**Reviewer 1:**

Reviewer #1 (Comments to Author (shown to authors)):

Ort et al. uses airborne observations of O3 and CO from 12 aircraft campaigns to investigate the tropospheric zonal distribution of those two species and their ratio (with an emphasis on the northern hemisphere). The authors identify high O3-CO ratios in the subtropics (23° - 40° latitude) that extend deep into the tropospheric column. In addition to stratosphere-troposphere exchange processes (STE), the authors hypothesize that these high ratios in the subtropics are also due in part to lightning NOx (LNOx) emissions in the upper tropical troposphere that produce more O3 and OH (thus depleting CO) and is subsequently transported to the subtropics via the Hadley circulation. A standard run of a global 3D atmospheric chemistry model (EMAC) is able to reproduce the pattern of high O3-CO ratios in the subtropics. Moreover, the authors perform a sensitivity run using the EMAC model by turning off LNOx emissions and show that the O3-CO ratio is reduced by upwards of 40% in the northern subtropics. A model investigation on the seasonality of the O3-CO ratio is also included. Overall, in addition to STE, the authors show that the chemical composition of the tropospheric subtropics is influenced by LNOx emissions in the tropics.

This manuscript would be of interest to the readership of ACP, and I recommend publication after attention to the following comments:

We thank Reviewer #1 for the time to review our manuscript and the positive feedback and interest on our work.

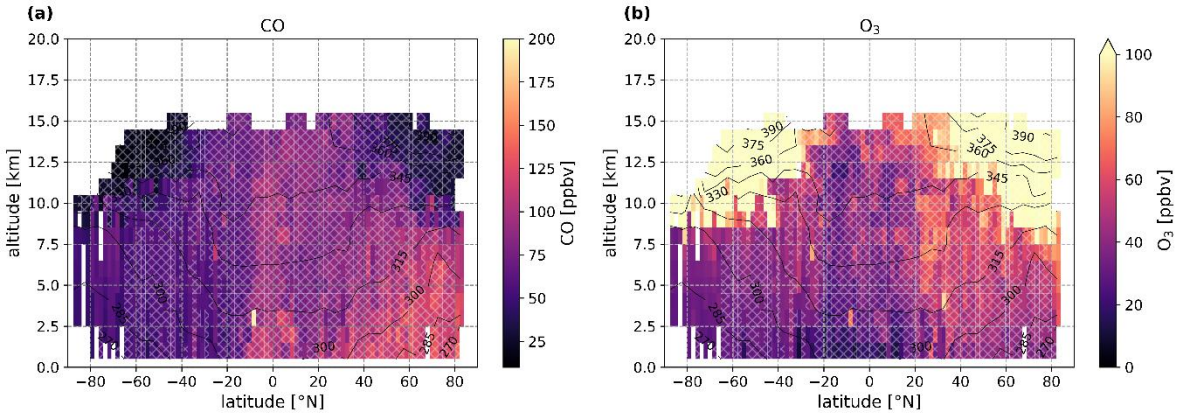
- Lines 125-126: "statistically relevant data can be found in the northern hemisphere": Could the authors define what they mean by "statistically relevant data"? Is there a way to show on Figs 2 and 3 which grid boxes are "statistically relevant" (e.g., using thatching of some sort)?

With “statistically relevant data” we define a sufficient number of airborne observations per grid cell, and that we are within a confidence interval of the variability of the observations. We have clarified this in the manuscript.

Furthermore, to show which grid cells are “statistically relevant”, we have added hatching in Figure 2, which highlights the grid cells, where two conditions are given: 1) enough data points per grid cell (≥10 counts per cell) and 2) good to moderate precision of the means (standard error of the mean relative to the mean in percent: RSEM ≤ 15 %). Keep in mind, that each count already is an average of the minimum time resolution up to 60 s and contains therefore more data points, depending in the time resolution of the individual measurements (see Table 2 in manuscript). We chose a RSEM threshold of ≤15 %, as we have to take seasonal variability also into account.

The hatching of white ‘xxx’ shows us that on the northern hemisphere, we are within the confidence interval and have enough data points per grid cell. On the southern hemisphere, however, this full coverage is not provided, especially not in the subtropics and polar latitudes. Furthermore, we have added a Figure in the Supplement, which shows us those two components (counts per grid cell and RSEM)

separately. Here, we notice that in the southern high latitudes, RSEM is small (<5 %), but the threshold for the counts per grid cell (≥10 counts) is not given.




