Reply to the Editor

We appreciate the editor's comments. As the editor mentioned, one referee gave a full review to our paper in the round of the quick-reviews. The editor suggested "The (full) review of referee #2 with some minor points can be considered by the authors now before publishing in ACPD- if they like - or these points can be considered later in the revised version for ACP". We had revised the manuscript following the referee's comments before publishing in ACPD. Here we would like to post our reply to the referee #2.

Reply to the queries and comments of Referee 2

We very much appreciate the Reviewer's efforts in considering our manuscript and making suggestions for improvement. In our detailed reply below, we reproduce *the reviewer's comments in blue italics*, while our replies are in a standard font.

This paper represents the state of the art in depicting the equatorial stratopause (and mesopause) semiannual oscillation in global reanalyses. Its focus is not on dynamical cause or dynamical diagnosis, but on differences among the reanalyses, in an effort to increase our understanding of "what is the best description of the lower mesosphere and upper stratosphere"? The authors are international experts who have been working intensively on this inter-comparison for several years. The exposition is logical and the technique of comparing standard deviations is helpful for concisely describing differences among data sets. They present helpful detail, both in assessing differences, and in attempting to diagnose the cause of the differences among analyses. I would like to suggest to try adding even more commentary regarding the likely causes of the differences among analyses. In particular, it would be helpful to add a few sentences in the introduction or data and analysis section which describes the altitude range of reliable data for MLS and SABER, and the degree to which they are used in reanalyses, since these are the primary contributors toward improved representation of winds and temperatures in the USLM. It might be helpful (but certainly not required!) if a summary graph could be included showing an idealized version of the altitude range of reliability for each of the analyses, with specific labelling and footnotes which offer likely reasons for diminished quality. If the authors have access to the following information, it would also be interesting to clarify the degree to which analyses are more strongly regressed toward Singapore radiosondes compared to other tropical stations with comparable accuracy and frequency of launch. I recommend publication with minor revision.

We appreciate the Reviewer's positive overall response to our work and for the valuable suggestions

provided.

1. l17-18: This basic difference in structure is probably just the latitudinal profile of the Coriolis parameter. It's too bad that there are very few rocket wind profiles to include. That means that differences in low latitudes among analyses may have a lot to do with differences in the manner of treatment of building up heights from temperature soundings. The fact that SD for temperature is larger in the polar regions may reflect modest differences in sampling of actual large-amplitude Rossby and gravity waves.

We agree with your comments.

2. l19-21: This, and other things that I have read, strongly suggests that there is an embedded preference in algorithms underlying most global analyses which favors Singapore, simply by using a higher weighting factor, compared with other stations. This seems to be a legacy of respect, but should perhaps be relaxed, particularly if there are other reliable radiosonde stations in the tropics (please state them, perhaps near p2. L15-16).

In fact the Singapore radiosonde observations are rather special even in terms of availability even over the recent decades (not just in the more distant past). At least in terms of observations that are archived in the IGRA data set we know that during 1979-2001 at 10 hPa, Singapore is the only station with data available are over 80 % of the time (Fig. 5b in Kawatani et al. 2016), resulting in large impact of Singapore station on zonal winds in other longitudes. As far as our knowledge, each reanalysis center assimilates radiosonde archive data after doing quality control. We suppose that in any reanalysis system the weighting and area of influence of data from each location would be determined entirely by an objective algorithm (for the JRA-55 reanalyses we have explicitly confirmed this with the scientists involved). We suppose therefore that there is no likelihood that the legacy of the Singapore observations in the distant past has affected the weighing for the Singapore data. What our results do show, of course, is that, in practice, the influence of the Singapore data seems to be strong in all the reanalyses even above 10 hPa. Whether this finding should be regarded as a deficiency to be corrected/adjusted somehow is an issue for the reanalysis centers.

3. l24-27: This is a key theme in the paper. You mention sponge layer differences and the interesting JRA-55 and 55C difference. Do you know of any specific algorithm-based differences in how different satellite data streams are dealt with in different reanalyses?

We confirmed that the S-RIP team (one of co-leads is Masatomo Fujiwara, a co-author of this paper)

did not collect the information about algorithm-based differences. Section 5.2 in Fujiwara et al. (2017) explained satellite radiances assimilated in reanalyses in detail. Wright et al. (https://jonathonwright.github.io/pdf/S-RIPChapter2E.pdf) present an overview of how the satellite data is used in four of the most recent full input reanalysis systems (their Table 2.23, p81).

In our revision we point the reader to the relevant discussion in the Fujiwara et al. and Wright et al. S-RIP introductory papers. Specifically, we added the sentence "Detailed information such as assimilated satellite datasets used in each reanalysis were provided by the S-RIP project, notably summarized in Fujiwara et al. (2017) and Wright et al. (https://jonathonwright.github.io/pdf/S-RIPChapter2E.pdf)" in the introduction of the revised manuscript.

4. p2, l20-25: Can the influence of HRDI be pointed out in the figures? Are there any lidar temperature profiles that are ever included in global analyses?

