

Note: Reviewer's comments are presented in black font; authors' responses are presented in blue plain font; manuscript text quotations are presented in blue bold font.

Anonymous Referee #1

We would like to thank Reviewer #1 for her/his time devoted and the constructive and helpful comments.

General comment:

In the manuscript, CAMS reanalysis tropospheric ozone profiles are evaluated during folding events using ozonesonde and IAGOS aircraft data in Europe. A control run without data assimilation is then also used to understand the differences, especially in the upper troposphere, between the modelled and observed ozone concentrations.

The manuscript is well written and gives a nice overview of the current knowledge about STT and tropopause folding. The scope of the manuscript is very focused, and the methodology very clear, although the knowledge of earlier studies by the authors is almost a must in the description. The results are interesting, but I had the feeling at several places that the authors could go more in depth. The authors observe, but do not interpret their findings enough. I will give examples here below.

We would like to thank Reviewer #1 for the general comments. Regarding the not so extended interpretation of some findings, we understand the rationale of the comment, yet, there is not enough flexibility in that direction as several sensitivity experiments are required which in the framework of the CAMS reanalysis are unfortunately not feasible. Nevertheless, the possible reasons for the CAMSRA O<sub>3</sub> overestimation in the troposphere documented in the manuscript, arise from previous sensitivity experiments and experience with the IFS modeling system. More details on this and our point-by-point responses to the Reviewers comments are presented below.

- In the introduction, you might refer to the work by Zhao et al. above Asia as well <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020JD033955>, <https://doi.org/10.1016/j.atmosres.2020.105158>, and Luo et al. (<https://www.hindawi.com/journals/amete/2019/4375123/>).

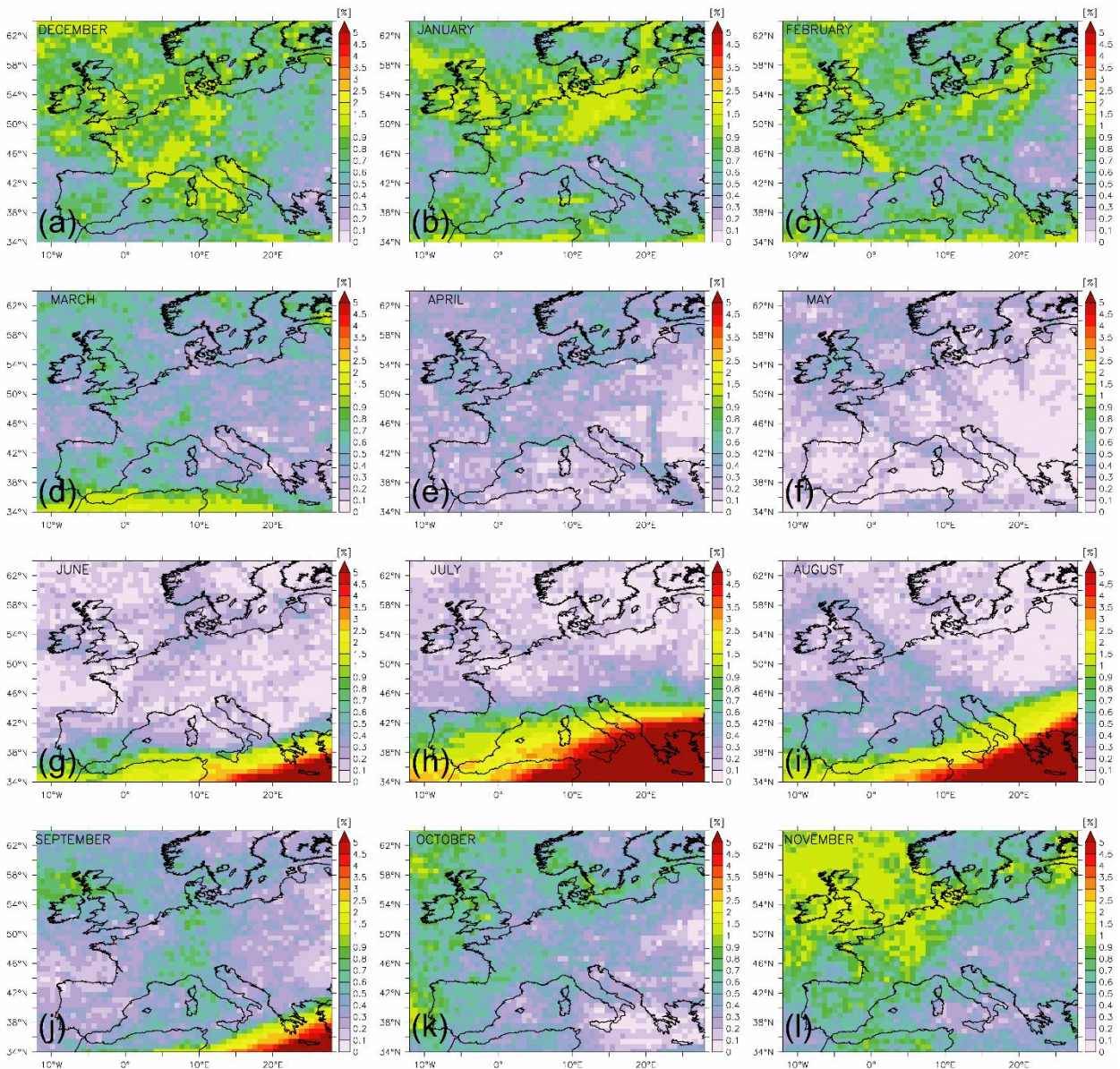
We have included the suggested references in the Revised Manuscript (RM).