"Figure 2. Zonal median distribution of CO (a) and $O_3$ (b) from the campaigns listed in Table 1. All airborne measurements have been averaged over 1 km of altitude, and 1.875° of latitude. The scale has been adjusted for $O_3$ to focus on tropospheric values. Black lines indicate layers of constant potential temperature, calculated using Equation 1, and averaged over 5° and 1 km. Grid boxes, which represent averages with good to moderate standard error relative to the mean (≤15 %), and at least ten points per grid cell, are hatched with white "xxx". Note that each count represents an average of 60 s, depending on the time resolution of the measurements."

Lines 124-128: "Overall, the twelve campaigns listed in Table 2 provide a good global coverage of in situ measurements from the boundary layer up to the lower stratosphere. In particular, good coverage can be found in the northern hemisphere, especially in the northern subtropical region. As the data sets have been measured using various time resolutions, all sets were averaged to 60 s before they were combined. Furthermore, the measurements were transferred into the same grid resolution as the modeled data. More statistical information can be found in the Supplement Material."

Lines 176-184: "The grid cell has been hatched with white "xxx" if two conditions are met: 1) at least ten counts per grid cell (counts ≥ 10) and 2) good to moderate precision of the mean (standard error of the mean relative to the mean in percent: RSEM ≤ 15%). Note that each point is already an average to 60 s depending on each time resolution of the individual measurements, and that the precision includes seasonal and atmospheric variability. More details and the zonal distributions of the grid cell counts and the RSEM for each tracer individually can be found in the Supplement Material. Both zonal distributions show a good coverage of the troposphere, despite some data gaps in the polar high latitudes, especially in the southern hemisphere. The observations made in the northern hemisphere are, with a few exceptions, contained within the confidence intervals previously defined (i.e., counts ≥ 10; RSEM ≤ 15%). Conversely, the southern hemisphere exhibits lower coverage within these conditions, particularly in the subtropics and polar latitudes."

Additionally, we have tested with the model, if the extracted flight tracks are comparable to the full modeled data, which is shown in Figure S2 in the Supplement (previously Fig. S1). As the model along the flight tracks and the full climatology do

not differ much, as described in the Supplement in Lines 164-166, we infer that the observations can represent a climatology.

