDGED run (Fig.1c), we hypothesize that the troposphere is already affected by the weak polar vortex influence. This is also consistent with Jucker and Reichler (2023), who characterized the life cycle of SSWs in the SH. They found that the polar vortex starts to decelerate more than 50 days before the SSW date, and, consistently, a very strong weakening of the ASL and a generally negative phase of the Southern Annular Mode appears in the troposphere one month before the occurrence of the SSW.

Considering the agreement across models, we have briefly discussed these results in the revised version of the manuscript in L507-515 of the revised manuscript.

6) Lines 538 ff: I suggest changing the y-label in Fig. 12 to hPa instead of Pa to be consistent with the text. I also suggest marking the areas discussed in the text in Fig. 12 (e.g. 10-3 hPS). Otherwise the discussion surrounding this figure is hard to follow.

Figures 12 and 13 have been modified in the revised version of the manuscript. In particular, the refractive index has been removed and both figures have been merged and extended up to 10ºS. The corresponding description has been modified accordingly. The y-label of the new Figure 12 has been changed to hPa.

7) Lines 549 ff.: Again, it is difficult to follow the discussion here. Which “shift towards the pole”? Again, marking the corresponding areas in Fig. 12 would improve readability of this part.

As mentioned in the reply to the Reviewer’s second major comment, Section 5 has been reduced and the description of the refractive index has been eliminated to improve the readability of this section.

8) Line 601: what is meant by “misrepresentation of the zonally symmetric stratospheric state in models”?

We refer to the differences in the representation of the zonally symmetric stratospheric state by the models with respect to ERA5 already shown in Figs. 11 and 12 and described in the corresponding text. We think that these differences may have contributed to the lack of a strong WN2 burst in the stratosphere since the F_z shows larger values in the NUDGED runs, where these discrepancies have been eliminated, than in the FREE ones.

The sentences have been clarified in the revised version of the manuscript.

<u>References:</u>

Esler, J., & Matthewman, N. J.: Stratospheric sudden warmings as self-tuning Resonances. Part II: Vortex displacement events. J. Atmos. Sci., 68, 2505–2523, 2011.

Jucker, M. and Reichler, T.: Life cycle of major sudden stratospheric warmings in the Southern Hemisphere from a multimillennial GCM simulation. J. Clim., 36, 643-661, 2023.

Shi, C., Xu, T., Guo, D. and Pan, Z.: Modulating effects of planetary wave 3 on a stratospheric sudden warming event in 2005. J. Atmos. Sci., 74, 1549–1559, https://doi.org/10.1175/JAS-D-16-0065.1, 2017

Smith, A.K., Gille, J.H., and Lyjak, L.V.: Wave-wave interactions in the stratosphere: Observations during quiet and active wintertime periods. J. Atmos. Sci. 41, 363-373, 1984.

Taguchi, M: Comparison of subseasonal-to-seasonal model forecasts for major stratospheric sudden warmings, J. Geophys. Res.: Atmos., 123, 10231-20247, 2018

Turner, J., Phillips, T., Hosking, S., Marshall, G.J., and Orr, M.: The Amundsen Sea Low. Int. J. Clim., 33, 1818-1829, 2012.