Fujiwara et al. (2017) and Wright et al. show the lists of assimilated satellite data. There is no information about HRDI and indeed it seems unlikely that the very experimental HRDI data, lasting only for a brief period, would be considered in the reanalyses. We have no indication that any middle atmosphere lidar observations were assimilated (for the JRA-55 reanalyses we have explicitly confirmed this with the scientists involved). Again, readers of this paper could refer to the introductory S-RIP papers for this background on the details of the assimilations.

5. p3, l1-5: When I was doing this for LIMS data, I tried several smoothing techniques for building up Z in the tropics to obtain a good match for zonal winds with rocketsondes. I found that a 1-2-1 smoother in latitude for temperature applied at each level before integrating thicknesses upward yielded better agreement than smoothing at each level independently. I also tried smoothing more at each level, which degraded the comparison. I also tried smoothing across different ranges of latitude, and decided that within 8s-8n is about right, so as not to include thickness information from the subtropics, spreading inward and upward. I didn't like the results from using a cubic spline, which can yield larger amplitudes at higher altitudes. I only mention this because the growth with altitude of SD among analyses for zonal wind is largest at the equator, making it likely that such simple differences in how smoothing is done may explain quite a bit.

We thank you for your comments. The procedure and/or assessment for derived zonal wind from observed geopotential height in satellites are beyond scope of this study. SD are calculated from reanalyses data only and satellite derived zonal winds are not included in this calculation. The details of how the "observed" mean winds we derived from satellite temperature are in Smith et al. (2017).

As we explained in our manuscript none of the reanalyses assimilate SABER temperature data and only MERRA-2 assimilates the MLS temperature data above 5 hPa after August 2004. So the growth with altitude of SD among reanalyses of zonal wind does not result from differences coming from smoothing procedure.

6. p3, l14: Could point out the lack of raobs in the central and eastern pacific.

We revised the sentence as follows:
"This can be attributed to the following: the relative paucity of in situ data (especially in the eastern and central Pacific area with few stations, see their Fig.5), "

7. section 2: Several kinds of sources for differences are mentioned, including sponge layer treatment, and non-orographic gravity waves in MERRA-2. Is it possible to give provide more information about how different satellite data streams are treated in different analyses? On p3, l28 it was suggested that such is not possible, but to whatever extent the authors are aware of helpful information in this regard, please do describe further. Otherwise, it may be time to send in investigative reporters to find out what is in those black boxes, anyway.

We appreciate your suggestions. As mentioned, in our revised Introduction we point the reader to the S-RIP introductory papers (Fujiwara et a., 2017 and Wright et al., 2017) for what we know about the details of the reanalysis procedures.

8. section 2: If any reanalyses include SABER or MLS data, please describe. If so, can their effect be seen in the figures?

We have added the following sentence to section 2 (P6L3-4).
"Note here that MERRA-2 is the only reanalysis that assimilates temperature data from Aura MLS but only at pressures less than 5 hPa, and none of the reanalyses assimilate SABER data (Fujiwara et al. 2017)."

9. p6, description of Figure 1: Are there any other features of interest to point out besides the lack of information above 10 hPa in JRA-55C and increasing disagreement at higher levels?

We later discuss the difference of SAO structures between JRA-55 and JRA-55c in Figs. 8-12. So, we confined ourselves to a brief mention of this in our discussion of Figure 1.

10. Figure 2: Hitchman and Leovy (1986) summarized what was known from rocketsondes about the time mean vertical mean profile, which includes time mean westerlies in the lowest stratosphere, easterlies in the middle, and westerlies in the lower mesosphere, as shown here. The differences in time mean profiles in the MS shown here could be more strongly emphasized as a theme (cf. Fig. 3b). The figure 2 caption and figure need to be reconciled (there are two dashed profiles). It is very hard for this reviewer to distinguish the differences among the colors chosen for profiles. Please try to distinguish profiles more clearly, perhaps with dash-dot or thickness variation.

This is very helpful comments. We add the sentence here as follow: "The long-term mean of the satellite zonal winds showing mean easterlies in the middle stratosphere and westerlies in the lower mesosphere is in good overall agreement with that computed from rocketsonde observations at low latitudes (Hitchman and Leovy, 1986)."We also improved Figure 2 following your suggestion.

11. p7, discussion of Fig. 4: I don't see a special change in 1999. Please discuss why you include the MERRA panels.

Kawatani et al (2016) show a clear drop of zonal wind SD around 2000 at 50 hPa (their Fig. 13). However, the special change around 2000 is not very clear in Fig. 4. So we write here "Both zonal wind and temperature SDs decrease with time, a trend that is particularly clear at levels from 70 hPa to ~3 hPa." and remove "drop" in the Summary section in the revised manuscript. As discussed in Kawatani et al. (2016), both satellite data availability and the number of radiosonde observations (their Fig. 15) contribute to reduce the SD among reanalyses. To explain better why we included the MERRA panels, we revised the sentence explaining the SD as follows: "The SD among both six and five reanalyses is discussed up to 1 hPa whereas the SD between MERRA and MERRA-2 is discussed up to 0.1 hPa (i.e., the maximum altitude of pressure data provided)".