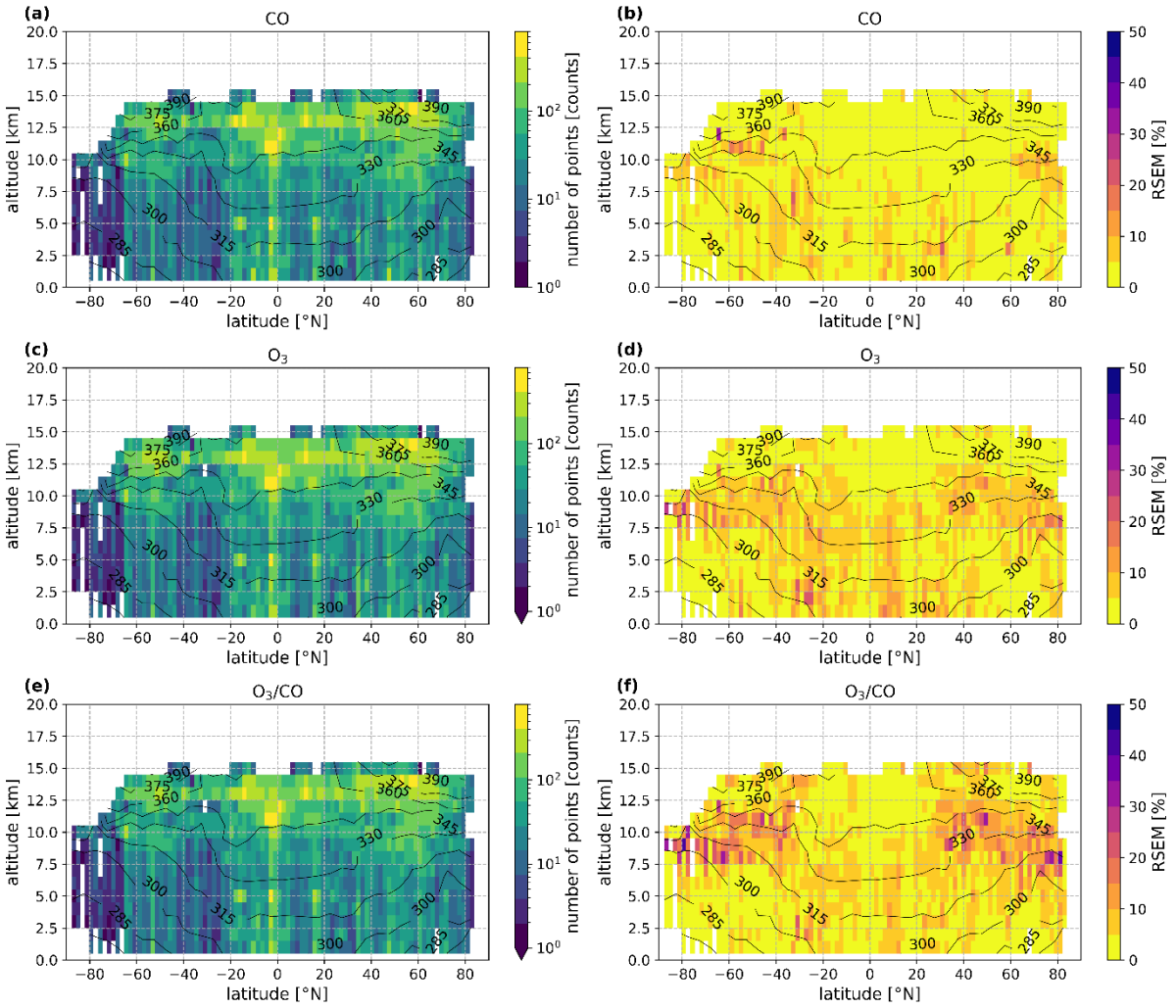
Lines 164-166 in the Supplement: “Taking this into account, there is good agreement between the model-generated climatological data and the flight track extracted data of the model across the troposphere. This confirms that there is sufficient observational data for our analysis, and the selected campaigns are considered statistically representative.”

- Lines 126-127: "all sets were averaged to 60 s before they were combined": Since individual data sets have different uncertainties (Table 2), do some data sets carry more weight than others when combining them or are they all treated equally?

The individual datasets have indeed different uncertainties, which have to be taken into account. Nevertheless, due to methodological differences in uncertainty estimation, no weighting was applied in the averaging, ensuring a consistent and unbiased treatment across data sources.

- Line 128: "sufficient number of data points for each grid box": The authors should add a figure in SI similar to Fig 1 that shows the number of data points used to calculate the value in each grid box. For example, the ~200 ppbv CO in Fig 1(a) seems out of place compared to surrounding grid boxes, and I'm curious if this is due to limited observations biasing that particular grid box. Moreover, what is the standard deviation of the data points in each grid box?

Thank you for this suggestion. We have added the following figure in the Supplement (Fig. S1), showing the number of points per grid cell and the standard error of the mean relative to the mean (RSEM calculated via Eq. S1 in Supplement) for CO, $O_3$, and the $O_3$-CO ratio. We also added a section addressing the measurement statistics in the supplement.




"Figure S1. Zonal distribution of CO (a, b), $O_3$ (c, d), and their ratio (e, f) showing the number of data points per grid cell and the standard error of the mean relative to the mean (RSEM) in percent, calculated via Eq. S1, respectively. Black lines indicate layers of constant potential temperature, averaged over 5° and 1 km. "

Lines 135-158 in the Supplement:

"S2 Measurement statistics

The left panel of Figure S1 (a, c, e) shows the number of data points used to calculate the concentration of CO, $O_3$, and their ratio for each grid box, available from the twelve campaign data sets used in this study. Note that before averaging, each data set had a different time resolution and therefore already represents an average of different time resolutions onto 60 s (see Table 2 in main study). On the right side of Figure S1 (b, d, f), the standard errors of the means relative to the mean (RSEM) in percent are displayed for CO, $O_3$ and their ratio, respectively. The RSEM is given by:

