Response to Reviewer 2

We thank the reviewer for thorough reading of our manuscript and providing valuable feedback. We understand the two main concerns as (i) the representativeness of the ozone climatology for CARIBIC data is not sufficiently discussed and (ii) that a more thorough discussion of methodological differences between tropopause definitions and corresponding sensitivities was needed.

In the following, we address all comments in detail (Reviewer's comments in italic, quotations of the corresponding revised text passages in blue).

Based on suggestions from all reviewers, we have added a new Section 3.4 titled "Key differences in tropopause definition methodologies". Additionally, the quantification of tropopause-sharpness has been changed from a calculation of the cross-tropopause gradient to the curvature at the tropopause, with the title of Sec. 4.3 changed to "Vertical profiles and curvature at the tropopause" to reflect this change.

General Comments

(i) **Representativeness of Ozone Climatology**
The chemical ozone tropopause definition is based on a climatology that links ozone mixing ratios to a vertical distance above the thermal tropopause. The ozone profiles used to construct this climatology are from one particular station. It is not clear and not sufficiently discussed if this ozone climatology is representative for the regions of the flight campaigns. The manuscript could provide some sensitivity analysis (potentially as a supplementary material) on how climatologies based on multiple station data or satellite data would impact the presented results.

Thank you, we agree that a more thorough discussion of the ozone climatology representativeness would certainly be interesting. The study on hand focussed on the differences between tropopause definitions as they are currently commonly implemented. We therefore chose to use the chemical ozone tropopause that is available as part of the CARIBIC data set and applied the same processing to the PHILEAS observations for consistency. The sensitivity analysis would therefore, while certainly adding an interesting component, be better placed in future studies. To clarify the studies' intent, we have added the following in Lines 53–55 to the manuscript:

To allow comparability with previous studies, we put a focus on tropopause definition parameters as available in published data sets, rather than separate implementations..

(ii) **Chemical ozone tropopause application**
The motivation of the use of a chemical tropopause (given in section 3.3) argues that that the same data set can be used for the definition of the tropopause and the trace gas of interest. But this is not the case in this manuscript, or? The motivation also argues that the chemical tropopause avoids mismatches in dynamical parameters. However, with a one-station ozone climatology applied to a large geographical region it seems that these advantages of the chemical tropopause do not apply. Furthermore, it seems that the chemical tropopause is used in a different way (based on a climatology) while the dynamical tropopauses are based on the respective dynamical situation. All of this needs to be clarified and dis-

cussed more clearly.

The argument for using chemical tropopauses refers to the approach of deriving the tropopause from one tracer and to apply this result to other observables. The reference to the same data set in this case refers to the type of measurement being in-situ measurements / air sampling, rather than trying to match this with modelled or remote-sensing-derived quantities. The ozone climatology provides the basis for calculating representative height-differences to the tropopause for the chemical ozone tropopause, however specific measurements are then used to match each observation with a representative tropopause value. This stands in contrast to the modelled thermal and dynamic definitions, which provide the general larger-scale meteorological context for each observation on a pre-defined grid. The tropopause height is then an independent variable to the observations.

We have extended the discussion of differences between tropopause definitions through the newly added Section 3.4 "Key differences in tropopause definition methodologies".

(iii) **O_3 and N_2O typical profiles**

To illustrate the ozone based and N2O based tropopause definitions, it might be helpful to show a latitude-altitude cross section of each gas. Or some typical profiles.

Thank you, we have added the mean profiles of O_3 and N_2O showcasing the typical vertical distribution of these gases as Fig. A1 in the appendix. The figure is also shown as Fig. 1 of this response.

