Reply to Reviewer 1

We thank the Reviewer for the careful reading and evaluation of the manuscript and the good comments which helped a lot to further improve the paper. In the following, we address all comments and questions raised (Reviewer's comments in italics).

We see two main concerns raised by the Reviewers, regarding (i) the complexity of the PVG tropopause determination algorithm and (ii) the analyses being carried out in zonal mean view only, which masks regional variations. A short overview of the related changes in the revised manuscript is:

(i) Simplification of the algorithm

We acknowledge the concern regarding the computational expense of determining the PVG tropopause from sub-daily global reanalysis data with several additional criteria concerning the height of maxima, latitude boundaries, and wind speed. In response, we have conducted additional analyses to examine the feasibility of simplified methods using monthly averaged data and the potential impact on the accuracy of the PVG tropopause latitude estimation. Specifically, we tested the following alterations:

1. Monthly reanalysis climatologies
2. Monthly reanalysis climatologies without the wind criterion
3. Monthly and zonal mean reanalysis climatologies
4. Monthly and zonal mean reanalysis climatologies without the wind criterion

These methods were applied to ERA5 reanalysis data, and we compared the results with the standard method that uses six-hourly, global reanalysis data with the wind criterion.
Our findings indicate that using monthly averages does not significantly alter the climatological latitude of the PVG tropopause between 340K and 370K, though deviations are observed at lower and higher isentropes. Omitting the wind criterion leads to increased noise, corroborating findings by Nash et al. (1996), due to the stabilizing redundancy of the PV gradient and horizontal wind peaks. Zonally averaging the reanalysis data and computing the PV gradient with respect to geographical latitude instead of equivalent latitude resulted in substantial changes to the results, as the method is designed to work with equivalent latitudes (cf. Nash et al. 1996, Kunz et al. 2011).
Based on these results, we do not recommend using zonal mean reanalyses or omitting the wind criterion. Averaging reanalyses over months yields similar results for the tropopause climatology near the tropopause break between 340K and 370K, but notable deviations above and below. The trends computed from monthly climatologies qualitatively show a similar vertical structure, but deviate numerically from the trends computed from sub-daily data. Therefore, using monthly climatologies may be sufficient if the focus is on the climatological structure of the tropopause break, but trends should be computed with sub-daily data.
The addidtional analyses of method simplifications are detailed in two new sections in the manuscript: Methods section 2.5 and Results section 3.5.

(ii) Regional climatology and long-term trend analysis

We agree that using a global zonal mean in our analysis could mask significant regional (longitudinal) differences. To address this, we have conducted a supplementary regional analysis focusing on these variations in the climatology and long-term trends of the PVG tropopause.
By identifying the PV contour for the tropopause in ERA5 data at each isentropic level, we determined global surfaces of the tropopause $\theta(\phi, \lambda)$. Poleward of the 320K contour, we continued the PVG tropopause by setting the tropopause as the PV isosurface corresponding to the PV value of the tropopause on 320K.
We computed zonally resolved long-term trends by applying multilinear regression to the tropopause intersection at each longitude and isentrope. The results confirm that the mid-latitude tropopause widens poleward overall in both hemispheres, while the tropical tropopause narrows equatorward. These trends vary noticeably with longitude. Comparison with Martin et al. (2020) reveals a similar longitudinal pattern; specifically, the strongest narrowing trends are observed over the east Pacific.
The corresponding methods and results are described in additional subsections of the Methods chapter 2.6 and Results 3.6.

A more detailed reply to all comments and description of the changes in the revised manuscript is given below.

1 Overall comments

Tropical circulation and so-called "tropical width" changes are important phenomena to better understand in the context of climate variability and change. The paper "Variability and trends in the PV-gradient dynamical tropopause" by Turhal et al. examines a new definition for tropical width based on a potential vorticity-based definition that is applied to several reanalyses. This paper is very well written and thorough, and I have no doubt it will be of great interest to ACP readers. I have mostly minor suggestions to clarify a few points in the manuscript.