- Why did you restrict your evaluation of the CAMS reanalysis ozone during tropopause folds to Europe? This is clearly not the region with the highest number of tropopause fold events, so I guess the availability of the ozonesonde and IAGOS profile data has driven the choice of the study area. Please clarify your choice for Europe in the introduction (lines 38-44 are not that convincing for the current focus of the manuscript).

The present work was performed within the framework of a postdoctoral fellowship of the first author by the Greek State Scholarships Foundation (IKY) (Reinforcement of Postdoctoral Re-searchers - 2nd Cycle" (MIS-5033021)) co-financed by Greece and the European Union (European Social Fund—ESF). The aim of the project is the investigation of stratospheric intrusions and their role on tropospheric ozone levels and air quality over Europe with the synergistic use of CAMS reanalysis and observational data. Thus, and as a prerequisite step in this direction, the evaluation of the CAMS reanalysis O3 during such events is limited over the European region.

- It would also be nice to give some additional climatological information on tropopause frequency, spatial and temporal variability of STT events over your study domain (Fig. 1) in Europe.

To present both the spatial and temporal variability of tropopause folds frequency over Europe we have included as a Supplement the fields of CAMSRA monthly mean folds frequency (%) for each month over the period 2003-2018. Very shallow folds with a vertical extent  $\Delta p < 50$  hPa are excluded. The following sentence is now included in the RM (L151-153): **"The spatial distribution of CAMSRA monthly mean tropopause folds (with  $\Delta p \geq 50$  hPa) frequency over Europe for the period 2003–2018 is presented in Figure S1 of the Supplement."**



**Figure R1.** CAMSRA monthly mean tropopause folds (with  $\Delta p \geq 50$  hPa) frequency (%) over Europe for a) December, b) January, c) February, d) March, e) April, f) May, g) June, h) July, i) August, j) September, k) October, and l) November over the period 2003-2018.

- It is not clear to me which selection criteria have been used for the ozonesonde and IAGOS sites. For instance, the Prague ozonesonde dataset, which you have been using in an earlier study, is absent. If the data criterion for using ozonesonde time series is number of observations available throughout the 2003-2018 time period (line 71), I do not understand why, for instance, Observatoire Haute Provence (OHP), Sodankylä, Valentia (?) data were not selected. The same question arises for the IAGOS airports: what is the data temporal coverage (line 77) criterion used to include data from an airport or not?

We agree with the comment raised by the Reviewer. Initially we included the ozonesonde sites with the most available observations (in terms of number and years). Inadvertently we didn't include some sites which we do now in the RM. In the RM, the criterion used for both ozonesonde and IAGOS sites selection is to exhibit at least 500 profile observations. This subjective criterion was applied in order to ensure a sufficient number of both observational sites and folding events to be selected for analysis. As a result, except the initially examined sites two additional ozonesonde sites are included in the analysis; Lindenberg (Germany) and Prague (Czech Republic). Please mind that Figures 1, 3, 4, 6, 7, 8 and 10 are revised, yet without changing the main findings of the study. The following is now included in the RM (L92-94): **“The selection of both ozonesonde sites and IAGOS airports was based in the availability of at least 500 profile observations throughout the 2003-2018 period. This objective criterion ensures a sufficient number of both observational sites and folding events to be selected for the analysis”**.

- As you included a control run without data assimilation, please specify the sources (which satellites? which products? during which period of the 2003-2018 time frame) of partial column and profile ozone retrievals that are assimilated in CAMS reanalysis.

The partial column (PC) and profile (PR) ozone retrievals assimilated in CAMS reanalysis are presented in Table R1. This information is provided in Table 1 of the CAMS reanalysis evaluation study by Wagner et al. (2021). The following sentence is now included in the RM (L106-107): **“More details on the satellite retrievals (product, satellite, period) assimilated in CAMSRA can be found in Table 1 of the CAMSRA evaluation study by Wagner et al. (2021).”**

**Table R1.** Satellite retrievals (product, satellite, period) assimilated in CAMSRA

Product	Instrument	Satellite	Period
PC 13L	SBUV/2	NOAA-14	200407-200609
PC 13L	SBUV/2	NOAA-16	200301-200706
PC 21L	SBUV/2	NOAA-16	20111201-20130708
PC 13L	SBUV/2	NOAA-17	200301-201108
PC 13L	SBUV/2	NOAA-18	200507-201211
PC 13L	SBUV/2	NOAA-19	200903-20130708
PC 21L	SBUV/2	NOAA-19	20130709-20181231

PR	MIPAS	ENVISAT	20030127-20040326 20050127-20120331
PR	MLS	AURA	20040803-20180312

- In section 2.3, I would expect more details on the fold detection algorithm. Now, the summary is very limited. Major clarifications: how is the stratospheric source of air identified, what is the weight of the specific humidity content, and are the ozone concentrations used in the detection (I guess not, but please confirm clearly).

