Response to Reviewers

A High-Level Changes

As discussed in our previous Comment, we have made several structural and methodological changes since the original version of the study.

A.1 New u calculation method

Previously we used a directionally-projected estimate of u derived from individual profiles. Following separate discussions with reviewer Isabell Krisch, our revised manuscript uses a new method to estimate the true value of u, which we describe in a new Appendix A. We feel that this is a major contribution to the study that provides significant additional confidence in the results and represents a significant methodological advance, and have accordingly offered her co-authorship on the article to ensure she receives intellectual credit for this new method.

This new method pairs profiles from different overpasses, which cancels several directional terms and hence allows us to constrain the direction of u. This provides a more accurate estimate of the true value of u, but with reduced temporal precision from an instantaneous value to a several-hour mean. Since in practice all our plots average over at least one day and often several, this is an acceptable tradeoff for the increased measurement accuracy.

A new and separate manuscript ("Manuscript 1", to use the term from our previous online comment) focusing on the details of this method has been prepared, and will be submitted to Atmospheric Measurement Techniques in the very near future - a complete manuscript has been prepared and is undergoing final checks between co-authors. This has been led by Dr Krisch, and focuses on both comparisons between different methods and validation of them against reanalysis data.

In general our results are morphologically very close to those in the original manuscript but with larger absolute values, consistent with the change from a projected to a true estimate.

A.2 Removal of Content

As suggested by the reviewers, the previous Section 6 on Mesoscale Forcing has been removed, together with discussion of and cross-references to this section in other parts of the manuscript. We have also removed the two appendices (one on methodological testing, one on comparisons to ERA5). Sections 7.1 and 7.2 have also been merged to significantly reduce duplication.

A.3 Use of MLS GPH

ERA5 GPH has been replaced with MLS GPH in the figure showing GPH patterns on specific days of the 2021 event (Figure 9 in both the original and new numbering), and the text updated to reflect this. In practice the morphological differences due to this are small, and hence so are the text edits. The figure showing GPH anomalies relative to a long-time average, Figure 10a in both numberings, retains the use of ERA5 GPH, as the much shorter period of MLS GPH would very significantly change the meaning of the analysis and introduce inconsistencies with the surface climate variables.

B Specific Responses - Manney

B.1 General Comments

As discussed in our previous comment, we broadly concur that the manuscript should be split, and accordingly have done so. The revised manuscript now represents ‘Manuscript 3’ to use the parlance of that comment, i.e. an observationally-focused discussion of the (early) 2021 SSW.

B.2 Specific Comments

1. ***Line 22: Don’t think “wider” is the best word here, perhaps “broader”, or “midlatitudes”. More importantly, the polar jet does not separate the polar vortex from the rest of the atmosphere, it defines the polar vortex.***

Sentence changed to “separates the cold polar stratosphere from midlatitudes”, addressing both issues.

2. ***Line 24: It would be good to cite Manney et al (2015a; doi:10.5194/acp-15-5381-2015) and Manney et al (2015b; doi:10.1002/2015GL065864; cited later in this MS) here, since both of these papers specifically focus on the differing effects SSWs can have on polar chemical processing and ozone loss (Manney & Lawrence, 2016; doi:10.5194/acp-16-15371-2016, is also relevant).***

The first two of these references have been added.

3. ***Lines 26–27: I would say “...SSWs can trigger extreme... Also, work such as that of Lee et al. (2019; https://doi.org/10.1029/2019GL085592), and references therein, suggests that SSWs may not be as directly related to North American surface weather extremes.***

Text modified as suggested.

4. ***Line 31: I would not say “usually” in light of, e.g., the discussion in Butler et al. (2015; doi: 10.1175/BAMS-D-13-00173.1) and the utility of other definitions for some purposes.***

Replaced with “often”.

5. ***Lines 42–43: I think it would be worth a brief mention of the UARS WINDII and HRDI instruments and their limitations (particularly altitude ranges) here for context.***

As the reviewer comment alludes, HRDI and WINDII were designed to measure a very different part of the atmosphere - therefore, a discussion of these instruments and their limitations is probably not fully germane to the discussion here.

6. ***Lines 48–52: As mentioned in the general comments, I don’t see that MLS data are really needed here, and ERA5 data are doing far more than “support”/***

This sentence has been modified to reflect the structural changes made described above by removing the word support and slightly rephrasing the prose around it, but as discussed in the previous online comment all three datasets remain.

7. ***Lines 55–59: Unfortunately, this “parallel” approach ends up being very unclear, as well as repetitive and incomplete at the same time. It is also very cumbersome for the reader as it entails flipping back and forth many times between widely-separated figures.***

This paragraph has now been removed, and structural changes made that change this.

8. ***Line 80 / footnote 2: Might these conditions occur during an extreme event such as***

an SSW where vertical winds are highly perturbed (e.g., Manney et al, 1994 and many others)?

This is arguably true, but as it is fundamental to the data as supplied by ESA we would argue that it is significantly beyond the scope of this study to properly disentangle this effect. Providing this important caveat up-front in this way seems to us to reflect a good balance - it gives the reader the information they require to make up their own mind, while keeping the work presented here (comparatively) focused on the science questions. To clarify that this is an upstream choice, we have changed the phrasing to make clearer that this assumption is made in the retrieval rather than by us.

9. ***Lines 105–110: Schwartz et al (2008; doi:10.1029/2007JD008783) is a better reference for MLS GPH and T. How do you do the gridding of MLS data onto the 5x20 degree grid? Also, please explain why it is desirable to (line 109) "provide a spatial weighting roughly equivalent to that for MLS temperature".***

The MLS data are gridded via 2D binning onto a regular grid indexed at the lower-left corner. The spatial weighting being roughly equivalent is desirable as it makes the two datasets more fairly comparable at a bulk level.

