We are thankful to the reviewer for these comments which help us improve the quality of our manuscript. Our response to each comment is written below. We note that we have noticed a small coding error: averages were made between 50-85N instead of 45-85N as indicated in the manuscript. After correcting this mistake, there is a small change in the classification of High-Agreement and Low-Agreement events (please see the correction to the table where SSW events are listed) but this does not alter our interpretation of the results.

General comments

This is an interesting study comparing several reanalysis datasets to analyze the momentum equation during sudden stratospheric warmings (SSW). In particular, I find useful the idea of analyzing the latest reanalysis ensemble separately. Having so many available datasets, an evaluation of how they perform depending on the topic of the study is required. I have enjoyed reading the manuscript and I appreciate the advises and recommendations regarding the uses of different reanalysis in stratospheric dynamics. The contain of this manuscript would be valuable for the scientific community so I recommend it for publication. I only have some minor questions listed next.

Specific comments

Page 6, line 10: What happens if a SSW is detected in less than 4 datasets? Is it excluded? Does this

happen in many cases?

If the event is not detected in at least 4 datasets it is excluded, which we now indicate in the manuscript. It happens in February 81, February 95 and February 2002. This means only three events out of a total of 25.

Table 2: Why the analysis stops in 2010? There is no data for some reanalysis considered herein?

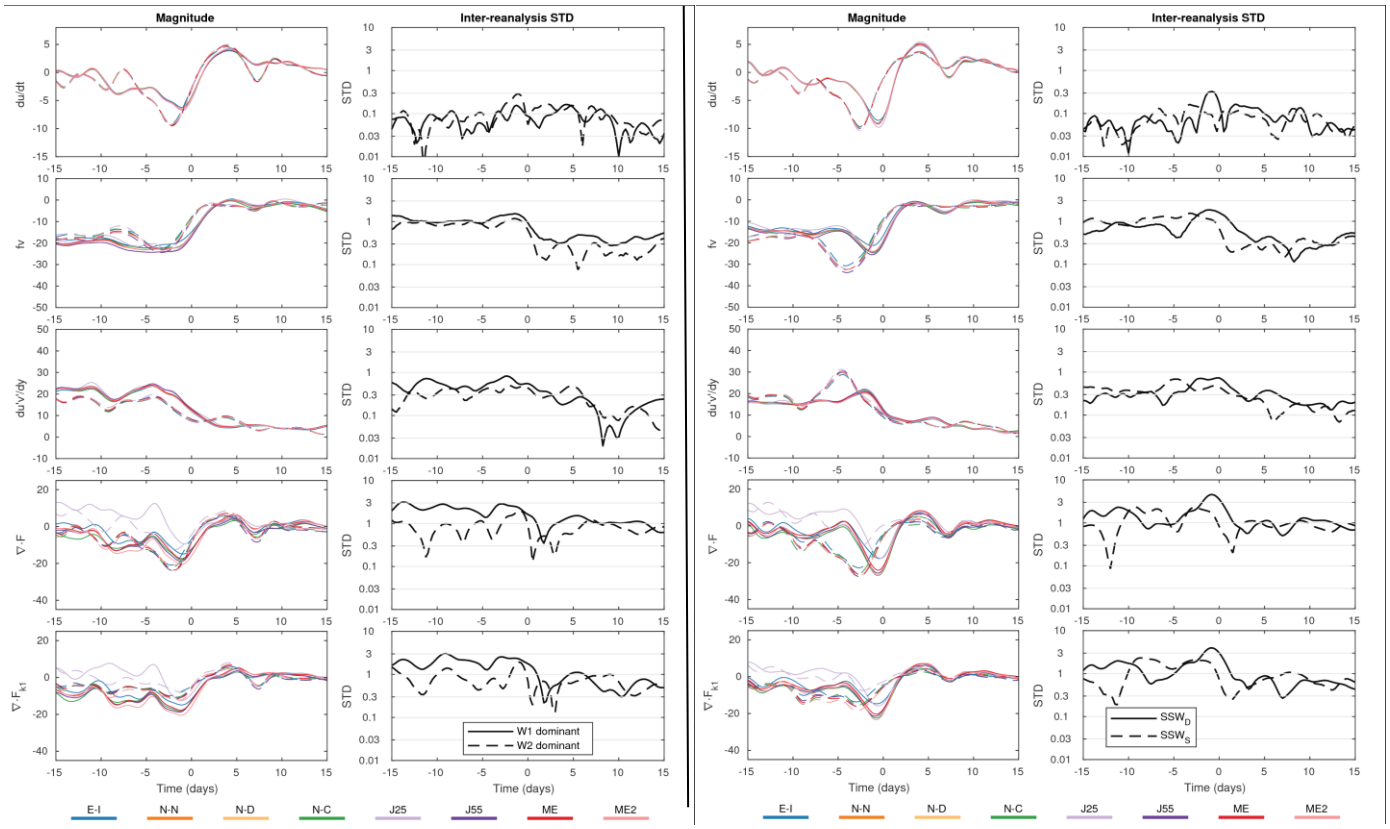
The last SSW listed is in 2010 because we use the official comparison period of the S-RIP project which ranges from 1980 to 2012 (mentioned in the data section). No events are detected in 2011 or 2012. Only one more event would be included if the analysis is extended to 2013 after which JRA-25 is not provided.