Thank you for this positive evaluation of the manuscript!

2 General comments

The only "major" point I would like to suggest is that the authors consider whether their rather more involved methodology (which involves computing PV, etc. from the reanalysis model level fields at high temporal frequency and complicated peak finding rules) is really necessary in order to accurately quantify the seasonal to multidecadal variability in the PVG tropopause latitudes? I have no objection to applying this more "expensive" approach to analyzing the data if it is truly necessary, but it would be helpful for future users to know whether it really is necessary, or whether more simplified approaches with averaged fields could be used without a loss of fidelity (e.g., monthly mean PV/wind fields on pre-determined isentropic levels, as is commonly provided by reanalysis centers)? The answer to this question has implications for how easily this type of analysis might be extended in time or to other reanalyses, as well as how straightforwardly it might be applied to model output fields from multi model experiments (where high temporal / vertical resolution fields are often not made available). If the authors could provide some insight on this issue, I think it would be a valuable contribution to the community.

This is a very good point! Determining the PVG tropopause from sub-daily reanalysis data over 40 years is indeed computationally expensive. As explained above in our main point (i), we have therefore run additional tests with monthly climatologies of the ERA5 reanalysis and examined the influence of using zonal mean data and omitting the wind criterion, as detailed in a new section 2.5 in the Methods chapter. We computed the seasonal climatologies, mean seasonal cycle and long-term trends from these additional runs, which are presented in a new section "3.5 Simplifications of the PVG tropopause determination method" in the results chapter.

3 Specific comments

Line 53: I suggest defining PVU at its first use (i.e., the SI unit equivalent)

Thank you! Rephrased the paragraph at lines 44-48 as follows and included the unit definition of PVU:

"Another common definition of the tropopause, the so-called *dynamical tropopause*, is based on the *potential vorticity (PV)*, an analogue of angular momentum in air flow introduced by Rossby (1940) and Ertel (1942). PV is measured in *Potential Vorticity Units (PVU)* with $1\,\mathrm{PVU} = 10^{-6}\,\mathrm{m^2 s^{-1} K\,kg^{-1}}$. An invertibility principle holds which allows inferring the flow velocity field from the PV distribution. PV is therefore closely linked to atmospheric dynamics (Hoskins, 1985), which makes it particularly valuable for transport studies. "

Line 84: Suggest citing Waugh et al 2018 here.

Thank you for the suggestion! Replaced the other references with "Waugh et al, 2018" (now line 87).

Line 97: Suggest citing Santer et al 2003 here

Added the reference, thank you!

Line 114: I think "contracted" might be a more appropriate word here than "converged"

Thanks for pointing out this misleading wording! Changed it to "while the STJ of both hemispheres converged towards each other".

Line 119: "to which extent" → "to what extent"

Fixed!

Line 140 – 145: This intro section confused me at first because it seemed to not be a detailed enough description, but then I eventually figured out that the material was discussed in further detail in the following subsections. You might consider referencing the subsections here to make it clear that further details are provided there.

*Line 140 – 141 (and elsewhere): I found myself confused a few times in the manuscript regarding some of the uses of words like "gradient" where the direction of the gradient was not really specified. In this sentence (and throughout the paper) the gradients being referenced are meridional gradients on isentropes. Similarly, there's a bit of potential confusion around the discussion of the PVG dynamical tropopause being something that is defined based on a meridional gradient, versus what most people think of which is the tropopause being defined based on a vertical gradient/threshold. It might be worth considering the language in the paper and whether it would be more appropriate to use the term PVG dynamical tropopause *width* when referring to the latitude, rather than just calling it the PVG dynamical tropopause.*

Indeed! We reworked the intro section (now lines 143-151) and added references to the individual subchapters, and clarified that we compute a meridional gradient:

"The PV gradient-based (PVG) dynamical tropopause (Kunz, 2011) is determined as a contour on surfaces of equal potential temperature (i.e., *isentropes*) from the meridional gradient of potential vorticity (PV), combined with the location of the subtropical jet streams. This study is based on four different meteorological reanalyses: ERA-Interim, ERA5, MERRA-2 and JRA-55, which are described in Section 2.1. From six-hourly datasets, we employ the potential vorticity (PV), potential temperature (θ), zonal and meridional wind speeds (u, v) to compute the PVG tropopause, as explained in Section 2.2. The PVG tropopause is compared to the WMO thermal definition (Section 2.3). Variability and trends of the tropopause are examined via multilinear regression, as detailed in Section 2.4. To reduce computational effort and improve usability, we explore simplifications of the method, which are explained in Section 2.5. We conclude with a regional analysis of the PVG tropopause climatology and trends, which is described in Section 2.6."

Line 147: Why does the analysis only go through 2017?

We chose the period 1980-2017 for this analysis as it is the maximum common period for which we have the data from all four reanalysis available.

Line 175: I don't see the Bosilovich refence in the bibliography, but I think a more appropriate reference is Gelaro et al.

I seem to have mixed up the references there – thanks for pointing that out! Fixed now.

Line 180: What is the temporal resolution of the reanalysis data sets used here? The issues regarding the use of full resolution reanalysis data versus isentropic-interpolated monthly mean data could be at least partially addressed in this paragraph.

Thank you! We changed the sentence to:

"The PVG tropopause is calculated based on the zonal and meridional wind speeds (u and v), potential vorticity (PV) and potential temperature (θ) from six-hourly reanalysis data. A computationally faster alternative using monthly climatologies of reanalyses is discussed in Section 2.5 below."

Line 197: I think the lapse rate should be -dT/dz here?

Yes, thank you! Changed that in the manuscript.

Line 200 and 204: Would it be helpful to be more specific about the dimension (vertical or horizontal) which is being referred to here in the context of gradients?

Thank you for pointing that out! Changed the wording to "meridional PV gradients" in both lines.

Line 215: One thing that I was curious about in the PVG tropopause definition here is why both windspeed and PVG are used rather than just simply PVG. Perhaps this is addressed in the Kunz paper but could the authors mention somewhere in the paper why both are necessary? This relates to my comments around whether or not one really needs to implement the most maximalist analysis of the data in order to capture the important variability, or whether a more simplified definition (e.g., based on monthly mean PV data alone) would suffice?

Thank you for this insightful remark! In Nash et al. (1996), where the idea of using the PV gradient and wind speed to diagnose the edge of the polar vortex was first introduced, the wind speed is included because multiple prominent maxima can occur in both the PV gradient and wind speed, but only one common peak appears at the vortex edge. Therefore, including the wind speed reduces noise in the diagnostic. We tested this for the PVG tropopause by running the algorithm omitting the wind criterion, and indeed found that using only PV gradient leads to a substantial increase in noise compared to considering the product of PV gradient and wind – see the new Methods section 2.5 and Results 3.5.

Line 228: Again, going to the simplicity of the definition, is it really necessary to compute the break on equivalent latitude rather than geographical latitude?

This is also a good question! Considering geographical latitude instead of equivalent latitude is a substantial change to the method introduced by Nash et al (1996) and Kunz et al (2011). We tested the algorithm on zonal mean reanalysis data and computed the PV gradient with respect to geographical latitude, which yielded substantially different results than computing the PV gradient from globally resoved data and equivalent latitude (see Methods 2.5, Results 3.5). We therefore conclude that these two methods are not comparable, and computing equivalent latitude is necessary.

Line 235 – 248: These seem like a complicated set of rules and bring up several questions to me. First, do the authors really need to identify the PVG tropopause width at each reanalysis time step? If the goal is to create monthly means, why not average fields first? I suspect that doing something like this could allow the peak finding to be more straightforward and not require so many ad hoc rules.