10. ***Lines 114–117: Why do you not calculate/use meridional winds from MLS? Not only zonal (in many papers, some of which should probably be cited here, e.g., McDonald et al, 2011, doi:10.1029/2010JD014719; Smith et al, 2017, DOI: 10.1175/JAS-D-17-0067.1; Harvey et al, 2018, https://doi.org/10.1029/2018JD028815), but also meridional winds have previously been calculated from MLS, and though meridional winds are more uncertain than zonal winds, they have been shown to be reasonable in the middle stratosphere and below (e.g., presented in posters by Millán et al at Fall AGU 2018, AMS MAC in 2019; also discussed / shown indirectly in calculated PV in Manney et al, 2008; doi:10.1029/2007JD00909, 2008). Given the uncertainties in meridional winds derived from Aeolus, I would suggest that meridional winds derived from MLS are no worse, if not better.***

All meridional winds have now been removed from the article.

11. ***Lines 149–151: I like this approach!***

Thank you for this - I (Wright) was a bit nervous structuring the paper this way, but both reviewers approving so strongly makes me more confident about doing so in future. The data have now been replaced with ERA5, with no changes to the data, and the section accordingly removed.

12. ***Figure 1 caption: is this a running or stationary mean?***

Modified from 'smoothed' to 'stationary-boxcar-smoothed' (the mean taken is neither stationary nor rolling as the data are shown as daily values only and hence the two are functionally equivalent, but the question is a meaningful one for the smoothing).

13. ***Lines 163: there are numerous other papers that could be cited here, including several that are particularly relevant since they use winds derived from MLS data (e.g., Smith et al, 2017, 2019; Harvey et al 2018; 2019).***

Smith et al 2017 and the two Harvey papers have been added. I'm afraid I (Wright) have not been able to locate the Smith et al 2019 reference mentioned by the reviewer as it is not forward-linked from the others and the very common surname (a problem I have myself!) makes the specific paper hard to find on online databases without further identifying information. Given the reviewer has been excellent about providing DOIs for all cited studies, this is clearly just a minor oversight and I am very happy to add it in a later revision (subject to reading it to confirm, as I do for all added references!)

14. ***Lines 168–170: I would be very cautious about terms like "underestimate" and "missing", since I see no evidence on which to base an assessment on which is more correct.***

Both of the features these comments described are no longer present in the data, and thus the comments have been removed. The 'underestimate' was due to differences in spatial sampling - the ERA5 values were implicitly averaging over a larger area as Aeolus data at 22 km are very sparse equatorward of 60°, and the changes to the data analysis method described above removed the brief reversal seen in Aeolus. The text has been replaced to describe some new differences seen now that the comparison is fairer, and greater care has been taken to avoid implying one assessment is better than the other.

15. ***Line 180: Should note that, for MLS, "poleward of 60N" means from 60 to 82N because of the sampling pattern.***

Bracketed comment "(but note that MLS data only extend to 82°N)" added to clarify this.

16. ***Line 181: In fact, zonal mean winds are rarely closely related to "vortex-edge winds" in the Arctic, where the vortex is never really symmetrical or pole-centered, and especially not during / around SSWs. (See, e.g., Lawrence & Manney, 2018, https://doi.org/10.1002/2017JD027556, for discussion of vortex-edge winds and their evolution during SSWs.)***

The sentence has been modified to "to focus on the strength of winds in regions closer to the nominal edge of the polar vortex under undisturbed conditions.".

17. ***Lines 183–185: Although it doesn't hurt to have this information here, it is most important to have it in the figure caption.***

Additional text now duplicated in caption - the information provided by the additional text was already present in the figure as solid lines on the colourbar (hence not adding it previously), but this avoids any ambiguity.

18. ***Lines 182-190, and Figure 2: There were, during these years, also 4 late January / early February SSWs, 2 late February ones, and one early-mid-March (so in the day range shown in Figure 2) major final warming. Some of these must contribute to some of the extremes in the climatology, and it isn't clear how that related to your statement about a "typical" SSW here (in fact, the use of "typical" in relation to SSWs is problematic, since each one has very different characteristics, including having a wide range of influence on the troposphere (eg, Butler et al, 2020, DOI: 10.1002/qj.3858, discuss that in the context of two of these events). Since one of the two previous early January SSWs in this period was in 2019, it would be best to cite something more recent than Butler et al (2017) for that (perhaps Butler et al, 2020, mentioned just above). (Also one set of Figure 2 x-axis labels must be wrong, since the top says "from 5 January" and the bottom "from 1 January" but both tick marks line up; and adding a horizontal zero wind line to the right panels of Figure 2 would be very helpful.) Per general comments, it would be more effective to simply show ERA5 for both winds and temperatures (which would also allow you to show temperatures at the lowest levels included).***

[Climatological context] Agreed: the other SSWs will contribute to the climatological range. The text referring to a 'typical' SSW being above/below the 82nd/18th percentile but not necessarily exceeding the full range was intended to make this point, but was clearly insufficiently clear. To correct for this, text has been added to make this intended point explicit rather than implicit within the text, and this rephrasing has also removed the word 'typical'.

[citation] This is a bit of a tricky judgement call. The actual data are derived from Dr Butler's webpage, which is up to date. However, the webpage explicitly asks that users cite the 2017 paper, as the method is explained and derived there and the webpage is "grey literature" rather than part of the public (and, more importantly, formally archived) scientific record. So, for now we have left this reference as-is, but are happy to reconsider this choice if the editor or reviewers disagree on this course of action.

[axis labels] This was a labelling error - earlier versions of the Figure were based on 01/01 and this was modified to 05/01 at a relatively late date. The upper label has now been corrected to 05/01.

[Change to ERA5] As previously discussed in the online discussion, this change has not been made.

19. ***Lines 193–196: first, I believe "above" at the end of line 193 should be "below"? Further, per comment above, strong zonal mean wind tells us virtually nothing about the strength of the polar vortex, since it has very little to do with vortex-edge wind speeds or potential vorticity gradients, and largely reflects the degree of symmetry and displacement from the pole of the vortex. In addition, a cold vortex is by no means synonymous with a strong vortex (a classic example is in the 2004/2005 winter, where the vortex was unusually cold, leading to more chemical ozone loss than usual, but was also unusually weak / permeable, leading to stronger mixing than usual, e.g., Manney et al, 2006, doi:10.1029/2005GL024494; Schoeberl et al, 2006, doi:10.1029/2006JD007134).***

For the first point, yes - this has now been corrected. For the second point, the text has now been removed.