Ozone concentrations are not used as an input to the 3-D labeling algorithm, thus are not used as a proxy in fold detection. The tropopause is defined as the isosurface of  $PV=2$  pvu or  $\Theta=380$  K, whichever is lower, so initially, an air mass is considered as stratospheric based on the criterion  $PV > 2$  pvu or  $\Theta > 380$  K, respectively. Yet, not all masses with  $PV > 2$  pvu should be considered to belong to the stratosphere, such as stratospheric cut-offs, surface-bound PV anomalies, and diabatically produced PV anomalies. To this end, some physical and geometrical criteria are used to categorize the air masses in five categories (5 labels): tropospheric (label=1); stratospheric (label=2); stratospheric cut-off or diabatically produced PV anomaly (label=3); tropospheric cut-off (label=4); surface-bound PV anomaly (label=5). The diabatically produced PV anomalies merged with the stratosphere are distinguished using a specific humidity threshold of 0.1 g/Kg. Further details of the applied algorithm and the criteria used for air mass labeling can be found in Škerlak et al. (2015). In the RM (L139-150) we have replaced the sentence of L113-114 with the following: **“The 3-D fields of pressure are constructed and the pressure level of the dynamical tropopause (Holton et al., 1995; Stohl et al., 2003) is determined using the lower of the isosurfaces of PV at 2 PVU and potential temperature at 380 K. Subsequently, the vertical profile for each grid point is examined and a fold is assigned when multiple crossings of the tropopause are identified. Still, there are specific cases where air with  $PV>2$  PVU is either not connected to the stratosphere (stratospheric cut-offs) or is not of stratospheric origin (diabatic PV anomalies or surface-bound PV anomalies) which should not be considered as stratospheric. To this end, the 3-D labeling algorithm, using physical and geometrical criteria, labels the air masses as follows: tropospheric (label=1); stratospheric (label=2);**

stratospheric cut-off or diabatically produced PV anomaly (label=3); tropospheric cut-off (label=4); surface-bound PV anomaly (label=5). The diabatically produced PV anomalies merged with the stratosphere are distinguished using a specific humidity threshold of 0.1 gkg<sup>-1</sup>. Further details on the criteria used for the 3-D labeling can be found in Škerlak et al. (2015). Therefore, a fold is identified when a 2→1→2→1 or 3 transition is detected on a vertical profile (from top to bottom), with the algorithm outputting a binary variable (0:no fold, 1:fold) for every grid point and time step.”

- The selection of STT events (section 2.4) in ozonesonde and IAGOS profile data seems to be rather indirect, based on your database of STT events detected in CAMSRA (with 3D-labeling algorithm). I assume this algorithm is not directly applicable to ozonesonde and IAGOS “2D” data? Please specify. However, algorithms exist to detect tropopause folds in ozonesounding data as well (e.g. Van Haver et al., <https://doi.org/10.1029/96GL00956>), so how did you confirm the presence of an STT event independently from CAMSRA in the ozonesonde and IAGOS data? Details are missing how a fold is found in the ozonesonde profile (line 126) and in the IAGOS profiles (lines 131-135).

No, the algorithm is not applicable to ozonesonde and IAGOS data. The comparison between CAMSRA and observations indicates that during the selected folds the observed ozone exhibits a clear increase compared to the rest of events confirming that in principle folds are also present in observations. Furthermore, the visual inspection of the individual IAGOS vertical ozone profiles for the 318 selected STT (fold) events over FRA indicates that in ~93% of the profiles clear ozone increases were seen in some part of the troposphere. In the rest of the profiles, the increase might be small in both observations and CAMSRA, due to small impact of the specific folds or issues related to temporal and horizontal resolution of CAMSRA.

- At the end of section 2.4, it would be nice to include some statistical information: how many of the CAMSRA STT events are also identified in the coincident ozonesonde and IAGOS profiles? Are the detection rates site dependent? Dataset (ozonesonde vs. IAGOS) dependent? What are the relative ratios between the STT events and rest of events? Again: site or dataset specific? Consistent?

Based on our methodology the fold events are identified in CAMSRA only, so such statistical information is not feasible. An explicit detection of folds from the observational data would rather serve as an evaluation of the 3-D labeling algorithm which is beyond the scope of this study. Moreover, a methodology for folds detection based on observed ozone concentrations would be not directly comparable to the 3-D labeling algorithm detection of folds which is based on CAMSRA meteorological data.

