Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2018-945-RC1, 2018

Atmospheric Chemistry and Physics
Discussions

Interactive comment on "Global Tropopause Altitudes in Radiosondes and Reanalyses" *by* Tao Xian and Cameron R. Homeyer

Anonymous Referee #1

Received and published: 17 November 2018

This paper documents the lapse-rate primary and double tropopause reported in the most current reanalyses. The purpose of the paper is suitable for ACP, especially for the SRIP special issue. However, I request a minor revision to the manuscript because there are a few places that need to be clarified.

In my understanding, this paper targets two parts of validation: 1) comparing to radiosonde data, how high the primary tropopause and how often the double tropopause showing in reanalysis, and 2) how are the long-term variability of tropopause inferred from the reanalyses comparing to that from radiosondes.

The first part of validation is very useful for the UTLS community because lots of analysis relies on using the tropopause level as a reference. Even though the detailed tropopause heights might be tremendously different from what's showing in the ra-

diosondes (e.g., Fig. 1), the variability of tropopause in reanalyses might not be that different, because the variability is more or less determined by the model settings and core dynamics that are consistent along the run. This is clearly shown in the analysis that in terms of primary tropopause altitude, MERRA-2 behaves much worse than the others, but in terms of trends, MERRA-2 is not that bad at all. Therefore, the long-term trend of tropopause is potentially useful to the community connecting the UTLS studies to the big scenario of climate change – although at this point the more urgent request would be to disentangle what caused the discrepancies in trend analysis among different realayses.

I am glad that the authors took my suggestion and added Fig. 1 to the revised manuscript. This figure illustrates how the relatively coarse vertical resolution of reanalyses could have distorted the tropopause analysis. This enhances our voice for the modeling community that adding more levels around the sharp gradient (both physically and chemically) tropopause is absolutely necessary.

My major comment is that the authors should have included detailed analysis for both bias and RMS while validating. While the authors tend to emphasize the RMS statistics, I hope they know that the RMS, as a loss function, gives a relatively high weight on large errors since it is more sensitive to extreme values on long tails/outliers due to the fact that the errors are squared before they are averaged. So, sometimes very few extreme values can completely change the statistics, which is not desirable for this analysis.

In contrast, the bias is sometimes more intuitive because it tells us how much of absolute differences between the radiosonde and the reanalysis. However, the bias analysis is not perfect because the positive and negative biases will cancel out. I would separate the bias analysis into positive and negative, with positive means the reanalysis primary tropopause showing at a higher altitude, and negative means the reanalysis primary tropopause showing at a lower altitude. Meanwhile, adding the frequency (with respect to total samples considered) of positive and negative bias. In this way, we know that on average how frequent and how much the reanalysis would overestimate/underestimate the tropopause height. I think this detailed analysis is more meaningful to the community.

In this sense, it is fair to always include both bias and RMS error analysis. I think Fig. 2 should include bias on the first panel and RMS error on the second panel. For each panel, different shapes represent different months, but please do include one more statistics for all-season averages. Then, add another figure that repeats a similar analysis for the double tropopause. Having the easy visualization of the statistics, still, keep Table 1 of detailed numbers for easy reference.

Another major comment is on the accuracy of IGRA data, and its ability to precisely document the lapse-rate tropopause is crucial for this study. It helps if the authors could iterate in more details on how the <= 50m vertical resolution of raw observation are eventually reported in only 1.5-2.5 km vertical resolution at the UTLS (although in the revised manuscript the authors changed to > 1km). The "> 1 km" is still less desirable for studying the vertical variability of temperature records - it will miss effects of both gravity waves and the Rossby waves acting on the temperatures. Given the reported resolution, why not using GPS/COSMIC temperature records that has a better coverage? The focus of the paper is from 5-20 km, in which COSMIC is totally capable of seeing waves on temperatures.

If I understand it correctly, Figs. 1-2 and Table 1 are the only places that the authors performed apple-to-apple comparison by collocating the reanalysis to the radiosonde locations. For all other analysis, the authors just reported trends inferred from gridded results at each latitude-longitude box, so the results could be biased by sampling sizes. So, the first part of validation is more meaningful to my sense. That said, personally I am not interested in the trends reported. For example, what does a trend of +/- 50 m/decade in primary tropopause mean? What does a positive trend of double tropopause frequency mean in specific reanalysis? Unless you can elucidate the possible cause of trends with proof, I don't think the trend numbers themselves have significant meaning. On the positive side, the fact that different reanalyses showing

different trends is meaningful in that they imply how unreliable the reanalyses are as to the tropopause analysis. This makes me wondering if it is necessary to include the trend analysis, especially in such a large portion of the paper. If I were the authors I would report the bias and errors in more details to help the community to understand the different performances of the reanalyses.

A last comment is that I do hope the authors could put more emphasis on the physical meaning/causes of the (large) differences among different reanalysis. So, beyond the vertical resolution, could there be any other reasons that caused the discrepancies? The current version seems to be less scientific and more like a technical report.

Minor wording comments: 1. P1L8, attributed –> attributable 2. P1L9: observations 3. P1L9: and reanalyses –> and the reanalyses 4. P1L9: analysis period 5. P2L10: –> the UTLS composition 6. P1L13-15: this sentence doesn't make sense 7. P3L6-7: this makes sense because of the existence of the ozone layer, but can you be more specific about it? 8. P4L1: I don't understand the logic here. If the radiosonde data is so limited, why bothering using them instead of COSMIC data? Plus, this part sounds like belonging to the discussion part. 9. P5L10-12: all reanalyses are reported in sigma or eta coordinates. From conversion you might get temperatures on pressure levels easily, but how did you get them in altitude coordinates? Did you use simultaneous geopotential heights to interpolate the data? Be more specific about how you preprocessed the data. 10. P5L16: –> quality-controlled 11. P5L23-24: one comment is that this linear interpolation doesn't change the shape of the profile, at all. So, you typically end up with the same value without doing interpolation. 12. P7L18: tropopause altitudes in MERRA-2 → primary tropopause altitudes in MERRA-2. 13. P8L16: is maximum –> its maximum 14. P13L11-12: how did you reach this conclusion?

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-945, 2018.