20. ***Line 201: the "rise in the climatological mean" at this time is partly (perhaps primarily) due to the later SSWs I mentioned above.***

Text has been added clarifying this.

21. ***Lines 208–210: In line with comments above, I don't think there is much, if any, information gained by saying an SSW is "broadly typical", especially since you have not even defined what you are calling typical.***

Sentence removed.

22. ***Section 4 Overall: This section would be much more effective and focused if you simply used the ERA5 dataset to describe the zonal mean evolution during the SSW. You could provide a complete picture from the surface into the mesosphere. In the troposphere through the mid-stratosphere in the extra-tropics, ERA5 winds are very well constrained by data. The only place where an argument could be made that winds derived from one of the satellite datasets might be of as good or better quality is MLS-derived winds in the upper stratosphere and mesosphere, where ERA5 is not as well-constrained by data.***

As discussed in the previous online comment, ERA5 has not been switched in to replace MLS here - other changes have been made to this section though based on additional comments below.

23. ***Figure 3: Should explain how stratopause and tropopause heights are obtained sometime before this figure. Also, it would be very helpful in the discussion of this figure if you (in addition to showing ERA5 entirely) showed both fields and anomalies for both temperature and wind speed. As it is we have no idea of what the actual temperatures are that these anomalies are from and no idea how anomalous the winds shown are.***

Stratopause and tropopause height calculation: papers which describe the methods used to do this are already referenced in the second paragraph of the Section (line 224 of the as-submitted version), with clear text stating this, viz. "derived from ERA5 as described by [**Wright2020**] and [**France2012**] respectively."

Anomalies vs absolute fields: agreed, these have been added as a supplementary figure, and referenced in the text where the anomalies are introduced.

24. ***Lines 231–232: Figure 3 does not show "vortex-edge winds" and gives us no information on whether those winds were "strong and zonal throughout the stratospheric and mesosphere" – while they can be presumed to be strong if the zonal mean winds are strong, the opposite is certainly not the case.***

The sentence has been rephrased to" The zonal mean zonal wind at 60–65°N during this period is generally large and positive throughout the stratosphere and mesosphere, ...".

25. ***Lines 240–242: Foreshadowing something that isn't shown until two sections later and giving possible interpretations of it is not effective presentation. This is one of many areas where reorganizing and focusing on a complete picture from ERA5 would be very beneficial.***

Sentence removed; see also restructuring comments above.

26. ***Line 250: According to the ticks on Figure 3, the zero wind line has not descended to 10hPa by 5 January.***

The change to a 60-65° geographic average has changed this feature of the figure, and now the datasets agree. Note that, although the comment is correct, the dashed line does indicate the top of the next figure rather than the 32 km level, and this may be a source of confusion that made a small discrepancy here look larger than it was in the original version.

27. ***Line 252: MLS temperature / GPH data are, however, considered to be scientifically useful at these levels according to the MLS team's quality documents, and have been used to calculate zonal mean winds in numerous previous papers that have demonstrated this (e.g., McDonald et al, Harvey et al, Smith et al references mentioned above; Manney et al, 2008).***

Agreed; comment rephrased.

28. ***Lines 269–270: It would be good to include France et al (2013, doi:10.1002/jgrd.50218) in this list (though they do, demonstrably falsely, claim to be the first to show the zonally asymmetric nature of elevated stratopauses).***

Added.

29. ***Lines 277–280: All showing the Aeolus data does here is unnecessarily interrupt the flow of the description of the evolution of the circulation, when all that discussion could have been done in a systematic and complete way by simply showing the full vertical range of ERA5 zonal mean winds and temperature in Figure 3, as suggested above. Further, to demonstrate the utility of Aeolus zonal (and meridional) wind data, much more detailed comparison with other sources of wind information is needed, including more detailed comparison with ERA5 – and this should be in the body of that paper, as it would be critical to interpretation of those results. That cannot be accomplished in parallel with giving a lucid and "simply-connected" description of the dynamics of the SSW.***

As discussed in the previous online comment, the decision to use Aeolus rather than ERA5 data for this paper is an active choice (and indeed the core focus of the paper), as is the decision not to focus the paper on ERA5 comparison/validation.

30. ***Lines 285–287: More precisely, it occurs earlier at lower altitudes. This does not***

tell us whether or not anything "propagates upwards".

Rephrased to avoid a statement of causality.

31. ***Lines 287–288: Same as previous comment but re "propagating downwards".***

Rephrased to avoid a statement of causality.

32. ***Line 289, How do you determine what is "typical"? Where "in the UTLS"? If you meant near the altitude of the upper tropospheric jets I'd expect it to generally be higher than that. Also would expect that if you are talking about the stratospheric subvortex region.***

"Typical" was intended to be relative to the earlier record within the same year, but has been removed to avoid confusion. "In the UTLS" has been clarified to make clear that we are referring specificaly to heights near the tropopause, with the text around rewritten to clarify vertical extent and morphology.

33. ***Lines 294–295: "...five days after it does so at 10hPa. . . " – after it reverses in MLS derived winds? In ERA5 winds? You've already said / shown there are differences in timing of several days in some cases, so simply focusing all this discussion around one dataset (and ERA5 is the only one that is complete enough) would help a lot with understanding relationships of timing such as this. Also, you say the zero wind lines reaches a minimum altitude on 15 Jan, but it looks more like about 22 Jan to me in Figure 4.***

The revision to the paper to cover the region 60°-65°N in wind data rather than 55°-60°N has in general removed the 5th January time mismatch. The text has been rewritten around the duration of the lower height limit, but in any case this may be a difference in visual interpretation of the original figure: the original data reached a near-minimal value around the 15th of January and then remained broadly flat until the 24th or so, but a true minimum was reached around the 22nd, so it depends on whether the reader takes the start of the flat period as the minimum and assumes any movement in this flattish feature is below measurement accuracy (i.e. of limited meaning), or whether the formal minimum is taken. Either way, the data and text have changed now anyway so the point is moot!

