Review of “Dynamical and Surface Impacts of the January 2021 Sudden Stratospheric Warming in Novel Aeolus Wind Observations, MLS and ERA5” by C Wright et al.
(Reviewed by Gloria L Manney)

Recommendation: I recommend splitting this work into (at least) two papers; each of these in turn would need some fairly major revisions. See details following.

General Comments: This is a very long paper with two major objectives, to characterize the dynamical features and impacts of the Jan/Feb 2021 major SSW, and to assess the value of Aeolus wind data for studying this extreme event. There is a lot of good material in this paper that would be valuable and interesting to publish. Unfortunately, not only are the two main goals sometimes in conflict, but also in trying to fit / organize all this material into one paper, the presentation has ended up very disjoint, often repetitive, and there are many parts where justice is not done to one or the other goal because the appropriate background / citations are not included. Given the limitations of the Aeolus and MLS data used here, and the demonstrated robustness of modern reanalysis datasets such as ERA5 for large-scale dynamics studies such as that of SSWs and their tropospheric / surface impacts (e.g., see references given in specific comments below), the best way to characterize the SSW and its surface impacts would be to use only the ERA5 dataset. This is especially true since the main focus of this paper is on the middle stratosphere and below, where reanalysis temperatures are well-constrained by data and are usually recommended over those from most satellite datasets (including MLS, eg, Schwartz et al, 2008, doi:10.1029/2007JD008783), and mid to high-latitudes, where that constraint on temperatures provides a constraint on winds as well. On the other hand, the analysis of how an extreme event such as this is represented in the Aeolus wind data is very interesting and highlights this new dataset, but as currently presented is diluted by the focus solely on the SSW, with the result that a lot of information that is important to interpreting the results of the comparisons is omitted or unclear. I thus strongly recommend that the authors divide this material into two papers, one on each major objective. (One might argue that the Aeolus-focused paper would fit better in a different journal, such as AMT, but since it would be a “validation” focused on representation of the dynamics of extreme events, I think the argument could be made either way.) In this case, the major improvements / changes I see as important for each of these would be (specific examples / questions related to these are detailed in the specific comments):

For an SSW dynamical conditions / impacts paper:

1. Focus on analysis of ERA5 dataset.
2. Showing maps (including the 3D isosurface plots) of potential vorticity on isentropic surfaces or GPH and windspeeds (that is, the magnitude of the horizontal wind vector, sqrt(u^2 + v^2)) and/or vectors, rather than zonal winds, will give a much more clearer / more complete picture of vortex evolution.
3. Rearranging the figures so that the time series showing zonal mean winds / T, eddy heat and momentum fluxes, vortex moments (centroid latitude and aspect ratio), GPH anomalies, and regional characteristics (near surface T, snow cover) in a small number

of consecutive figures, and rearranging the discussion around this “comprehensive” picture will make the story much clearer, much less repetitive, much easier to follow, and therefore more effective in providing a complete picture of the SSW and its impacts.

4. Improving deficiencies in citing relevant literature, especially on stratosphere-troposphere coupling and wave fluxes, and therefore in motivating the choice of diagnostics shown and their interpretation.
5. Please note regarding Section 6 that the eddy heat and momentum fluxes calculated herein are **not** “mesoscale” fluxes; in fact, as shown in a vast body of literature, in the stratosphere they are dominated by effects of planetary scale waves.

For an Aeolus validation / representation of the SSW paper:

1. Again, because of the demonstrated quality of the reanalysis datasets, comparing in detail with ERA5 would be most convincing. This would mean showing differences of both zonal and meridional winds. This may sound odd (and in fact feels somewhat disloyal) coming from a member of the MLS team, but I don’t really understand the value / point of using MLS data here -- while derived winds from MLS (including quite decent meridional winds) have been calculated and used in numerous studies, they are not as robust as the winds in modern reanalysis datasets in regions where those are well-constrained by data; further, in and below the middle stratosphere, modern reanalysis temperatures are generally recommended over those from MLS.
2. More care needs to be taken in explaining / clarifying the conversion of LOS wind to zonal and meridional winds. Since these “derived” zonal and meridional winds are the focus of the Aeolus analysis presented, their quality needs to be demonstrated more thoroughly. To that end, I suggest that both the current appendices should be part of the main paper. There should also be more comprehensive direct comparisons with ERA5 winds (for example, why compare Aeolus-derived wind vectors with the morphology of / gradients in ERA5 GPH contours when you can compare with ERA5 wind vectors?), as well as better explanation of the situations for which the assumptions the conversion is based on are valid (for example, in a situation common to CAOs such as those studied here, where the upper tropospheric jet becomes nearly meridional, bringing cold air down from the north, it would seem that any assumptions based on zonal wind dominating over meridional wind would be invalid).
3. More thorough, clearly explained, and quantitative comparisons are needed with ERA5 to, first, verify the LOS winds, and, then for the derived (or guesstimated in the case of meridional wind) zonal and meridional winds.