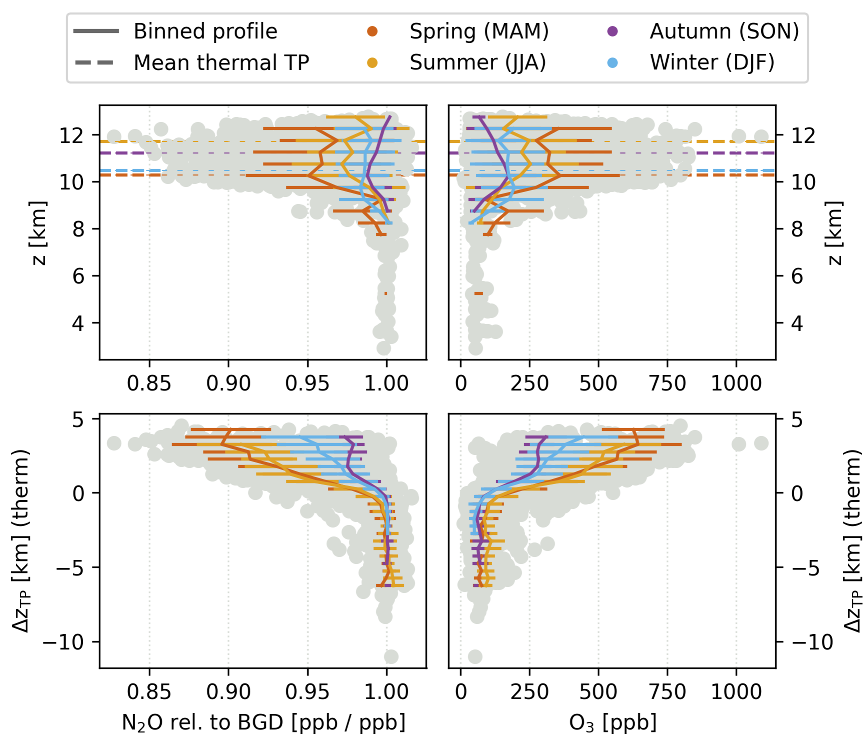

Figure 1: Seasonally averaged profiles of O_3 and N_2O in CARIBIC data. Top two panels show vertical distribution over geopotential height, while the lower to panels show the averaged distribution relative to the local thermal tropopause.

(iv) **Cross-tropopause gradient**

The use of the cross-tropopause gradient in Figure 7 seems not sufficient to assess the quality of separation into tropospheric and stratospheric air as the authors explain starting in line 374. Instead of using a supplement and whose interpretation needs to be combined with the interpretation of Figure 7, it would seem easier to evaluate the second derivative instead of the first.

Thank you, we agree that the second derivative provides a much stronger indication of this separation. We have revised this section and now present the curvature at the tropopause in Section 4.3. "Vertical profiles and curvature at the tropopause".

(v) **Data availability**

Some of the data is not freely available as stated in the manuscript. Following the link to the HALO database, MISSION: PHILEAS webpage, I found 'Begin free data access: 2028-09-29'.

We agree that this should be stated clearly. Moreover, we have since made the ERA5 interpolated data accessible online. The data availability section in Lines 536–539 now reads:
The whole-air-sampler data of the IAGOS-CARIBIC project are available at `https://doi.org/10.5281/zenodo.8188548`, the merged (MS) high resolution IAGOS-CARIBIC data at `https://doi.org/10.5281/zenodo.8188548`. The observational data of the HALO flights during the PHILEAS campaign are available upon request via the HALO database (`https://halo-db.pa.op.dlr.de/mission/138`). The interpolated ERA5 reanalysis data for CARIBIC-2 and HALO campaign data are available at `https://doi.org/10.5281/zenodo.15076519`.

Minor comments

Section 2 Data: What is the time period considered? How many flights are included in the analysis?
Thank you, we have added the following sentence in Lines 70–71:
The dataset used for this study includes 278 individual flights, taking place regularly between 2005 and 2020.

Line 21 (significant uncertainties in our understanding) and line 25 (knowledge gaps and uncertainties): I do not disagree, but these general statements need to be explained and backed up more. What are the uncertainties and remaining questions? Describe the general knowledge gaps and give some examples illustrating why these are relevant.
Thank you for the comment, we have revised the paragraph with additional information. Lines 20–27 now read as:
" However, in-situ observations in this complex region are rare, leading to significant uncertainties in our understanding of processes such as stratosphere-troposphere exchange (STE) or long-term trends in atmospheric circulations. Despite growing scientific interest and research in this region over the past few decades, key knowledge gaps remain: For example, while changes in the strength of the jet streams and the Brewer-Dobson circulation are expected, the extent and implications are not yet well understood (Butchart and Remsberg, 1986). Ozone trends in the exUTLS are also difficult to quantify due to the complex interplay between processes such as large-scale circulations and STE, which contributes to a large natural variability. These uncertainties extend to pollutant transport and consequently the overall impact of aerosols or short-lived trace gas emissions."