34. ***Lines 295–306: There are many other places in this discussion where the dates in the text do not appear to match well with what the figures show. This should be checked carefully, but it would also help the reader a lot if, in all of the time series figures, you had more obvious labeling of the dates (eg, more frequent labels/major ticks, more frequent minor ticks, and/or more vertical lines at ticks).***

Agreed - minor ticks have been added, and the dates are now provided at double the original cadence. Some of the original dates in the text were indeed slight misreads due to the low precision of the original axes and/or changed by the modification to the data analysis pipeline and geographic averaging region, and have been adjusted accordingly.

35. ***Line 304: It is not clear what "This minimum" refers to here.***

Revisions to the areal coverage of the data (primarily in Figure 1 above, rather than here) have led to a change in the data morphology at this time index, and as a result this paragraph has been removed.

36. ***Lines 308–309: It looks to me from Figure 2 as if the winds recover to being stronger than average, which is typical of early major SSWs in the middle to upper stratosphere, and in the lower stratosphere in those that are early enough for the longer recovery times scales at those altitudes to have an effect before the spring final warming (e.g., Manney et al, 2008; Hitchcock et al, 2013; Hitchcock & Shepherd, 2013; Manney et al, 2015b; numerous others).***

Thank you - this comment has been added to the text for additional context.

37. ***Line 314: Need to give references for different effects on surface of splits and displacements. There is a lot of work on this subject and from what I've read (by no means all of it) there is not a clear consensus as to how different those effects are and under what other circumstances (eg, differences depending on when in the season the SSWs, the geographic location of the surface impacts, etc).***

Three references have been added.

38. ***Figure 5 and Section 5: Splits and displacements are in fact just as (or more? Haven't counted #s of papers) frequently determined using potential vorticity on isentropic surfaces (e.g., Matthewman et al 2009; Lawrence & Manney, 2018, and references therein), often but not always in the middle stratosphere (Lawrence & Manney, 2018, looked at them throughout the stratosphere).Further, there are several methods used to determine them, with the methods of both Matthewman et al (2009) and Seviour et al (2013) being indirect – that is, they do not directly measure whether the vortex splits or not. Lawrence & Manney (2018) compare these methods with direct identification using closed contours of splitting, and show some serious limitations of these methods.Further, there are numerous previous cases where an SSW is "hybrid" in some sense – either splitting at a limited range of levels (sometimes, as in early Jan 2019 and March 2016 at some levels, into more than two pieces) or having several "pulses" (similar to the 2021 SSW, but also the 2010 SSW), with whether it is classified as a split or displacement depending on which of the pulses is identified as "the" major SSW. There are many cases where different reanalysis datasets may "see" different types of events. Since any method of identifying these depends on one or more "thresholds" (in the case of direct methods, the threshold is the exact GPH or PV value chosen to represent the vortex edge; for methods like Seviour's, empirically – and subjectively – chosen values of centroid latitude and aspect ratio), they are all very sensitive to small differences. These and other nuances of the split / displacement classification are not discussed here and are needed to put this section in context, and to clarify that the idea of "mixed" or hybrid or evolving types of SSWs is not unique to the 2021 event.***

[Use of GPH vs PV] We agree with the point that split/displacement events are often, and perhaps preferably, defined using PV rather than GPH. We have, however, chosen to use GPH here in order that more direct comparison may be made with satellite measurements (in particular, MLS), as in Fig 9, in keeping with the focus of the paper.

[Vertical level of classification] We have altered the text to reflect the point that the classification at 10 hPa may be different at other levels, and added the relevant reference to the Lawrence and Manney (2018) paper.

[Threshold classification] We agree with the inherent limitations pointed out regarding use of thresholds, and that more advanced techniques such as that of Lawrence and Manney (2018) may be preferable. Our aim in Section 5 was to describe the split/displacement classification of the 2021 event according to published classification methods available to us (those of Seviour et al. 2013 and Gerber et al 2021), rather than to delve into the relative merits of these methods. We have, however, added a sentence to highlight the limitations of these methods.

39. [Several comments were raised here about Section 6, which has now been removed as discussed above. Accordingly, we do not include specific responses to these comments.]

40. ***Section 7 Overall: Because of the way the sub-sections are arranged, and the Aeolus fields shown/discussed, this section is not only very repetitive (e.g., the seven***

stages are walked through twice, once in relation to Figures 7 & 8 and once in relation to Figure 9), but does not convey a clear, coherent picture of the synoptic evolution of the vortex during the period surrounding the SSW. Showing zonal winds is a much less effective way to describe the synoptic evolution of the vortex than showing wind speed magnitudes (on either isobaric or isentropic surfaces), GPH on isobaric surfaces, or potential vorticity on isentropic surfaces. Especially, showing 3D plots of zonal winds as opposed to potential vorticity (scaled to have a similar range at each level) or wind speed maxima (or PV gradient maxima, or any of a number of other quantities that have been used that would "outline" the shape of the polar vortex – you simply cannot see things like (line 468) "vortex begins to break down" from zonal winds alone, they do not tell you the shape and position of the vortex without knowing the meridional winds as well.In addition, the discussion includes specific statements such as "helical structure begins to develop in wind..." and "wind speeds have reached values as low as. . . ", when referring to zonal winds only – in the former case it is unclear what this implies for the full wind field, in the latter it is simply incorrect. Further, to even effectively show the information content in Figure 8, the isosurfaces / view would need to be adjusted (including transparency) so that most of the isosurfaces are not hidden most of the time, and a more appropriate (higher) zonal wind value would need to be chosen for the positive isosurface so that it actually represents something near the vortex edge rather showing nearly the whole hemisphere poleward of 60N (which is far too high a cutoff, since parts of the vortex commonly extend or are displaced farther south than that during SSWs). Again, these sections would much more effectively convey the synoptic evolution of the vortex during the SSW if you simply used the ERA5 dataset and showed something that more directly represents vortex evolution (such as a version of Figure 9 with ERA5 wind vectors at both levels).

