We thank both reviewers for their detailed evaluations and positive assessments of our paper. We have carefully considered each comment and have made minor revisions to the dialogue where appropriate. Our detailed responses to each comment are provided in **blue** in the following pages, with original reviewer commentary in **black**.

Reviewer 1

This study investigates present trends of tropopause characteristics from GNSS radio occultation measurements for the time period 2003-2025, including height and temperature of the primary tropopause, tropical width, frequency of double and triple tropopauses and tropopause sharpness. Specifically, the authors focus on regional differences and claim to be the first to address tropopause sharpness trends as well as multiple tropopause trends from GNSS-RO observations. The mentioned tropopause characteristics are compared for two (three) tropopause definitions in the extratropics (tropics), namely the commonly used WMO lapse rate tropopause (LRT), the more recent potential temperature gradient tropopause (PTGT) and the cold point tropopause (CPT) in the tropics. The in-depth comparison of LRT and PTGT features is another novelty of the article. As such, the paper makes an important contribution to our understanding of the tropopause and its undergoing changes.

The manuscript is already in good shape. I suggest only minor revisions. While the conclusions already offer some discussion and outlook, certain statements throughout the manuscript would benefit from further elaboration, and the placement of the findings within the broader literature could be expanded in a few places (see detailed comments below).

My major criticism concerns the generalization that the findings are (apparently) insensitive to the tropopause definition (e.g., L8 and L274-275). While this holds for the definitions considered here, it should be made clear that the same behavior cannot necessarily be expected for dynamical or chemical tropopause definitions. Furthermore, given the close similarity of the results obtained with the LRT and PTGT definitions, the advantages of PTGT over LRT the advantages of PTGT over LRT should be articulated more explicitly (rather than only referring to Tinney et al. (2022)).

Thank you for the detailed evaluation and identification of areas where the text could be made clearer or fairer. We have addressed all feedback with the intent to simultaneously resolve this general concern.

In addition, I recommend adding a code availability statement.

We have followed all ACP policies.

Details:

L8: In my opinion, it should be made clear that “regardless of the definition” means: LRT, PTGT (and CPT in the tropics).

Done.

L22: For completeness, a reference regarding STE should be added.

Done.

L63-68: Beyond tropopause-break-based proxies, other approaches to characterizing tropical width and its changes exist (e.g. Davis and Birner, 2017; Waugh, 2018; Turhal et al., 2024). Notably, Turhal et al. (2024) and Weyland et al. (2025) report evidence of tropical narrowing in the SH based on zonal mean data and even a shape change of the tropopause break.

Exclusion of a mention of other proxies was only motivated by a focus on the tropopause, but we agree it is fair to at least acknowledge such and have added a sentence at the end of this paragraph (lines 71–73 of the revision). Also, thank you for referencing the Turhal paper as it aligns with the regional variability emphasis of our discussion here. We have added it to the cited papers.

L78: What about studies based on reanalyses? Trends derived from reanalysis data must, of course, be interpreted with caution. However, since key variables such as temperature are constrained by the assimilation of (satellite) observations, tropopause characteristics are generally captured fairly well in modern reanalysis products.

Reanalyses are captured by the more general use of "model output" here, but we have clarified that to make it more obvious.

L87: It would be helpful for the reader to make it clear at this point that tropical tropopause frequency is considered here to examine tropopause break latitude. You could simply move the brackets from line 92 up.

Done.

L119-120: In my opinion, these failure cases should be specified here (instead of merely referring to Tinney et al. (2022)), not least to highlight the advantages of the PTGT. How are these failure cases handled/why can the PTGT resolve these failure cases?

Done (lines 134–137 of the revision).

L169: A reference should be included here.

Done.

L179-181: The CPT altitude trends are (significantly) positive at the locations where LRT and PTGT trends are negative. This should be mentioned and discussed briefly.

Done (line 201–203 of the revision).

L185: Personally, I would prefer a title that clearly states that this section addresses the tropopause break latitudes/tropical width trends.

Revised Accordingly.

L198: I suggest showing the mentioned global frequency distribution. For example together with Figure 7.

We have included global frequency distributions for 1 month per season of one year as an example in a new appendix Figure A1 and referenced it here.

L210-212: A brief discussion of possible causes for this local tropical narrowing would be interesting.