- In section 3.1 and 3.2, you mention some spatial differences between the CAMSRA and observational ozone profile differences, e.g. in lines 145-148, in lines 157-159. But, you do not give any explanation why features in the differences arise at some sites, and not at other sites. Is this related to data quality issues at some sites, instruments used, differences in the spatial representativeness of the tropospheric ozone observations at some sites compared to others, etc? Some discussion and/or thoughts would be helpful here.

Regarding the climatological comparison between CAMSRA and observations, all stations exhibit an overestimation of CAMSRA O<sub>3</sub> in the upper troposphere which is the main feature. Attributing the different overestimations among the examined sites in specific reasons is difficult, but potential factors can indeed be discussed. To our knowledge no site-specific data quality issues are reported, but both ecc and Brewer Mast ozonesonde measurements introduce uncertainties which might affect in different ways the comparison among the examined sites. In addition, the proximity of the selected grid points to the respective ozonesonde site, and the CAMSRA 3-D spatial representation of the IAGOS take-off/landing are also another factor to consider. Accordingly, we have included the following sentence in the RM (L184-187): **“The differences seen in the comparison between the observed and CAMSRA O<sub>3</sub> concentrations among the examined sites are subject to the uncertainties introduced by the ozonesonde instrument measurements, as well as the proximity of the selected grid points to the respective ozonesonde sites, and the CAMSRA 3-D spatiotemporal representation of the IAGOS take-off/landing routes.”**

As for the differences of CAMSRA-obs comparison during the selected STT events, in addition to the aforementioned factors and even more importantly is the vertical location and geometrical characteristics of these events. More

specifically, the pressure level in which a part of the fold appears over the examined site, directly affects ozone concentrations, and shapes its vertical profile. Thus, for observational sites with not so extended number of stt events the individual dynamics and ozone increases in different levels of the troposphere can form somehow unique structures of CAMSRA and observed O<sub>3</sub> deviations. In contrast, sites where more stt events were selected (e.g. Frankfurt and Munich) exhibit more smoothed profiles for both CAMSRA and observations.