[Text arrangement] 7.1 and 7.2 have been merged, and the text rearranged. This has significantly reduced the degree of duplication.

[Variable choice] As discussed in the previous online comment, the use of observed winds rather than simulated PV is an active choice, and has not been changed.

[Description] The offending phrases have been excised, and "wind" changed to "zonal wind" throughout (it's possible one or two instances of the latter have been missed in the heavy rearrangement - happy to correct if found).

[Figure design] The following changes have been made to these panels: (i) the outer isosurface has been moved to 15 m/s; (ii) a second inner isosurface has been added at 25 m/s; (iii) the outer isosurface has had its opacity dropped significantly; (iiii) the interior grey cylinder has alse been reduced in opacity; (v) the viewing angle has been dropped down. Combined, these changes are intended to address most of the specific issues the reviewer identifies with this figure. The data cannot easily be extended equatorward of 60°N as there is a substantial drop in maximum altitude of Aeolus data here, introducing a visually-confusing step in data height here if the data are included. The data have not been changed to ERA5.

41. ***Figure 7: Panels are too small, wind vectors are very hard to see; if you are going to show the wind vectors (that is, use the meridional wind values you have guesstimated), it would be helpful to simply show the wind speed magnitude instead of the zonal wind – that would show directly how these winds derived from Aeolus data represent the polar vortex structure.***

The wind vectors have been removed. As the $\overline{u}$ structures remaining are at much larger scales physically than the vectors were, this should also address the panel-size problem, but we are happy to address this by (e.g.) supplying a landscape version of the figure that can be shown on a rotated page if this would help.

42. ***Lines 464–465, "typical for winter Aeolus data": you don't really have enough yet to define what is typical.***

The sentence has been rephrased to "is similar to Aeolus data from earlier in the winter" to make clear that the comment refers to earlier in the same winter rather than to (e.g.) a climatological mean over many winters.

43. ***Lines 491: You can see that the vortex becomes more symmetric, that does not necessarily mean it is "spinning up".***

Replaced with "return to a more symmetrical form".

44. ***Lines 496–499: You simply cannot tell whether the "...flow...is again strong and circular around the pole. . . " from the zonal winds; in fact, the wind vectors suggest to me that it is elliptical and weak, and Figure 5 indicates an aspect ratio around 1.5, which I would hesitate to call "circular". (BTW, "normal" for the Arctic vortex is never "circular around the pole" and rarely centered near the pole.)***

Both fixed.

45. ***Lines 508–510: This doesn't make sense to me. So far as I know, there is no "mid-latitude" jet at 70hPa equatorward of the polar night jet that defines the polar vortex. The tops of the upper tropospheric jets just do not extend this high to any significant degree (see any of numerous papers on climatology of the UT jets, eg, for lack of time to look something else up, Manney et al, 2014, https://doi.org/10.1175/JCLI-D-13-00243.1).***

Sentence removed.

46. ***Lines 510–511: This is very subjective, and to my eye there are numerous places/times where they don't line up very well, e.g., the large vectors at the lowest latitudes shown on 12 Nov, 5 Jan, 20 Jan, 29 Jan, 19 Feb do not appear to be associated with strong GPH gradients; the vectors at high latitudes near 60–90E on 20 Jan do not line up with the contours, nor do those on 29 Jan near 60E crossing the red highlighted contour). For a comparison/verification of Aeolus zonal and meridional winds, a direct comparison of u and v wind components would be much more useful.***

The wind vectors have now been removed.

47. ***Line 515: I don't think I'd call an aspect ratio of 1.5 "roughly circular".***

Adjusted during the text rearrangement.

48. ***Lines 520–521: In fact the vortex in the lower stratosphere in November is not even close to fully developed, thus is very weak and highly variable.***

"consistent with expected atmospheric dynamics at this height" removed.

49. ***Line 522: I can't see a "a tight detour towards the pole" in the GPH contours.***

This is subjective, but in any case is flattened out by the geographically-coarser MLS GPH product, so the comment has been removed.

50. ***Lines 532–536: It doesn't really make sense to me to talk about a "vortex split" if the contour that splits is well inside the vortex edge.***

This section has changed with the changed data, so no direct response, but the comment has been taken account of in the new text.

51. ***Lines 539–540: But if you just used the ERA5 data for everything, you could show this!***

See“High-Level Changes”, above.

52. ***Lines 553–554: I would call it more than an “attempt” since for both of the contours you highlight to show where the vortex edge is, it is clearly split.***

See response 50, above.

53. ***Lines 558–561, Again, if you work solely with ERA5 for the dynamical description, you don’t have this problem.***

See“High-Level Changes”, above.

54. ***Lines 563–565: All of this appears to me to be well equatorward of the contours you show to represent where the vortex is. Also the winds appear to be directed significantly cross-contour.***

The text has been removed.

55. ***Section 8, Overall: To read and try to understand this section, the reader is constantly flipping back to virtually all of the previous time series figures in the paper. For this material to be communicated effectively, all of these figures should be introduced and discussed consecutively, and combined into as few separate figures as possible so that flipping back and forth and trying to eyeball lining up times is not constantly disrupting the flow of the narrative. It would also be much better focused if you weren’t constantly referring to one quantity from one dataset and another from another, leaving the reader to wonder how much it matters which dataset you show (e.g., around lines 579–580). This sort of reorganization would vastly improve the paper, and would be straightforward to do if the Aeolus results were separated out into a different paper.***

I (Wright) disagree with the core argument here - perhaps due to misinterpretation of the written comment, or perhaps due to philosophical differences in our views on technical writing. Back- and cross-referencing are entirely normal parts of scientific (and general) writing; the sections are not intended to stand completely independently, and doing so would force an extremely artificial and narrow scope on both each individual section and on the manuscript as a whole.