Line 83: in-fight -> in flight
Thank you, this has been changed in Line 88.

Line 98: I don't really understand the argument: 'the routes taken were not tailored to explore special features in the atmosphere'.
This line refers to the nature of the chosen flights as commercial routes, rather than scientific flights aiming to specifically sample air in "interesting" conditions. This approach stands in contrast to (scientific) flight routes aiming to "feature-hunt" and presents a more representative data set for typical states of the atmosphere. The same reason was applied when choosing the flights for the case studies: Rather than taking data from scientific flights during the PHILEAS campaign, the transfer flights were chosen to be more representative for the atmosphere. No changes have been made to the manuscript.

Line 110: Was 'MS data set' defines?
Thank you, we have changed the sentence to explain the type of the data instead:
The O_3 data from CARIBIC used here is part of the merged data set on 10 s time resolution

Line 250: This seems not really true north of 40N.
We agree and have changed Line 292 to read:
In contrast, both chemical tropopause definitions result in lower seasonal variability at latitudes between 30-50°N.

Line 258-263: Can downwelling impact tracer distributions and PV in different ways and therefore impact the differences in seasonality?
Indeed, the difference in the sensitivity to downwelling of different tropopause definitions will impact the seasonal differences between them. We have added the following sentences in Lines 305–307:
The different sensitivity of tropopauses to downwelling may additionally impact the seasonal variations. Changes in the chemical composition through increased downwelling in some seasons will immediately be represented by the chemical tropopauses, changes in PV may be more complex.
Line 269-270: The statement 'most notable ... spring' is not clear to me. We have changed the wording to express the content more clearly in Lines 312–315:
There are some slight differences between these definitions. In winter, the $\text{dynamic}_{3.5}$ tropopause identifies a larger fraction of tropospheric air in the data set, while in spring the stratospheric fraction is higher for the chemical_{O_3} tropopause than the other two. However, the overall seasonal variations are very similar for these definitions.

Line 286: Does 'with' need to be removed. Something seems wrong in this sentence. We have improved the readability of the sentence in Lines 330–331:
While the chemical_{O_3} tropopause results in a similar tropospheric fraction as the thermal and $\text{dynamic}_{3.5}$ tropopauses, the tropospheric fraction is much higher in autumn for the chemical_{N_2O} tropopause.

Line 286-291: Wouldn't this impact ozone and thus the chemical ozone tropopause in similar ways?
Thank you, we agree that this is not clearly explained in the text. We have added the following clarification in Lines 333–340:
This increased tropospheric fraction may be caused by the sensitivity of the statistical baseline al-

gorithm on the underlying dataset. The weakening of the subtropical jet and corresponding flushing of tropical air into the extra-tropics results in higher N_2O mixing ratios in the UTLS. The baseline algorithm may then identify the mixed air masses as tropospheric, while other definitions are less sensitive to these chemical composition changes. This effect appears to then not be apparent for the chemical$_{O_3}$ tropopause, as the climatology at the tropopause already includes this seasonality. Another effect causing this difference could be the seasonal variation of N_2O emissions propagating upwards. A systematic investigation of the statistical baseline algorithm across different seasons and conditions is recommended to better understand this behaviour.

Line 295 and other places: The term seasonal trends seems to refer to seasonal variations and might be confusing.

Thank you, we have changed this in Lines 315, 344 and 399.

Line 308-309: Does this show in the mean or in the standard deviation?

Thank you, this is an important point. We have added the following information in Lines 359–362:
For these definitions, 16% and 12% of stratospheric ozone mixing ratios were below 100 ppb, compared to 6% or less for other tropopause definitions. This results in strong decreases of the mean mixing ratio across all seasons and increases in the variability in regions close to the tropopause. Overall, the mean ozone value is shown to strongly depend on the tropopause definition, leading to a range of 60 ppb in the stratosphere and 21 ppb in the troposphere.

Figure 6: For comparability, it would be better to show the full y-range for the upper panels as well.

We have adjusted Fig. 6 accordingly.

References

Butchart, N. and Remsberg, E. E.: The Area of the Stratospheric Polar Vortex as a Diagnostic for Tracer Transport on an Isentropic Surface, Journal of the Atmospheric Sciences, 43, 1319–1339, https://doi.org/10.1175/1520-0469(1986)043<1319:TAOTSP>2.0.CO;2, 1986.