$$\mathrm{RSEM} = \frac{\frac{\sigma}{\sqrt{N}}}{mean} \times 100, \qquad (1)$$

with the standard deviation σ, the number per grid cell N, and the mean of each grid cell. The RSEM provides a variability of the averaged measurements relative to the mean value of each grid cell. Note that this also includes atmospheric and seasonal variability in averaging all campaigns. However, when RSEM is small (< 5%), and counts are high (≥ 10), the observations show very good agreement throughout all data sets. This is the case, for example, in the tropical upper troposphere, where seasonal variability is small. For regions with stronger seasonal variability, e. g., at the tropopause, RSEM is higher (≤ 15%), even if there is a high number of counts per grid cell. Conversely, small RSEM with small counts per grid cell, e. g., in the southern polar latitudes and southern subtropics, show regions with small variability, but averages from only a few flights through this area, which neglects annual variability. Therefore, we decided to categorize the data set using two conditions to show statistical relevant data, with a sufficient number of counts per grid cell (N ≥ 10) and a reasonable variability of the mean values (RSEM ≤ 15%). This has been indicated using hatching in Figure 2 in the main study.

The highest RSEM and therefore the strongest variability between the campaign data sets is found in the vicinity to the tropopause, especially for O3 (Fig. S1, d). We attribute this to the different altitudes of the tropopause throughout the years and the different campaigns. Especially close to the tropopause, the gradient of O3 is very steep, with increasing O3 towards the stratosphere, and therefore leads to such high variability. However, as we are focusing on the troposphere and filter out mixing ratios higher than 100 ppbv of $O_3$, following Prather et al. (2011)), these differences have a negligible effect on our results. In the RSEM of CO, only a few grid points show very large percentages above 15%. Those can be linked to extreme events (e. g., biomass burning), which were measured during single measurement campaigns and increase the standard deviation."

- Line 142: How does overestimated O3 in the model account for reduced NOx from lightning?

By reducing LNOx, O3 formation is significantly reduced. We understand that our phrasing caused confusion. We have rewritten and edited the sentence for clarification.

Line 142: "$O_3$ overestimation by the model is attributed to an overestimation of transport from the stratosphere, related to the limited horizontal grid resolution of the model (1.875° x 1.875°) (Lelieveld et al., 2018). However, by reducing LNOx emissions associated with deep convection, photochemical $O_3$ production declines in the upper troposphere."

- Line 261: "but more pronounced in the northern subtropics": Southern subtropics in the model look just as pronounced as well. Is there a way the authors can quantify the difference instead of giving a qualitative statement? Admittedly, the southern subtropics appears less pronounced in the observational data (Fig 3).

Thank you for this question. We have changed the wording.

Line 270-277: “The model shows a strong enhancement in both hemispheres, while the observations could not capture such ratios in the southern hemisphere, most likely due to the lack of observational data, especially towards the southern high latitudes. In the model, relatively high $O_3$ and low CO values, indicated by high $O_3$ – CO ratios, are found immediately above the boundary layer in both subtropical regions, compared to the values towards the tropics and midlatitudes. However, background $O_3$ and CO ratios differ strongly between both hemispheres, mostly caused by the difference in land cover (natural) and anthropogenic emissions. This results in slightly lower $O_3$-CO ratios in the northern compared to the southern hemisphere, ranging from 0.4 – 1.1 ppbv/ppbv and 0.45 – 1.4 ppbv/ppbv, respectively. Focusing on the northern hemisphere, the model captures well the features of observed zonal $O_3$-CO ratios.”

- Line 275: "smaller CO and higher O3 mixing ratios in the subtropics compared to the northern extra tropics and tropics": Could this statement be conditioned since it's not necessarily true at all altitudes.

Sure, we have clarified this statement.

Line 289-290: “a higher $O_3$-CO ratio in the subtropics compared to the northern extra tropics and tropics, particularly at lower altitudes.”

- Lines 282-283: "the decreasing gradients towards the tropics and mid- to high latitudes are reproduced by the model": Generally agree with the authors but it seems like the gradients are mainly reproduced for 0-6 km and not so much for altitudes higher than 6 km in Fig 6(c) (particularly for the higher latitudes). Why is this the case in the model?