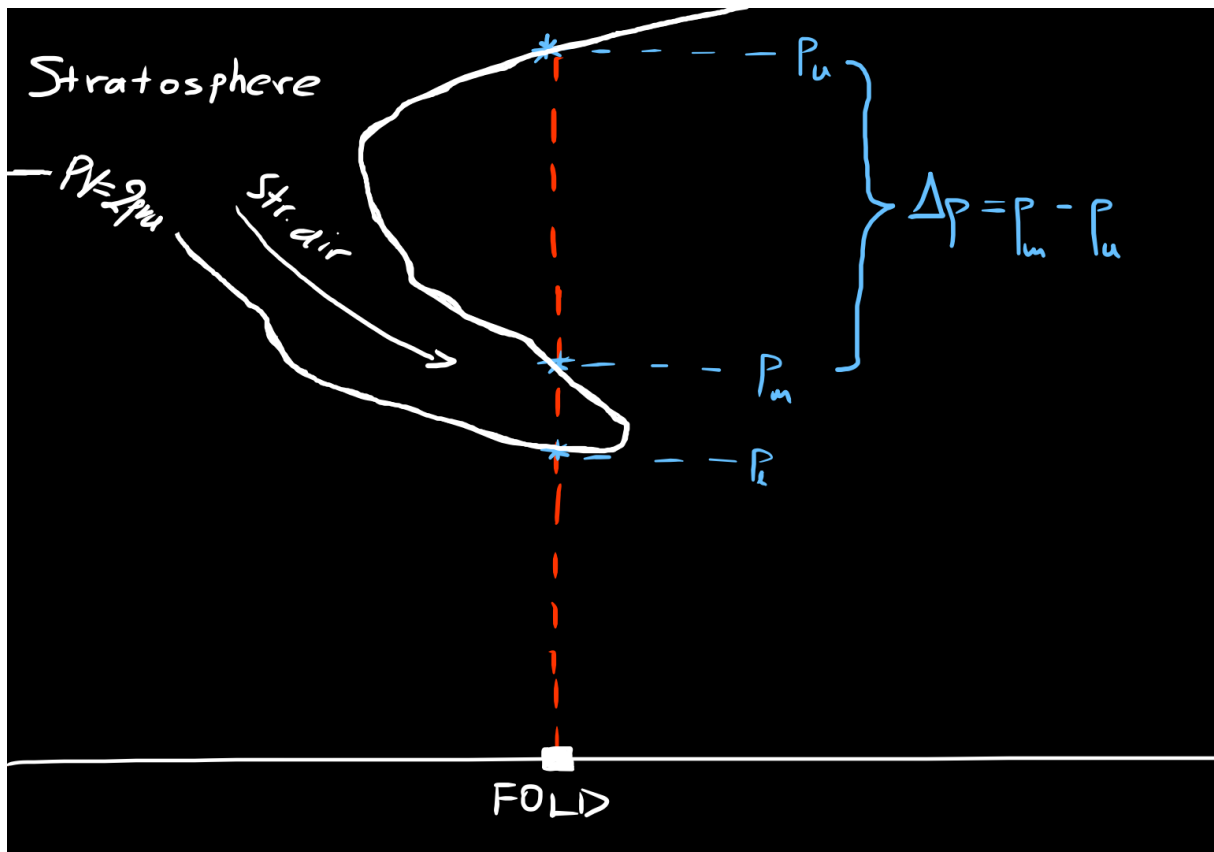
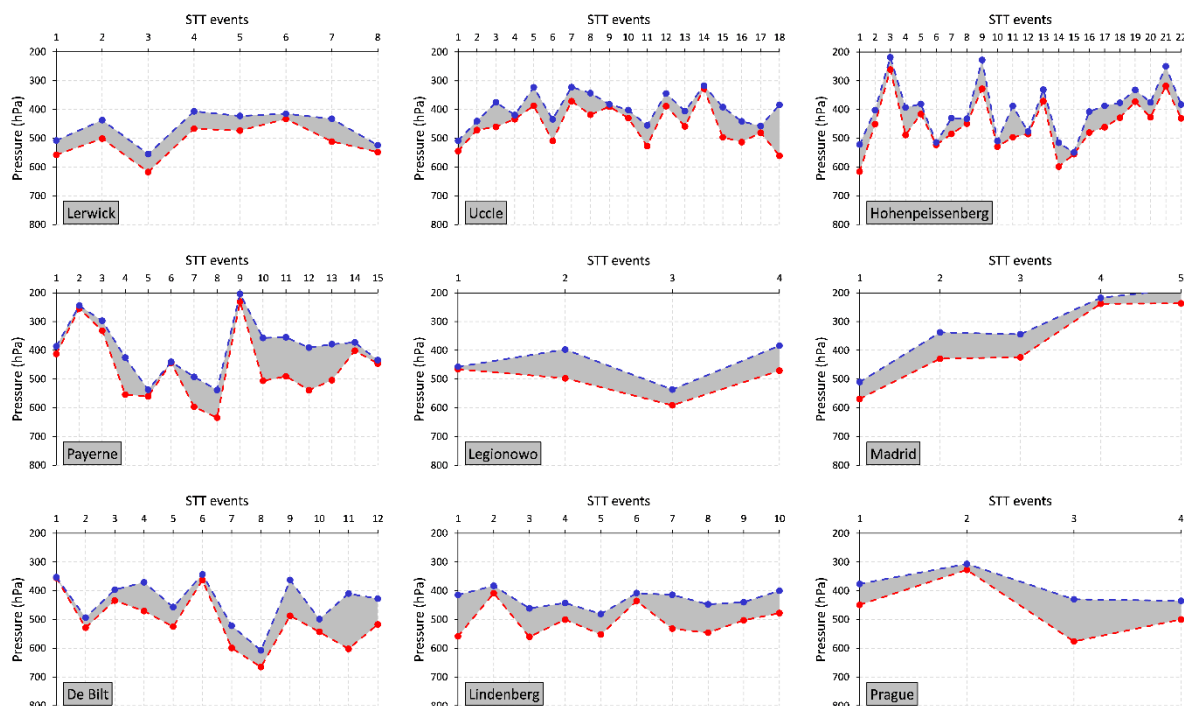


Figure R2. Schematic representation of fold detection.

In more detail, Figure R2 presents a schematic representation of fold detection, also depicting the multiple crossings of the stratosphere in vertical profiles. As mentioned in the manuscript the pressure difference between the middle and upper stratosphere crossing  $\Delta p = p_m - p_u$  indicates the vertical extent of the fold, while the vertical area between the  $p_m$  and  $p_l$  represents the area of the fold that directly affects the ozone profile over the underlying site. Such crossings during the stt events detected at the ozonesonde sites (not shown for IAGOS as several folds might be present during a selected event following the take-off/landing route) are depicted in Figure R3. The



following sentence is now included in the RM (L205-207): “In addition, the individual dynamics and the different vertical location and geometrical characteristics of the selected STT events, especially for observational sites with not so extended number of events, may form somehow unique structures of CAMSRA and observed O3 deviations”.