12. p7, l26: "geographical" suggest variation in (x, y) to me. Perhaps "variation in the meridional plane"?

Here we show latitude-height cross section in Fig.5 and horizontal distribution in Fig. 6. So we write "spatial variation of SD among reanalyses)"

13. p8, l2: The midlatitude maximum in SD for zonal wind may be related to the climatological mean maximum location for the polar night jet (cf. Fig. 6b).

We appreciate your suggestion. Figure A shows climatological zonal mean zonal winds in MERRA

and MERRA-2 for each season. We add the sentences as follows: “The local maximums of zonal wind SD around 50°S and 50°N at 0.1-0.3 hPa are due to the different shape of the mesospheric jets between MERRA and MERRA-2 (not shown). This may result from the inclusion of parameterized non-orographic gravity wave drag in MERRA-2. The observational constraints in the lower mesosphere are much weaker and the representation of the zonal winds may depend strongly on the model configuration used in each reanalysis”.

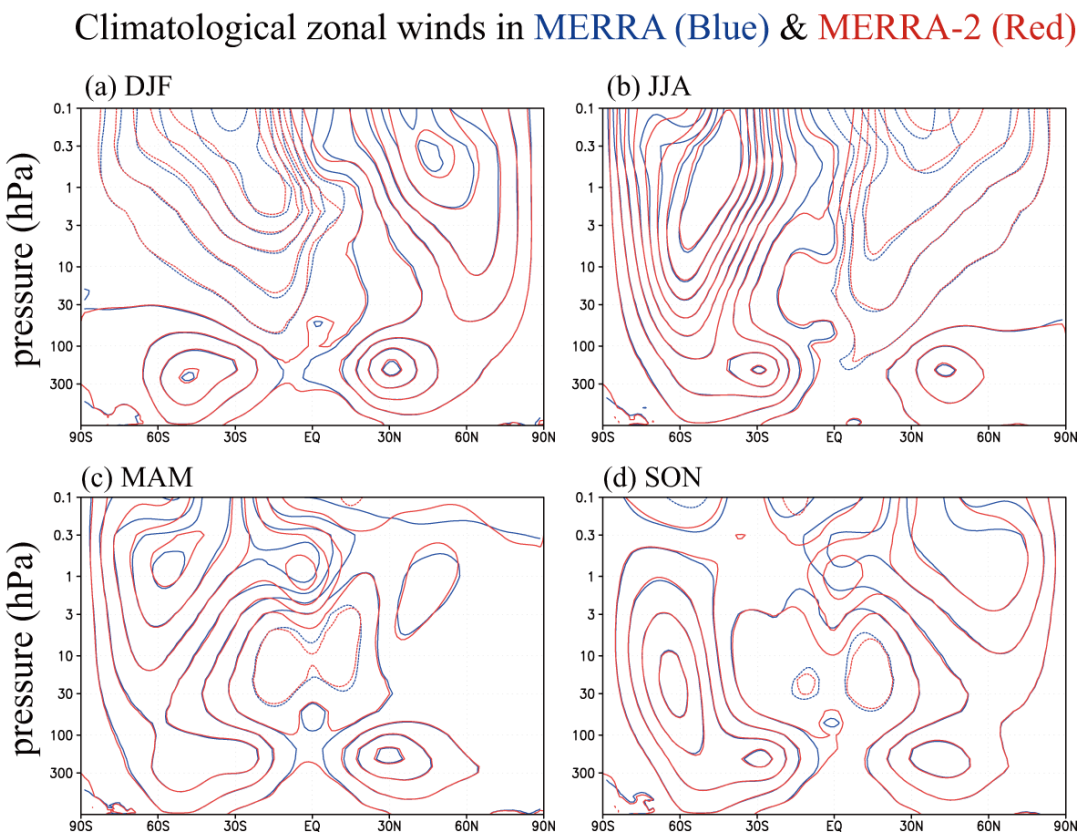

Fig.A: Climatological zonal mean zonal winds for (a) DJF, (b) JJA, (c) MAM and (d) SON in (blue) MERRA and (red) MERRA-2.

14. Fig. 7: Again, there are two dashed lines but only one is described in the legend.

We improved the figure.

15. p11, l22: Which other analyses include MLS? Do any include SABER? Please clarify early in the paper.

We have noted in section 2:

“Note here that MERRA-2 is the only reanalysis that assimilates temperature data from Aura MLS but

only at pressures less than 5 hPa, and none of the reanalyses assimilate SABER data (Fujiwara et al. 2017)".

16. p12, l1: Please weigh in with an editorial decision. My understanding is that an apostrophe takes the place of missing letters, such as in the word "doesn't", or indicates possession, but that plural never has an apostrophe, so it makes reasonable sense to write "1900s" instead of "1900's". Yet "1900's" is quite commonly used.

We have now changed our convention to not include the apostrophe – "1990s", etc….

17. p13, l11-12: This also seems to indicate a very strong regression coefficient for Singapore winds in many algorithms.

Yes indeed. See our comments on the assimilation of Singapore radiosonde data above.