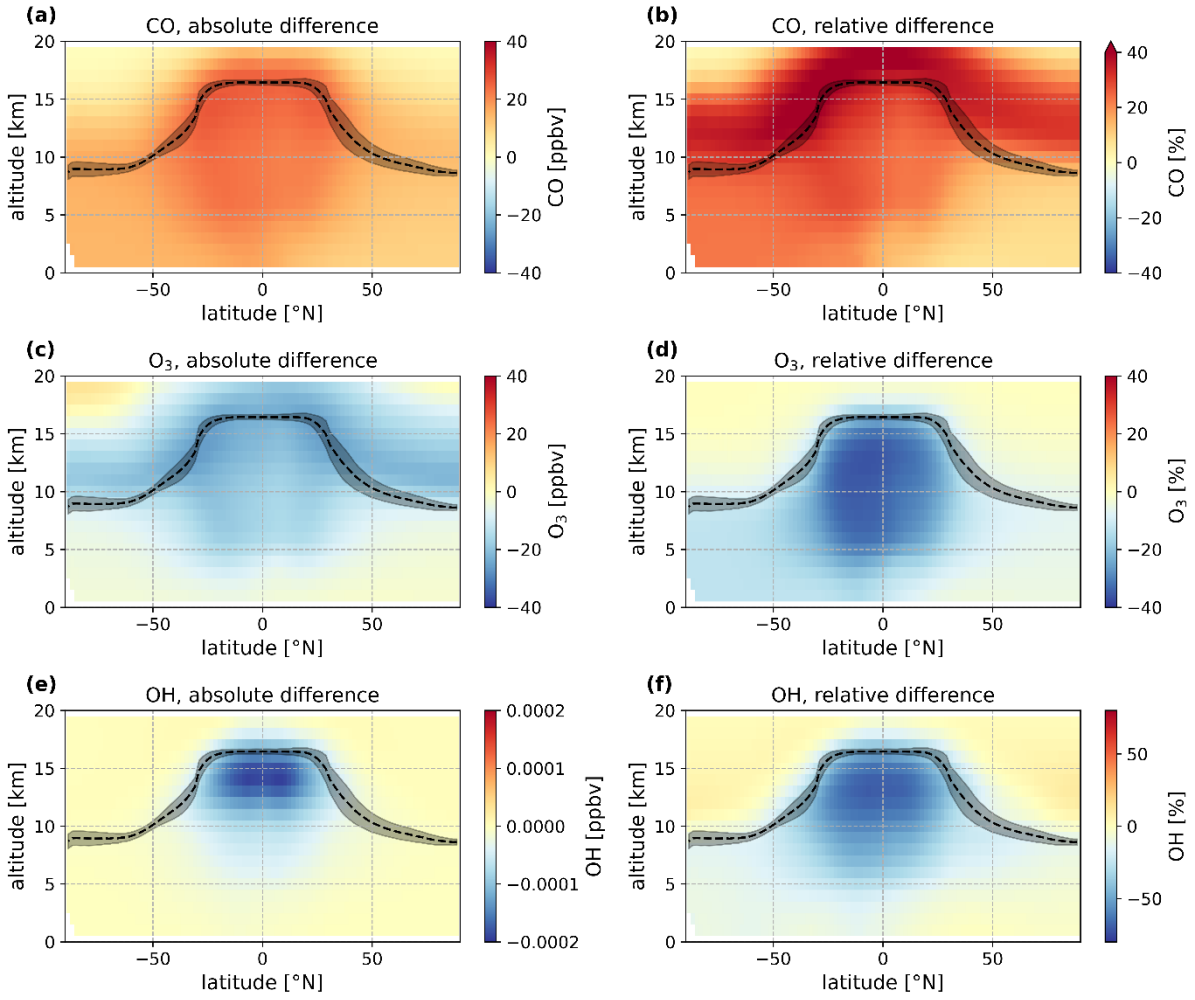
As we clarified in Line 142, the larger difference of the model towards higher altitudes is mostly due to the limited spatial resolution of the model. Small-scale mixing, tropopause folding events, and the exact position of the tropopause cannot be well simulated by the model (see for example Lelieveld et al., 2016 and Joeckel et al., 2016 for a detailed analysis of simulated ozone by the EMAC model).