56. ***Line 571: Would be good to say something about why you don’t use the full period available for ERA5 (back to 1950-something as I recall), especially in light of studies that show that modern reanalysis data for the “pre-satellite” era are useful for exactly this sort of large-scale dynamical studies (e.g., Ayarzagüena, et al, 2019, https://acp.copernicus.org/articles/19/9469/2019/; Hitchcock, 2019, https://acp.copernicus.org/a Gerber & Martineau, 2019, https://acp.copernicus.org/articles/18/17099/2018/).***

Both options would be good, and would be worth considering if starting the work again. We have chosen to stick with the existing climatological period to avoid unnecessarily repeating this analysis; since this choice of period is largely arbitrary between these two choices, any explanation would be inherently post-hoc rather than meaningful and accordingly one has not been added.

57. ***Figure 10, and discussion of snow cover: Showing snow cover is not useful unless the reader knows what is typical for these locations/seasons. Some indication of how anomalous these values are is needed, ideally showing them as an anomaly the way the other fields are shown. (Also, a zero Delta-T horizontal line would be really helpful in the regional panels.)***

A zero delta-T horizontal (dotted) line has been added to the regional panel; this line is also now the zero snowfall anomaly line.

58. ***Lines 574–578: The first sentence of this is quite repetitive. And wouldn’t you***

expect maxima, not minima, in wave fluxes when the vortex is starting to break down? (I also don't see the minima in all the places mentioned here, the timing doesn't line up to my eye – this would be much easier to see if all these figures were together or close together.) In addition, Z' appears to be positive in the UTLS throughout the period being discussed, so it seems this discussion doesn't really apply to that region?

This text has been removed in response to another comment.

59. ***Line 586: Is that local minimum expected, and if so, why?***

This text has been removed in response to another comment.

60. ***Line 588: Re "From the beginning of February" and "crossing 0 at the beginning of February", this timing depends strongly on what level you are talking about, which is not specified.***

Added "at the top of the shown height range".

61. ***Lines 591–592: Need to cite literature to justify why Z' is a good metric for strat-trop coupling.***

Two references and a discussion of the argument supporting this choice have been added to the manuscript in Section 7 (new numbering).

62. ***Lines 597, 600–601, 608, 627: As noted above, need some reference to climatology similar to that used here for 2m T' so that the reader knows how unusual the snow cover values are. You need to present or refer to something that provides evidence to say something is "extreme".***

Snowfall has now been converted to daily snowfall anomalies from ERA5, and normalised relative to climatology.

63. ***Figure 11: This would be a lot more convincing if you did something to assess the significance of these anomalies and/or the fraction of them that is congruent with Z' variations (e.g., similar to what was done for anomalies shown in Butler et al, 2017, doi:10.5194/essd-9-63-2017, or Lawrence et al, 2020, https://doi.org/10.1029/2020JD033271); especially re lines 605–606 where you say "surface impacts begin to appear", the congruence would help support the assertion that the surface changes are, indeed, related to the stratospheric changes. (Also, there are very strong regions of high anomalies in 2m T over the Arctic ocean / Siberia on 7 Jan, and over North America on 12 Jan; these are interesting and I think might be worth saying something about in the text.)***

We have added GPH contours to the Figure, showing that the positive 2mT anomalies are spatially co-located with the GPH anomalies.

64. ***Lines 612–613: How do we know that this is "high pressure over the Urals advecting warm air from the south"? I don't see anything in the analysis / figures that tells us that.***

This is now shown by the GPH contours on Figure 11.

65. ***Line 619–621: Again, I don't see how the material you have shown tells us this, so the statement needs further support.***

See response 65.

66. ***Lines 612–623: As presented, this is pure speculation. Should at least back this up by some brief statement about what is expected (based on the dynamics and the literature).***

It is speculation, but is clearly presented as speculation ("this feature <u>may</u> have acted...").

It is reasonable to expect that readers will accordingly interpret it this way and assign the appropriate degree of epistemological weight to the suggestion.

67. ***Section 9 (Conclusions) Overall: I'm not going to repeat here specific points that have been made above the analysis / data the conclusions are based on, but just to restate that a much better way to achieve each of these objectives would be to dedicate a focused paper to each one. Based on what has been shown in this paper, the agreement of zonal and meridional winds with ERA5 appears to be overstated.***

The conclusions have now been largely rewritten.

68. ***Abstract, page 1, line 8: should spell out Microwave Limb Sounder (and when / if you use the acronym, don't put "limb sounder" after it).***

Fixed.

69. ***Line 48: Replace "contextualize" with something less awkward (e.g. "put into context" is fine). (Also page 16, line 365.)***

This is a stylistic comment with which the lead author disagrees. We are happy to revise this at the copyediting stage if requested then for reasons of journal house style.

70. ***Line 71: Add comma after "ascending node".***

See response 69, above.

71. ***Line 98: The Schoeberl et al reference is not needed here, and does not contain the detailed information about MLS that you are citing it for.***

Removed.

72. ***Line 129: "however" should be set off by commas.***

See response 69, above.

73. ***Line 149: 737: "lengthscales" should be "length scales"***

See response 69, above.

74. ***Line 141: "due to" should be "because of", and a comma is needed after "observations"***

See response 69, above.

75. ***Line 142: "suggest" should be "suggests" (refers to "assessment")***

Fixed.

76. ***Line 143: "temperatures" should be "temperature"***

Fixed.

77. ***Line 162: "manney" should be "many" (Freudian slip?)***

Fixed.

78. ***Line 175: add "Aeolus" between "consider" and "data"***

This paragraph has been removed in response to a comment by the second reviewer.

79. ***Figure 5 caption: need "and" between "Centroid latitude" and "(blue line)".***

Word added.

80. ***Line 532: "g.km" as a unit is very confusing, just use "km".***

Done.