Indeed, the rules appear complicated. Thank you for the suggestion! We tested using monthly mean data, which yielded similar results than using sub-daily data, see method chapter 2.5 and results 3.5. Considering the set of rules, these have shown to be still necessary when using monthly climatologies, because strong wind and PVG maxima also appear near the polar vortex, which requires a poleward latitudinal boundary and choosing the most equatorward maximum. In order to mitigate noise, we kept the relative prominence criterion of maxima.

Also, related to peak finding, there are several methods outlined in the TropD software package (Adam et al., 2018). Is there a reason the authors chose a different approach rather than adopting one of the well documented methods outlined there?

Thank you for suggesting this software package, this is a valuable resource. Line 27 in Adam et al (2018) states that the methods are designed for zonal means and is not tested yet on gobally resolved fields. While we are using globally resolved fields, this would require further testing with our method and is out of the scope of our current study. We consider using the method for maximum determination based on weighted latitudes (Eq. 1, Adams et al. 2018) in one of our next studies and compare with our current algorithm. Thanks again for the suggestion!

Line 265-267: I find the discussion around the seasonal cycle term confusing here. It seems as though it is something that is fit from the fact that it is included in the equation, but then it sounds like it is not included in the fit and that what is fit to is actually the de-seasonalized data.

Thank you! The seasonal cycle $S(t)$ is indeed computed in advance and used as a regressor. The multlinear regression takes as input the time series, as well as S, QBO_{30}, QBO_{50} and $ENSO$ and outputs the coefficients $a_{...}$. We added a sentence in line 274 which hopefully clarifies this:

"The regressors S, QBO_{30}, QBO_{50} and $ENSO$ are determined in advance of the multilinear fit:"

Line 276: The QBO winds at 30 and 50 hPa are not truly orthogonal to one another, and I think that a better practice is to fit to QBO EOFs. That said, my guess it doesn't significantly impact the results that much.

Indeed, fitting to QBO EOFs would be a good alternative approach. We chose 30 and 50 hPa winds as QBO

regressors, as these have been frequently used in similar studies
(e.g. Stiller et al., 2012, ACP, https://acp.copernicus.org/articles/12/3311/2012/).

Figure 1: I appreciate that it is easier to look at 2 plots than 4, but I always feel like I am missing something when people only show 2 seasons rather than 4. Given that the authors are trying to document a new method, I think it is important to show results for all 4 seasons. This could easily be done with the use of supplemental figures if the authors feel like it would significantly detract from the manuscript.

You are right! We originally chose to show only the solstice seasons, since the changes in climatologies, as well as the differences in reanalyses, are strongest between DJF and JJA. However, we acknowledge that showing the equinoctial seasons is important as well, and included additional plots for MAM and SON corresponding to figures 1–4 (A1, A2, A3 and A4) in the appendix.

Line 322: It looks to me like PVG and WMO tropopauses agree well up to 360K, but not 370K in all seasons/hemispheres.

Indeed, thank you, that was a typo! Changed that to 360 K.

Line 376-378: Is this due to STJ/EDJ separation?

Thanks for the idea – that is likely the case and we added a corresponding sentence:

"The larger PV variance in summer is likely due to the weakening of the subtropical jet and coalescence with the eddy-driven jet (Manney and Hegglin, 2018)."

Line 426-435: Is this due to horizontal resolution differences? Line 460-461: Do you have an idea of why ERA5 is different. Horizontal resolution?

These are interesting questions which we also tried to find answers to in the literature. However, we currently have no certainty as to why PV differs between the reanalyses. This would probably require further in-depth analyses, which are out of the scope of our current study.

Line 439: The word "Conclusively" doesn't seem right here

Thank you for pointing that out! Changed the sentence to:

"We therefore advise using the same reanalysis for comparisons between the PVG tropopause and other variables."

Line 585: "Hereof" doesn't seem like the correct word to use.

That indeed sounds a little awkward! Reworded the sentence to:

"The seasonal cycle accounts for most of the variability, shifting the tropopause north- and southward by 5° to 15° latitude, which is concurrent with the shift of the ITCZ."