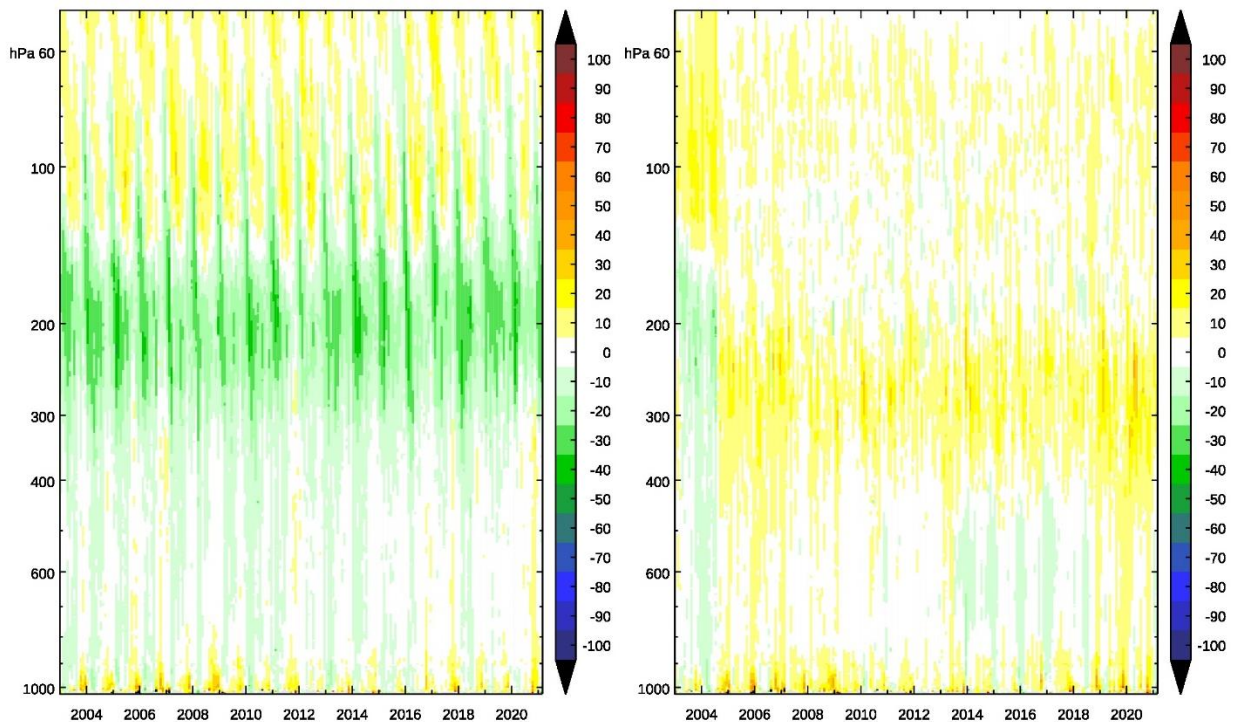


**Figure R3.** Vertical location of middle (blue) and lower (red) crossings of the stratosphere during STT events at the WOUDC ozonesonde sites.

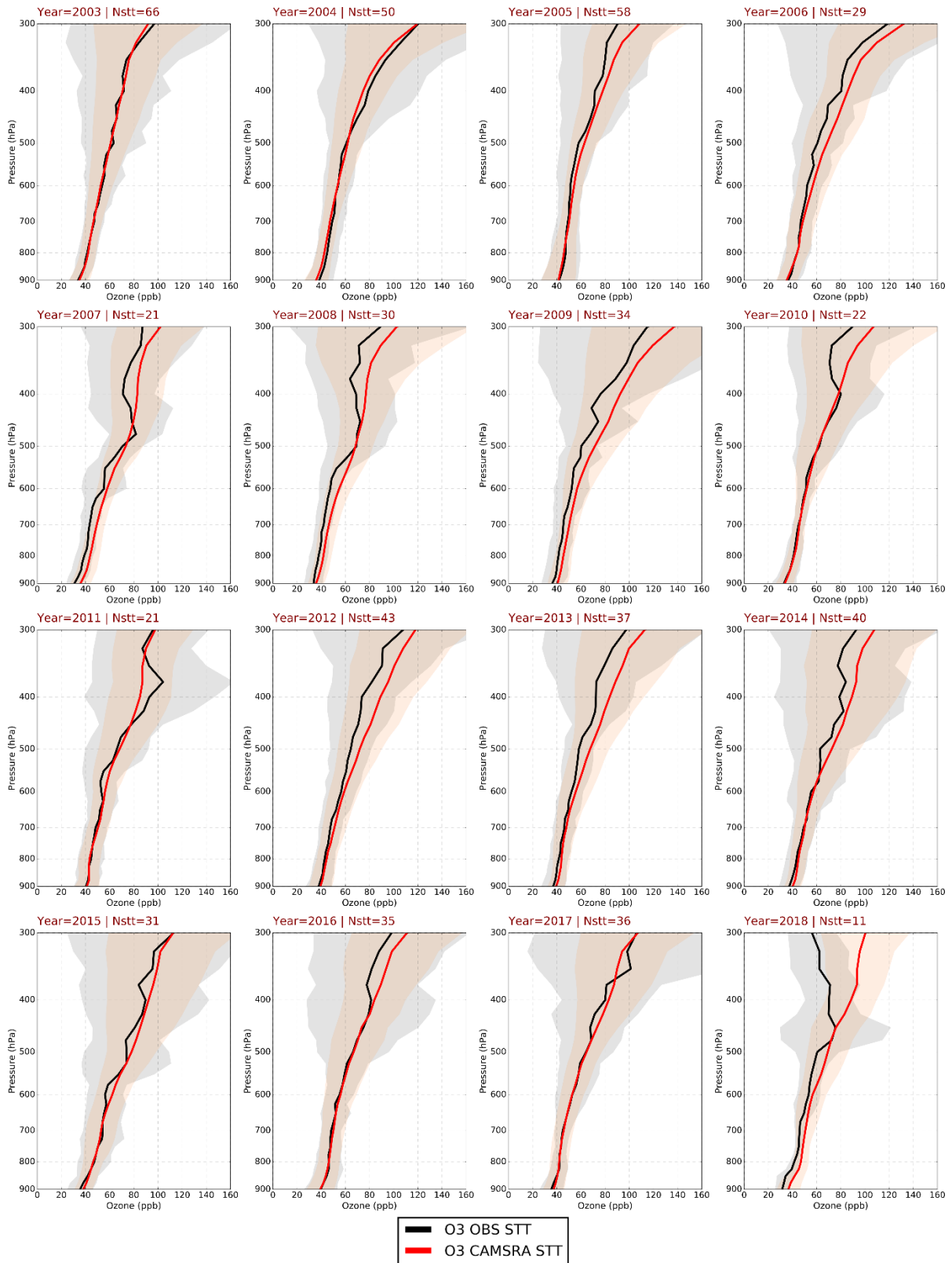
- My previous comment was on the spatial fingerprint of the CAMSRA ozone evaluation. But what about the temporal fingerprint? Is there a temporal evolution in the figures 3 to 10 that is smeared out by considering only the full 2003-2018 period? There might be a temporal component due to a change of the data quality of the observations, change of data assimilation source data, etc. Can you comment if you detected a temporal fingerprint?