*Lelieveld, J., Gromov, S., Pozzer, A., and Taraborrelli, D.: Global tropospheric hydroxyl distribution, budget and reactivity, Atmospheric Chemistry and Physics, 16, 12 477–12 493, https://doi.org/10.5194/acp-16-12477-2016, 2016.*

*Jöckel, P., Tost, H., Pozzer, A., Kunze, M., Kirner, O., Brenninkmeijer, C. A. M., Brinkop, S., Cai, D. S., Dyroff, C., Eckstein, J., Frank, F., Garny, H., Gottschaldt, K.-D., Graf, P., Grewe, V., Kerkweg, A., Kern, B., Matthes, S., Mertens, M., Meul, S., Neumaier, M., Nützel, M., Oberländer-Hayn, S., Ruhnke, R., Runde, T., Sander, R., Scharffe, D., and Zahn, A.: Earth system chemistry integrated modelling (ESCiMo) with the modular earth submodel system (MESSy) version 2.51, Geoscientific Model Development, 9, 1153–1200, https://doi.org/10.5194/gmd-9-1153-2016, 2016.*

- Line 293: "leading to higher OH concentrations in the convective outflow. This potentially shortens the lifetime of CO and other hydrocarbons": It would be helpful if the authors added a brief discussion on how much LNOx emissions is impacting the atmospheric oxidation capacity of the northern subtropical troposphere. Perhaps show the difference in OH that results from excluding LNOx in Fig 7? Also, can the authors calculate the change in lifetime for CO and a few other hydrocarbons after excluding LNOx emissions?

From Lelieveld et al., 2016, 2018, who have investigated the impact of LNOx onto OH mixing ratios with the EMAC model, we know that OH is strongly dependent of NOx and particularly on LNOx in the upper tropical troposphere where the recycling of OH from $NO+HO_2$ is the predominant OH source. Excluding LNOx from the simulations cause a decrease of OH concentrations in the upper troposphere by a factor of two to three (Lelieveld et al., 2018; Labrador et al., 2004).




"Figure 7. Relative and absolute differences of CO (a, b), $O_3$ (c, d), and OH (e, f) caused by excluding LNOx emissions in the global 3D model EMAC, shown on annual averaged zonal mean distributions. On the left side (a, c, e) the absolute differences (Eq. 2) of the gas concentration are shown and on the right side (b, d, f) the relative differences (Eq. 3) in percentages, normalized to the background concentrations. Note the scales for OH differ from CO and $O_3$. The black dashed

lines and shaded area indicate the mean height of the tropopause layer and its standard deviation, respectively, by taking the zonal and annual average."

This confirms an estimate we performed for the tropical upper troposphere, where we calculated a decrease in OH concentrations by a factor of 2.4 by excluding LNOx emissions. This changes the chemical lifetime of CO from approximately two months (67 days) to five months (161 days) in that region, taking only the major atmospheric sink through OH into account. Considering the relatively long lifetime of NOx in the upper troposphere of approx. 4-7 days, and meridional transport of tropical air masses along the Hadley circulation, OH recycling by NOx extends outside the tropics. Therefore, calculating the chemical CO lifetime for the northern subtropical troposphere, we achieve a change from 56 days to approximately 70 days, when LNOx is excluded from the simulation. Hence, by the absence of LNOx, CO lifetimes would increase by a factor of 1.24 in the northern subtropical troposphere.

We agree with the suggestion of Reviewer #1 to include the absolute and relative differences of OH in Figure 7 in the manuscript and have accordingly changed and added Section 3.3.1 (see manuscript).

*Lelieveld, J., Gromov, S., Pozzer, A., and Taraborrelli, D.: Global tropospheric hydroxyl distribution, budget and reactivity, Atmospheric Chemistry and Physics, 16, 12 477–12 493, https://doi.org/10.5194/acp-16-12477-2016, 2016.*