81. ***Line 603: "cooler" should be "lower"***

See response 69, above.

82. ***Lines 603–604: "however" should be set off by commas***

See response 69, above.

83. ***Line 615: “cold” should be “low”***

See response 69, above.

84. ***Line 629: Saying “Although” the data cannot show <x> followed by “therefore” the data strongly suggest <x> does not make any sense; I think this is just a matter of “therefore” being the wrong word here.***

The word therefore was incorrect, and has been removed.

85. ***Line 647: “downward-coupling” here should be “downward coupling”***

Fixed.

86. ***Line 745: Add a comma after “noise” (unless this is supposed to be read as a dependent clause, in which case change “which” to “that” to make that clear).***

The Appendix containing this line has now been removed.

C Specific Responses - Krisch

C.1 Major Points

A. ***The two different objectives of this paper make it sometimes difficult to read. The authors should therefore revise the manuscript and think about possible ways of restructuring to make it easier to read.***

Agreed: see "High-Level Changes" above.

B. ***The use of measurement data where possible is a large asset of this paper. However, this is not fully consistent throughout the paper. At many locations I would recommend to use MLS data (GPH and/or geostrophic wind) instead of ERA5 reanalysis data. More details on this are given below.***

Agreed for some Sections/Figures but not for others: see "High-Level Changes" above.

C. ***The calculation of the Aeolus projected zonal and meridional winds (which is in detail described in the appendix) should be revisited. More details can be found in the comments to the appendix. More detailed comments are given below.***

Agreed: see "High-Level Changes" above.

C.2 Specific Points

1. ***A paper on SSWs should probably include a reference to the most recent review on this topic, Baldwin et al. 2020 (e.g. to be added in line 35).***

This review paper was not used as reference material when preparing the introductory material (or indeed at all in the initial draft of the paper), and so should not be cited here as a reference used to do so. On other grounds, it is however now cited as methodological justification for using Z' in Section 7 (new numbering) and as evidence of the distribution of SSW types in Section 5 - see response to reviewer 1 comment 61.

2. ***L1-4: Major SSWs lead to a reversal of the wind. For SSWs in general (minor + major) the given definition is correct. The rise happens over just a few days, but has to remain reversed for a longer period. Maybe revise this sentence.***

Added ", and remain so for an extended period".

3. ***L5-7: Sentence itself is correct, but might be misleading as multiple other missions measuring temperature also provide wind products (even though not direct). This point is nicely explained in the introduction, but should also become clear in the abstract.***

The sentence has been modified to say instead "Due to the major technical challenges involved in measuring wind from space, while SSW-induced changes to the wind structure of the polar vortex have been inferred from other types of data, they never previously been directly observed at the global scale.".

4. ***L9-10: As you notice correctly in the introduction, already in Jan 2019 a major SSW was observed by Aeolus. Additionally, in Sep 2019 a minor SSW was observed over Antarctica. The data from Jan 2019 is reliable, but not yet available for scientific analysis due to large systematic biases. These biases need to be removed by reprocessing before the data can be used for scientific analysis. The reprocessing is planned for end-2021.***

Removed ", the first such event in the Aeolus data record" - this avoids an extended semantic discussion about whether the *first-released* or the *first-observed* is the first "in the ... data record"!

5. ***L17-18: Does this paper really show 3)?***

Arguably yes - a very very small contribution to this, but that's true of any paper. The mplicit pathway here is by improved knowledge of disturbed UTLS wind structures, which affect how the signal couples upwards and downwards.

6. ***L69: Maybe include Stoffelen et al 2020 here.***

Added.

7. ***L73/74: Are there no newer publications on ALADIN? 1989 is 30 years before launch!***

For ALADIN specifically it seems quite hard to find more recent papers in the open scientific literature about the hardware concept (a) as distinct from Aeolus as a whole and (b) that are focused on the instrument design and concept rather than validation of the built instrument. Also, while 1989 is a while ago, the papers do still describe the instrument design and conceptand are thus appropriate to cite, even if maybe not alone. To make this paragraph less jarring but avoid a potentially misleading inclusion of only tangential references, the Chanin reference has been moved to the end of the paragraph and merged with the other references about Aeolus, some of which include descriptions of ALADIN as part of the combined Aeolus system and are thus relevant.

8. ***L76: vertical resolution is 0.5-2km (0.5km is rare nowadays)***

Changed to 2 km.

9. ***L77: The hot pixel correction is described in Weiler et al 2020***

Added.

10. ***L78: Rennie and Isaksen 2020 could be updated to Michael Rennie, Lars Isaksen, Fabian Weiler, Jos de Kloe, Thomas Kanitz, Oliver Reitebuch: The impact of Aeolus wind retrievals in ECMWF global weather forecasts, QJRMS, 2021 (if this becomes available in time; the revised version was submitted recently and publication is expected soon).***

This reference has been added.

11. ***L90-94: As Aeolus data is one of the major components of this paper, I would propose to explain the vertical sampling of Aeolus a bit more in detail. First, the vertical sampling changes in height (maybe even include typical altitude profiles), then in location (the range bin settings are terrain following and change e.g. at 60°N), and last but not least also in time (however, this should not be the case throughout your study period). Especially the change at 60°N is worthwhile mentioning as this impacts all your estimates (means). Thus, I recommend plotting two sampling profiles at >60°N and <60°N to explain this change.***

A new figure (figure 1) has been added demonstrating coverage for one orbit, which highlights the 60°N transition clearly.

12. ***L69-94: The following information is missing here and should be added for clarity: Aeolus horizontal resolution, which data (baseline) was used, and which quality filtering has been applied.***

This information has been added to the text.

13. ***L102: Here you give precision, whereas for Aeolus you state the systematic bias (accuracy). As your analysis uses mean values, the precision of MLS is probably not representative for your data and the accuracy should be given instead.***

This information has been added to the text.

14. ***L130: Is MLS assimilated in ERA5/OPAI?***

MLS ozone is assimilated, but not temperature.