Page 7, line 5 and Page 24, last paragraph: The authors argue that split and displacement SSW-types result from different planetary-scale wave forcing which motivates that distinction. I wonder whether it would be more meaningful to distinguish between wavenumber 1 and 2 as SSW precursors instead of the split/displacement sub-classification. As shown in previous studies (e.g., Bancala et al. JGR-2012), he ratio W2/W1 events will be much smaller than the split/displacement type, but it could be worthy to check whether differences are significant in that case.

Thank you very much for the comment. It is true that the geometry of the vortex (split or displacement) does not necessarily reflect the wave drag that produced the SSW event. To investigate the role of the longitudinal scale of wave activity fluxes to the spread among reanalysis, we classify events according to whether they are dominated by wavenumber-1 (W1) EP flux or wavenumber-2 (W2) EP flux. We indicate the outcome of this classification in Table 2 of the revised manuscript.

We have added the following paragraph in the manuscript: *On the other hand, when considering the dominant fluxes of wave activity producing SSW evens, we find that out of 7 LASSWs, four are W1-dominant and 1 is W2 dominant and out of 7 HASSWs, 1 is W1-dominant and 3 are W2-dominant. This*

seems to indicate that wavenumber-1 wave drag is responsible for larger uncertainties in reanalysis datasets but a detailed analysis reveals that inter-reanalysis spread is not markedly different between W1-dominant and W2-dominant events (not shown) suggesting that it is the intensity of wave drag rather than the longitudinal scale of wave activity that is linked to uncertainties among reanalyses. Supplementary Fig. 1 shown here supports this analysis.



Supplementary Figure 1: Similar to Fig. 11 of the manuscript but for comparing SSWs that are primarily forced by wave-1 or wave-2 EP flux (solid and dashed, respectively; shown on left side) and SSWs that are displacements or splits (solid and dashed, respectively; shown on right side). Although there are small differences in the deceleration of zonal-mean zonal wind, the weakest stratospheric winds following the SSWs are similar in all types of events (not shown).

Page 12, line 8: What do you mean with "our interpretation of the evolution of SSW events"?

What we mean is that the uncertainties in the upper-stratosphere have a greater impact on how we understand the evolution of SSW events from observations. We rephrase this section to improve the clarity.

Technical corrections

Figure 2: It is difficult to distinguish the thin and thick lines. Maybe using dashed and solid lines as in figure 5?

We are now using dashed lines.

Figures 3 and 4: I have found difficult to distinguish contours with this color scale. Especially in a printed version.

We use a new color scale that makes interpreting these figures easier.

Figure 6 (right column): Note that some plots are out of range.

We extend the range of this figure.

Page 8, line 9: events or event?

Corrected

Page 10, line 25: quantities.,

Corrected

Page 15, 4: Here, and in other parts of the text: m/s/day. This notation is confusing, I would suggest using ms-1day-1

We now use $ms^{-1}day^{-1}$ throughout the text.