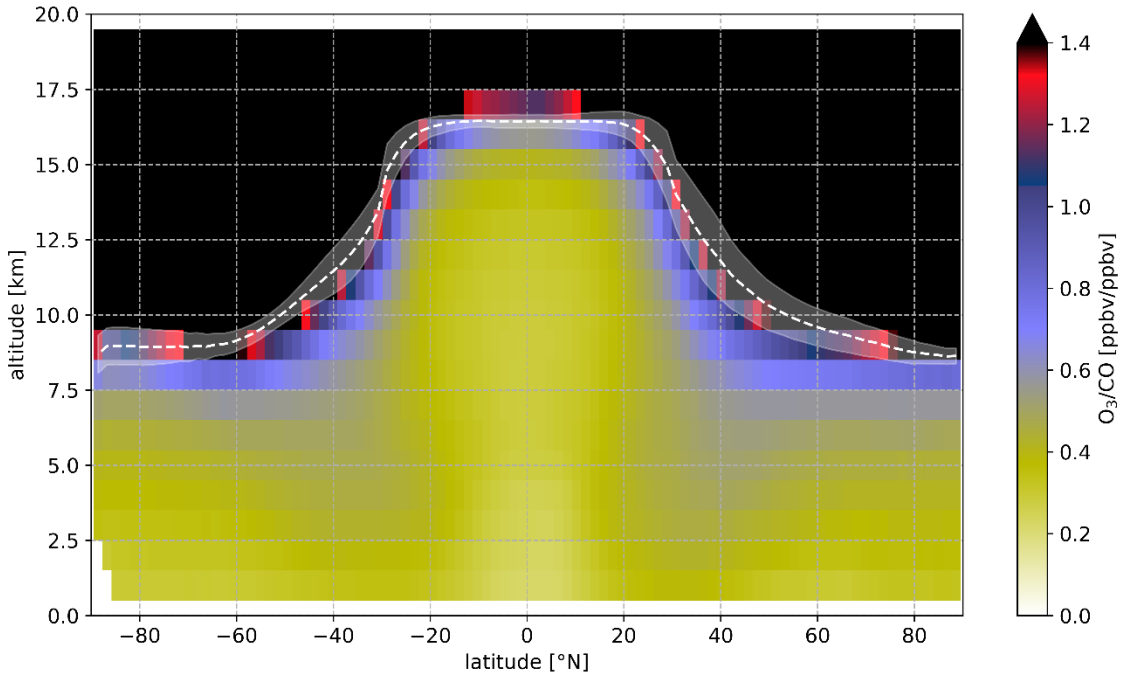
*Lelieveld, J., Bourtsoukidis, E., Brühl, C., Fischer, H., Fuchs, H., Harder, H., Hofzumahaus, A., Holland, F., Marno, D., Neumaier, M., Pozzer, A., Schlager, H., Williams, J., Zahn, A., and Ziereis, H.: The South Asian monsoon—pollution pump and purifier, Science, 361, 270–273, https://doi.org/10.1126/science.aar2501, 2018.*

*Labrador, L. J., R. von Kuhlmann, and M. G. Lawrence (2004), Strong sensitivity of the global mean OH concentration and the tropospheric oxidizing efficiency to the source of NOx from lightning, Geophys. Res. Lett., 31, L06102, doi:10.1029/2003GL019229.*

- Line 357: "the local high of the O3-CO ratio in the subtropics is still present, albeit weaker": Could the authors make a figure similar to Fig 5 but for the EMAC model run excluding LNOx and place it in the SI? This would help to show the local high of the O3-CO ratio in the subtropics is still present albeit weaker in a quick way. Moreover, it would be helpful when readers later see seasonal differences in Fig 11.

We have included the zonal distribution of the $O_3$-CO ratio from the sensitivity run, excluding lightning NOx, in the Supplement as Figure S4 and added a sentence in the manuscript referring to it.

Line 321-322: "A zonal distribution of the annual average $O_3$-CO ratio from the sensitivity simulation can be found in the Supplement."




"Figure S4. The median annual zonal distribution of the sensitivity simulation, excluding LNOx, for the ratio between $O_3$ and CO. All data points are averaged over 1 km altitude and 1.875° latitude. The scale has been adjusted to display tropospheric values. The white dashed line indicates the tropopause height, and the standard deviation is shaded in white."

- Line 400: "the O3-CO ratio is still shifted to higher values by excluding LNOx": Please double-check statement. It seems like ratio is shifted to lower values when excluding LNOx in Fig 11.

We have corrected this mistake.

Line 425-426: "the $O_3$-CO ratio is still shifted towards lower values by excluding LNOx"

- Lines 423-426: What other factors are at play for O3 since even though O3 depletion is strongest during the summer, the O3 mixing ratio from May-Aug remains relatively high compared to the rest of the year.

Indeed, $O_3$ mixing ratios remain relatively high during the summer months. This correlates with the high BG, which we have defined in the study and shown in Fig. 12 as every other $O_3$ source except for lightning NOx or the stratosphere. This includes other NOx sources, which can produce $O_3$ (e. g., anthropogenic emissions, PAN from long-range transported biomass burning).

- Lines 436-444: "However, storm tracks... stronger isentropic STE": This section needs clarification. Seems like the main message is that the STE influence on O3

has a minor effect in summer over the northern subtropics (Lines 445-446), so what is the main message you're trying to convey in Lines 436-444?