We thank the Reviewer for the comment. A potential factor affecting upper tropospheric ozone concentrations in CAMSRA is suggested to be biased assimilated data. In particular, it seems that when Aura (MLS V4 data) start to be assimilated in August 2004 an overestimation of CAMSRA ozone in the upper troposphere arises that is not seen in the control simulation (CAMSRA no DA) (Figure R4). This is also supported if we average all examined sites and split the comparison between CAMSRA and observations by year, for both stt and rest of events, as depicted in Figures R5 and R6, respectively.

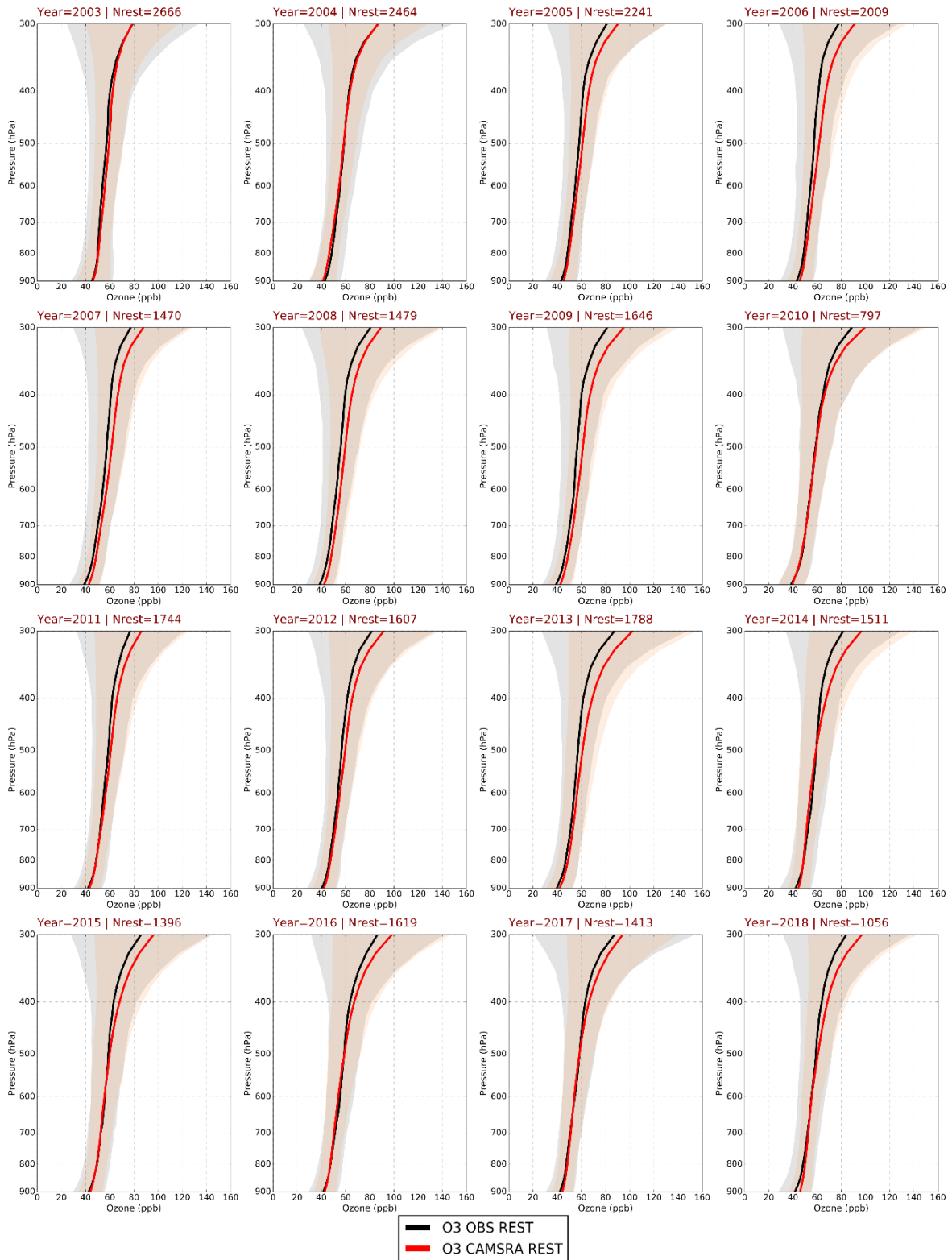
During 2003 and 2004 there is a better agreement (with even a small underestimation) in the upper troposphere between CAMSRA and observations suggesting that the inclusion of Aura data in 2004 is probably a driver of the upper tropospheric ozone overestimation. The following is now included in the RM (L202-205): “**In particular, CAMSRA O3 vertical profiles during both STT and rest of events exhibit a better agreement in the upper troposphere with observations during the years 2003 and 2004, indicating that the inclusion of the Aura data in the assimilation system from August 2004 and on is likely to result in O3 overestimation in the upper troposphere (Figures S2 and S3 in the Supplement).**”.