The bottom line is, the authors are trying to shoehorn far too much material into one paper, and because of this not really doing justice to any of it. There is a wealth of good material here, and I applaud the authors for all the work they’ve put into it; this material should be published, but the current MS is not on the right path to do so effectively.

Specific comments:
(These are in order of appearance, not importance, in the current manuscript, but I believe will demonstrate many of the reasons I believe this should be at least two papers as well as the deficiencies I see in the current pursuit of the two objectives.)

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.

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).

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.

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.

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.

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"/

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.

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; https://doi.org/10.1175/1520-0493(1994)122%3C1115:SOTFSS%3E2.0.CO;2; and many others)?

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".

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.

Lines 149--151: I like this approach!

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

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).

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.

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

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.)

Lines 183--185: Although it doesn’t hurt to have this information here, it is most important to have it in the figure caption.

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).

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).

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

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.

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.

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.

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.

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.

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

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).

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).

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.

Lines 285--287: More precisely, it occurs earlier at lower altitudes. This does not tell us whether or not anything "propagates upwards".

Lines 287--288: Same as previous comment but re "propagating downwards".

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.

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.

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).

Line 304: It is not clear what “This minimum” refers to here.

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).

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).

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.

Section 6 Overall:
First, as noted in the general comments, the wave fluxes you are calculating here are not “mesoscale” wave forcing. While in principle, the scales of waves represented is limited only by the resolution of the data used to calculate them (so you could in fact resolve some mesoscale wave fluxes from ERA5, though not so much from the other datasets used here), in the stratosphere including, eg, around 100hPa, often examined to assess wave forcing of the stratosphere from below) these are dominated by planetary scale waves, with even wave 3 (still planetary scale) fluxes usually being much smaller than those for wave 1 and wave 2. In the

UTLS, smaller scale waves of wavenumbers 4--6 can be more important, but these are still "medium scale" or "synoptic scale", not "mesoscale". (I am at a loss for references here because this is textbook knowledge -- perhaps Andrew, Holton, & Leovy, Middle Atmosphere Dynamics.) The language in relation to this (as well as the Section title) needs to be adjusted accordingly throughout the section. In line with this comment, the discussion in Section 6 overall needs to be supported by statements referring to the vast body of literature on this subject that tell the reader what we expect to see in the wave fluxes and why. A couple of particular points where this is critical are line 385 (and lines 428--431), where you need to not only tell the reader why you expect strong heat fluxes before the SSW and give references for that, but also you need to mention and cite literature that indicates that this is not necessary in all cases to trigger an SSW (e.g., de la Cámera et al, 2019, https://doi.org/10.1175/JCLI-D-19-0269.1, and references therein), and around lines 395--399, where the reader needs to know whether the timescales you are talking about for influence of wave fluxes are consistent with those in previous literature.

Second, the combination of different datasets to get u, v, and T for calculations of momentum and heat fluxes seems problematic and completely unnecessary. In terms of getting robust values for these, you need look no further than the values calculated entirely from the ERA5 data. Re the satellite data only, combining MLS and Aeolus is highly suspect given large uncertainties in derived winds from those sources, and even more so because of the different sampling patterns; while I see no value added to the scientific results in this paper by calculating heat fluxes this way, if you are going to do so, you need to provide much more detail for and justification of how you combine these datasets (the same applies for the estimates that combine all three datasets).

Figure 6, and discussion thereof:

If you include correlation coefficients, you need to do / show something to assess their statistical significance (for daily time series such as these, that needs to be some method that accounts for autocorrelation).

If you include SSWs as late as 24 Mar, why do you exclude the major final SSWs in early-mid March 2005 and 2016? Given the wide range of dates of these SSWs, how do you reconcile different expectations for wave forcing, the zonal flow it is affecting, and interactions with radiative processes for all these events? Also, is it even possible given the highly variable timing (not only time in the season, but duration, e.g., the February 2007 and 2008 major SSWs were very brief) to even meaningfully define (lines 412--413) "typical post-SSW values" or "largely normal wind speed" in relation to these events?

Since the grey shading showing the range does not tell us which years / events given minima and maxima are associated with, we cannot assess the veracity of statements such as (line 405) "This degree of variability is highly anomalous…" since a minimum and maximum in a given time period could come from the same or different years (and could in fact at any given time given the small number of events consistently be from the same year) so may or may not represent the range of variability for a single event.