With this section, we aim to clarify that the influence of STE on tropospheric $O_3$ seems minor in the subtropics despite the significant dynamical mixing processes, especially in late spring and early summer (Rossby wave-breaking, tropopause folds). We have changed the section slightly to make this statement clearer.

Lines 459-471: “The contribution of STE to the total tropospheric column $O_3$ budget in the northern subtropics reaches about 25 – 50 %, peaking in February and has a minimum in August. The STE influence is strongest close to the tropopause and diminishes towards the surface. Previous analyses showed that storm tracks and mountain chains in the subtropics are preferred regions for STE (Škerlak et al., 2014), being strongest in late spring-early summer, particularly over continents (Sprenger and Wernli, 2003). STE is strongly affected by the intensity of the jet streams, due to their role in the formation of tropopause folds, and the height of the tropopause. The location and strength of jet streams are strongly influenced by temperature differences between high and low latitudes. Jing and Banerjee (2018) found that anticyclonic wave-breaking events become more frequent in summer, and can be associated with stronger isentropic STE. Overall, STE seems to have a strong impact on the mixing ratios in the UTLS region; however, averaging over the whole troposphere, STE is found to have a minor effect on the seasonality driving the $O_3$ – CO ratio in the northern subtropics. In a climatological study based on ERA-Interim reanalysis data, Škerlak et al. (2014) showed that STE influence on $O_3$ can be found down to the planetary boundary layer throughout all seasons with global hot spots along mountain chains mainly in the subtropical summer. However, they identified an accumulation of $O_3$ below the tropopause, rather than originating directly from the stratosphere. Similar seasonality can be found in the southern hemisphere (not shown here).”

- Lines 458-462: Please clarify the main takeaway from this discussion on convection. Lines 458-459 imply that it is not a major factor given Fig 12(a), but then Lines 459-462 suggest otherwise.

We understand that this discussion seemed a bit confusing. We have rewritten this part.

Lines 478-485: “Strong convective activity transports surface emissions upward, with impacts depending heavily on the inflow region. In highly polluted regions with strong NOx emissions (e.g., anthropogenic, biomass burning), enhanced $O_3$-CO ratios can be transported upwards by convection (e.g., Asian Summer Monsoon, see Mickley et al. (2001)). While in remote areas, especially over the oceans and in the tropics, low $O_3$ is common at the surface through the strong sink of photodissociation and the reaction of $O(^1D)$ with water vapor, which leads to rather low $O_3$-CO ratios. Therefore, seasonal convective activity supports higher $O_3$-CO ratios in the tropospheric column, notably in summer (also noticed as a slight increase of $O_3$ from LNOx during summer in Fig. 12 (b)). Furthermore, deep convection is also linked to stratospheric–tropospheric mixing through overshoots and corresponding downdrafts (Frey et al., 2015), additionally mixing high $O_3$ and low CO air masses into the upper troposphere.”

- Line 472: "consistently influences the O3 mixing ratio": Please restate the magnitude of that influence on the O3 mixing ratio in the northern subtropics from tropical LNOx.

We changed the wording.

Lines 494-496: “Tropical LNOx, meridionally transported via the Hadley circulation, influences the $O_3$ mixing ratio in the northern subtropics, and possibly also the depletion of CO, throughout the year.”

Technical Corrections:

We have implemented the technical corrections in our manuscript.

- Line 88: Remove apostrophe in aircraft's

- Line 100: Remove the word "are"

- Line 136: Correct typo: miscirculationaneous

- Figure 2 caption: Should "white lines" be changed to "black lines"?

- Line 207, now 216: Fix typo in sentence: "Here, ppbv values a the low hundreds could already be observed."

- Line 234, now 244: "pols" should be "poles"

- Line 387, now 411: "spacial" should be "spatial"

- Line 393, now 417: "Below, all monthly medians are shown..." Please mention the figure that this sentence is referring to.

- Line 409, now 434: "lighting" should be "lightning"

- Line 417, now 442: Remove comma after effects

- Line 418, now 443: "depend on other mechanism" should be "depend on another mechanism"

- Line 503, now 525: "investigated" should be "investigations"

- Throughout manuscript: Fix end quotes (e.g., Line 157: "REF" and "without LNOx")