Done (lines 235–236 of the revision).

Figure 6: The color bar in panel a) is not ideal. The CPT altitude appears to be very uniform.

While we agree the altitude appears uniform, this was an intentional choice to be consistent with the global distributions of PTGT and LRT altitudes in Figure 2. Thus, we have left it as is.

L216: A reference should be added here.
Done.

L220-222: It would be interesting to have a short discussion on why these local maxima exist.
Given the weak character of this longitudinal variability and its minimal relevance to the broader analysis here, we've decided to not expound upon it. It is somewhat related to the jet structure argument two comments below, which we've focused on.

L225: "most locations equatorward of 50°", I would say.
Done.

L229-230: This is an interesting observation and hypothesis, worthy of further discussion. Could you elaborate on how the changing tropopause sharpness and the jet structure are related and add the respective references?
Done (lines 256–258 of the revision).

Figure 9: In this figure as well, the color bar limit is not ideal. Consider choosing a slightly higher limit.
We've tried alternative ranges, but believe this provides the best compromise given the inherent differences between potential temperature gradients bounding the LRT and PTGT definitions (see also response to Reviewer 2).

L241-243: Here as well, it would be worthwhile to discuss hypotheses for the observed sharpness changes (already before the conclusion).
Done.

L247: Regarding the wording: In my opinion, the motivation to address (or not address) multiple tropopauses shouldn't be that it has (not) been covered in previous studies, but rather if its generally of physical interest/of any added value compared to double and triple tropopauses.
Agreed. We have thus revised and strengthened this statement (lines 273–275 of the revision).

Figures 13 and 14: Consider merging the two figures in one for easier comparison.
We had originally considered this, but believe the panels become prohibitively small for useful interpretation/evaluation. Thus, we have left these as is.

L248-252: I find it interesting that even though the climatological tropical width is different between PTGT and LRT (Fig. 2), the latitudes of the highest double tropopause frequencies are basically identical for PTGT and LRT (Fig 13). Maybe you could discuss this briefly?
We believe this impression is mostly driven by the necessary coarseness of the analysis bins. We considered discussing here but have elected to keep the text focused on the multiple tropopause diagnosis and trend results.

L258-260: Is one of the definitions closer to reality in your opinion? Maybe you could elaborate.

This is a tricky question. Since either case is largely a product of the selected threshold, each is comparably "real". However, it is possible that the physical/dynamical significance of diagnosed multiple tropopause varies between the two given the differences in the stringency of multiple tropopause definition. Given the complexity and subjectivity of this issue, we have elected not to elaborate.

L268-269: This statement is very vague. Could you specify the mechanisms?

Done (lines 297–299 of the revision).

L273: ...and the CPT in the tropics.

A sentence has been added here (lines 303–304 of the revision).

L274-275: Again, the observations might be insensitive to the tropopause definitions considered in this study, but not necessarily any tropopause definition in general.

Acknowledged.

L286-288: From Fig. 10 it's not evident to me that the sharpness trends are confined to the tropics and mid-latitudes. Trends in the subtropics look similar. This is also stated in L303-304.

The text has been revised accordingly.

L301-302: I agree that it can be beneficial to consider tropopause trends relative to the tropopause break, depending on the research question. Locally, however, the average tropopause altitude does indeed increase with increasing tropical tropopause frequency.

Agreed. This is why we used the more general term of "meandering" here because, as demonstrated in this study, tropical tropopause frequency is not increasing everywhere.

References:

Davis and Birner, 2017: https://doi.org/10.1175/JCLI-D-16-0371.1

Turhal et al., 2024: https://doi.org/10.5194/acp-24-13653-2024

Waugh et al., 2018: https://doi.org/10.1175/JCLI-D-18-0108.1

Weyland et al., 2025: https://doi.org/10.5194/acp-25-1227-2025

Reviewer 2

The manuscript analyses long-term change of the lapse-rate tropopause (LRT) and potential temperature gradient tropopause (PTGT) characteristics based on 23-years of radio occultation data. In comparison to existing studies, trends of PTGT altitude, temperature and tropical tropopause occurrence are new results. Furthermore, the analysis of the tropopause sharpness and the occurrence of triple tropopauses are new results for both tropopause definitions. While the manuscript reproduces a relatively large amount of already known findings, it also presents enough new results to warrant publication. The manuscript is well written and clearly structured, and I recommend publication after the following comments have been addressed.