Finally, the Aeolus-related points (lines 420--435) disrupt the flow of the discussion of the dynamics of the event, and do not add any further insight into its dynamics -- this type of material would be very interesting (assuming one could justify the use of hybrid datasets to

calculate these fluxes) in a paper focused on how Aeolus represents the event in comparison to the reanalysis dataset.

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).
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.

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

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

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.)

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).

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.

Line 515: I don’t think I’d call an aspect ratio of 1.5 “roughly circular”.

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.

Line 522: I can’t see a “a tight detour towards the pole” in the GPH contours.

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.

Lines 539--540: But if you just used the ERA5 data for everything, you could show this!

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.

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

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.

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.

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/articles/19/2749/2019/; Gerber & Martineau, 2019, https://acp.copernicus.org/articles/18/17099/2018/).

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.)

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?

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

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.

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

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".

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.)

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.

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

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).

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.

Appendix A:
This “conversion” of LOS wind to zonal and meridional wind is crucial to a paper (such as the current manuscript) that uses those derived winds, and thus should not be relegated to an Appendix. The description of the conversion here is also very unclear. I went through this with my colleague Prof. Ken Minschwaner, who is, in part, an instrument scientist, and our student who is just starting graduate school in Atmospheric Physics, and even with ~ an hour’s discussion between the three of us, we could not figure out for sure exactly what your procedure is here. Two things that definitely need to be done are referring to the specific term(s) in the equations that you neglect in order to solve this underspecified system, and to either add a diagram beside (preferable IMO, because less noisy) or add lines / labels / angles to Figure A1 so that you can show clearly all of the angles / vectors that are involved with respect to both the measurement track and LOS direction and to latitude and longitude.
The method appears from my (limited) understanding to depend on the assumption that the meridional wind is much smaller than the zonal wind, but SSWs (in the stratosphere) and CAOs (in the troposphere) are exactly the type of situations where that is typically not the case, so something needs to be done to clarify why you expect this method to be reasonable in these situations. This particularly concerns me because Figure A2 and A3, which are used to “validate” the method, are both for January 2020 (something that needs to be added to the Figure A2 caption), which was the opposite extreme of an anomalously strong/symmetric stratospheric vortex -- so I’m concerned that this “validation” may not be that relevant to the opposite extreme of the dynamical situation. Ideally, this sort of comparison (as part of a paper dedicated to assessing the value of Aeolus winds for analyzing dynamics of SSWs, CAOs) would be done for as many different situations as possible, which in this case should definitely

include 2021 as well as 2020. This is especially important in trying to justify the “empirical” factor of 10 multiplier used for meridional winds.

Appendix B:
This material, and more extensive comparisons and discussion, would also be appropriate for the body of a paper dedicated to assessing use of Aeolus winds for such dynamical studies. I see no reason to include the MLS temperatures here, since this is not a paper on MLS validation. In many places, here and in the main text, the comparisons are stated as if you are assuming the Aeolus data are closer to “truth” than ERA5, which is not warranted (and should not be simply assumed even if you could show that it was).

Minor Points (typos, grammar, etc):
(I’ve made no attempt to be comprehensive about this, since necessary revisions will change many details of the wording, so these are just things I happened to notice.)

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).

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

Line 71: Add comma after “ascending node”.

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.

Line 129: “however” should be set off by commas.

Line 149: 737: “lengthscales” should be “length scales”

Line 141: “due to” should be “because of”, and a comma is needed after “observations”

Line 142: “suggest” should be “suggests” (refers to “assessment”)

Line 143: “temperatures” should be “temperature”

Line 162: “manney” should be “many” (Freudian slip?)

Line 175: add “Aeolus” between “consider” and “data”

Figure 5 caption: need “and” between “Centroid latitude” and “(blue line)”.

Line 532: “g.km” as a unit is very confusing, just use “km”.

Line 603: “cooler” should be “lower”

Lines 603--604: “however” should be set off by commas

Line 615: “cold” should be “low”

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.

Line 647: “downward-coupling” here should be “downward coupling”

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 following are issues of grammar, usage, or style (depending on who you are talking to in each case) that appear repeatedly in the paper. While I know not everyone considers these “rules”, these are all cases where the phrasing used results in lack of clarity:

“Due to” should be “because of” (capitalization as appropriate for each case): lines 36, 142, 174, and 175.

Add Oxford (serial) comma: Title after “MLS”; line 37 after “reanalyses”; line 74 after “aerosol”; line 115 after “reanalysis”; line 130 after “radiosondes”; Figure 1 caption after “ERA5 u” and after “22 km”.

“That” rather than “which” should be used for restrictive clauses: lines 22, 245, 374, 553, and 569.