15. ***L145-151: This approach is a good idea!***

Thank you for this - I (Wright) was a bit nervous structuring the paper this way, but both reviewers approving so strongly makes me more confident about doing so in future. The data have now been replaced with ERA5, with no changes to the data, and the section accordingly removed.

16. ***Figure 1 and others: Are you sure the differences at 22km are not due to the Aeolus sampling (only data available >60°N -> data plotted is for 62.5°N, not for 60°N). You should verify that the averaged latitude is consistent for the different datasets.***

To correct for this issue, this plot and other figures showing a 60°N zonal mean have been modified to instead average the region 60–65°N consistently across all datasets. Differences are generally small, and as a result only small changes have been made to the text to reflect these. There is one exception to this: the previous 'underestimate' of the zonal wind minimum in ERA5 at 22 km has gone - presumably it was due to this spatial sampling difference - and the text describing it has been removed.

17. ***Figure 1: Interesting, that just before the SSW onset (day 0), Aeolus shows weaker winds than ERA5 at 22km whereas stronger winds at 15km.***

The changes made to spatial sampling and the calculation of u from Aeolus have changed these graphs.

18. ***L162: many previous studies (even though they are by Manney et al.)***

Fixed.

19. ***L170/171: The paper is long enough as is, thus, I understand the authors here. Still for such a new dataset as Aeolus closer investigations of the differences at geographical scale could probably raise the confidence in the projected meridional winds.***

Agreed.

20. ***L173/174: Please see comment above.***

These lines have now been removed.

21. ***L173-177: From my point of view, this paragraph is not really necessary, wrong (reliability of Aeolus data) and could be removed completely.***

See response 20, above.

22. ***Figure 2: To be consistent with Figure 1: Why not use similar or even the same altitude levels? Especially, why not show 10hPa (32km) as this altitude is used for the main definition of SSWs?***

In response to this comment, Figure 2 has been modified to show the same three levels as Figure 1, with an additional tropospheric level (5km) added for additional context.

23. ***L187/188: Maybe citation of a 2017 paper not correct for 2 events out of which one is from 2019?***

This is a bit of a tricky judgement call. The actual data are derived from Dr Butler's webpage, which is up to date. However, the webpage explicitly asks that users cite the 2017 paper, as the method is explained and derived there and the webpage is "grey literature" rather than part of the public (and, more importantly, formally archived) scientific record. So, for now we have left this reference as-is, but are happy to reconsider this choice if the editor or reviewers disagree on this course of action.

24. ***L214: mission-to-date day-of-year median: Is this for the time period 2004-2021?***

Yes.

25. ***L220: Also validation campaigns do not show a strong altitude dependence of the bias and reasonable values down to 2km (e.g. Witschas et al 2020, Lux et al 2020)***

Both added, and sentence slightly modified to support inclusion.

26. ***Figure 4: Again, I am not sure if an average around 60°N is recommendable for Aeolus data above 17km as this might actually be the average at 62.5km. This sampling bias should be investigated and discussed in the paper. For validation of Aeolus winds, it would be good to include MLS winds in this plot as well.***

See response 16, above.

27. ***Figure 3 & 4: I would somehow propose to combine these two Figures or put them close too each other as otherwise the reader is constantly going back and forth.***

In principle this makes sense, but actually merging them would require a reduction in size (which would hide detail) and rearranging panels between them would just shift the problem rather than fix it. A possible better solution would be to push at copy-editing stage for the two figures to be placed on adjacent pages and hence be relatively easy to cross-reference - would this work? (and, as a note to the Editor - is this a practical option to request?).

28. ***Figure 5: Before only measurement data (MLS & Aeolus) was used. Why introducing here ERA5 if MLS GPH data is available? What is the additional information or do MLS and ERA5 differ significantly?***

See response to Major Point B, above.

29. ***L323/324: Citing from Baldwin et al. 2020 (link provided above): "About a third of the observed 36 major SSWs in the 1958–2012 period can be unambiguously classified across all methods as splits and another third as displacements (Gerber et al., 2021). The rest of the events are more ambiguous across methods, perhaps because in some cases, the polar vortex both displaces and splits within a period of several days (Rao et al., 2019)." Thus, the 2021 event just seems to fall in the third category.***

The text " This is reasonably common, with typically a third of SSWs neither clear displacements nor splits **[Baldwin2021]**." has been added.

30. [Several comments were raised here about Section 6, which has now been removed as discussed above. Accordingly, we do not include specific responses to these comments.]

31. ***L440: Why again using ERA5 instead of MLS?***

See response to Major Point B, above.

32. ***L464/465: Is there already "typical" Aeolus winter data?***

The sentence has been rephrased to "is similar to Aeolus data from earlier in the winter" to make clear that the comment refers to earlier in the same winter rather than to (e.g.) a climatological mean over many winters.

33. ***Figure 9: Again, why not use MLS? Does MLS give significant different results or is the data too coarse?***

See response to Major Point B, above.

34. ***Sections 7.1 and 7.2 are somewhat repetitive. Maybe they could be joined and all Figures could be discussed simultaneously?***

Agreed and done.

35. ***Line 532: units not consistent between text and figure.***

The text is correct, but since the units have been changed in response to Reviewer 1 comment 80, above, no further change has been made.

36. ***Line 539-540: Can MLS winds provide this missing information?***

No, as MLS data also do not reach the pole (the northern limit is $\sim$82°N).

37. ***L 563: 85° N***

Fixed.

38. ***L639: Referring back to my citation of Baldwin et al 2020: 1/3 of all SSWs are not easily classifiable. Thus, better remove "unlike many others".***

Removed.

39. ***L644: For clarity add 60° *N*.***

Fixed.

40. ***L670: "timing" double***

Fixed.

41. ***L670: Is there a way to asses if this is really a timing effect or an offset effect?***

The difference mentioned here is no longer present due to the changes in data analysis method and averaging region.

42. ***Appendix: [multiple comments]***

The appendices have now been removed.