**Figure R4** Timeseries of CAMSRA no DA (left) and CAMSRA DA (right) O3 relative differences (%) against 13 European ozonesondes (40-61N, 10W-24E).



**Figure R5.** Vertical profiles of observed (black) and CAMSRA (red) ozone concentrations (ppb) averaged over all examined Woudc and IAGOS sites during STT events for each year of the period 2003-2018. The grey and sandybrown shaded areas depict the  $\pm$  one standard deviation of ozone vertical profiles during STT events in observations and CAMSRA, respectively.



**Figure R6.** Vertical profiles of observed (black) and CAMSRA (red) ozone concentrations (ppb) averaged over all examined Woudc and IAGOS sites during the rest of events (no STT) for each year of the period 2003-2018. The grey and sandybrown shaded areas depict the  $\pm$  one standard deviation of ozone vertical profiles during rest of events in observations and CAMSRA, respectively.

- There is no clear explanation given in section 3.3 why chemical data assimilation deteriorates the comparison with the observations above 350 hPa. Only in the conclusions, a list of possible improvements in the data assimilation is given, which might explain the larger inconsistencies between model and observations at those pressure levels.

As also stated in our previous responses sensitivity simulations to isolate the impact of potential drivers of this overestimation are not feasible at the time in the framework of CAMS reanalysis. This deterioration reflects the CAMSRA O3 overestimation in the upper troposphere discussed in previous comments, thus in the RM we have included the following phrase at the end of the last sentence in section 3.3 (L241): “**..reflecting the aforementioned CAMSRA O3 overestimation in the upper troposphere.**”

- Please reformulate the expression “with a bias increase close to O3 increases closes to the upper troposphere”; its meaning is not fully clear to me.

The phrase has been replaced (L237-238) with “**with a bias increase close to O3 enhancements in the upper troposphe**”.

## References

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L.: Stratosphere-troposphere exchange, *Reviews of Geophysics*, 33, 403–439, <https://doi.org/10.1029/95RG02097>, 1995.

Škerlak, B., Sprenger, M., Pfahl, S., Tyrlis, E., and Wernli, H. (2015), Tropopause folds in ERA-Interim: Global climatology and relation to extreme weather events. *J. Geophys. Res. Atmos.*, 120, 4860– 4877. doi: 10.1002/2014JD022787.

Stohl, A., Bonasoni, P., Cristofanelli, P., Collins, W., Feichter, J., Frank, A., Forster, C., Gerasopoulos, E., Gaggeler, H., James, P., Kentarchos, T., Kromp-Kolb, H., Krüger, B., Land, C., Meloan, J., Papayannis, A., Priller, A., Seibert, P., Sprenger, M., Roelofs, G. J., Scheel, H. E., Schnabel, C., Siegmund, P., Tobler, L., Trickl, T., Wernli, H., Wirth, V., Zanis, P., and Zerefos, C.: Stratosphere-troposphere exchange: A review, and what we have learned from STACCATO, *Journal of Geophysical Research: Atmospheres*, 108, <https://doi.org/10.1029/2002JD002490>, 2003.

A. Wagner, Y. Bennouna, A.-M. Blechschmidt, G. Brasseur, S. Chabrilat, Y. Christophe, Q. Errera, H. Eskes, J. Flemming, K. M. Hansen, A. Inness, J. Kapsomenakis, B. Langerock, A. Richter, N. Sudarchikova, V. Thouret, C. Zerefos; Comprehensive evaluation of the Copernicus Atmosphere Monitoring Service (CAMS) reanalysis against independent observations: Reactive gases. *Elementa: Science of the Anthropocene* 21 January 2021; 9 (1): 00171. doi: <https://doi.org/10.1525/elementa.2020.00171>