Major comments:

1. It would be very helpful to have a summary of the new aspects of the study and what is already know. The authors provide this information in several places, but usually only for individual diagnostics. For instance, in the first paragraph of the conclusion section where all diagnostics are listed, add 2-3 sentences to make clear which of these diagnostics are truly new and which ones are updates of existing studies.

   We have added such a sentence to the opening paragraph of the Conclusions section (lines 309–311 of the revision).

2. Add a more thorough discussion of how the results here compare to existing results. This is done well for some aspects such as the tropopause temperature and altitude trends, but not for all diagnostics. For instance, how does the tropopause break analysis compare to the existing studies mentioned in the introduction (e.g., Martin et al., 2020; Zou et al., 2023).

   We have revised the conclusions carefully to add such commentary where appropriate.

3. Given that previous analysis demonstrated the sensitivity of the tropopause break diagnostic to the choice of threshold, the manuscript needs to address this issue. Ideally, the authors would include a sensitivity analysis by using varying threshold values for the tropical tropopause occurrence analysis.

   The foundational work on this topic, Birner (2012) - http://dx.doi.org/10.1029/2010JD014664, demonstrated that the tropopause break diagnostic is least sensitive to threshold choice when aligned with the frequency minimum in tropopause altitude between extratropical and tropical modes, which is what motivated the parallel analysis and ultimate threshold choice of 14 km in this study. As was true in Birner (2010) and many studies that have followed, the frequency minimum for the LRT is commonly found in the altitude range 13–15 km using reanalyses or observations and, consequently, 14 km is commonly used as a fixed threshold. For transparency sake, we have provided example frequency distributions as a new Appendix figure to show that 14 km is an appropriate choice for both the LRT and PTGT definitions (see also response to related comment from Reviewer 1) and have revised the text in Section 3.1 for clarity.

4. The manuscript needs to include the definition of the PTGT. Given that the PTGT is one of the two main quantities analyzed here and has only been defined recently, it seems strange that it is currently not included and readers are forced to look it up in other publications. The data and methods section is relatively short, which should make it easily possible to include the PTGT definition.

    We have commensurately expanded Section 2.2 to include the text of both definitions and have better acknowledged their differences in several places (see also responses to Reviewer 1).

Minor comments

Does the abrupt and relatively large change in GNSS-RO profile density in 2006 impact the trend results for the various diagnostics?

No. As originally noted in lines 107-110, all results we present are insensitive to this change.

Line 133-135: Do you show this here or is this known from other studies? Please clarify.

We identified this based on testing both the temperature lapse rate and potential temperature gradient. We have clarified this here (lines 152–154 of the revision).

Line 144: Using two terms for the QBO is usually standard practice in regression studies. Given the downward propagating nature of the QBO, two QBO terms are needed to correctly capture the QBO variability. These two terms are usually taken as orthogonal time series representing QBO variations, being constructed as the first two empirical orthogonal functions. Alternatively, an optimized time shift of the QBO proxy needs to be found and applied. Please clarify and improve if needed.

Good catch. We tested many configurations and the results were unmeaningfully different (almost identical) using one or two QBO terms. However, we agree that using a second time-lagged term is best practice and have revised the text for clarity (lines 162–166 of the revision).

Line 192-194: This sentence is not clear to me.

Clarified (lines 212–215 of the revision).

Line 222 states that differences between PTGT and LRT sharpness are greatest in the subtropics and tropics. Why is this case?

This is likely a result of the difference in definitions. The relationship between the temperature lapse rate and the potential temperature gradient has a pressure dependence, such that at a higher-altitude tropopause (lower pressure) the temperature lapse rate threshold of the LRT definition corresponds to an ~50% greater potential temperature gradient than the PTGT definition requires. We have added this explanation with citation here (lines 246–249 of the revision).

Line 264: Which ones are the aforementioned studies? I assume that this does not refer to the previous paragraph but to some earlier part of the manuscript.

Citations added.

Line 288 states that this is the case especially in the tropics and midlatitudes in regions where it is climatologically sharpest. It is not clear what this means. Is the sharpness largest in the tropics and midlatitudes in general or do the authors refer to specific regional features?

Sharpest in general. We have revised for clarity (lines 325–327 of the revision).