Review 1

The paper gives a comprehensive update on the comparison of long-term ozone data from balloon-borne ozone sondes and IAGOS measurements on commercial aircraft. As in previous studies, the authors find that ECC ozone sondes give 5 to 10% higher ozone values in the free troposphere (3 to 8 km). Brewer-Mast and Carbon-Iodine sondes give slightly lower ozone, 0 to 5% less, in the free troposphere. Monthly means from all three sonde types show high correlations, more than 0.7, with IAGOS monthly means. Indian ozone sondes usually give 25 to 35% higher ozone than IAGOS, with generally poor correlation, around 0.4. The authors find little to no dependence on season or distance between sonde station and IAGOS airport. Overall this is a well written paper which is relevant and should be published in ACP.

Response: We appreciate the time and effort that the editor and the reviewers dedicated to providing feedback on our manuscript. We are grateful for the insightful and valuable comments on our paper. We have incorporated most of the suggestions made by the reviewers. Those changes are highlighted in the revised manuscript. According to your suggestion, we modified the manuscript in detail and marked the revised contents with red font.

I do have a number of generally minor suggestions:

For several parts of the paper, I would prefer a clearer separation between three altitude regions and would like to see more specific results for these altitude regions. In many places, e.g. Fig. 3, there is much better agreement for the 3 to 8 km region, less agreement for altitudes below 3 km and above 8 km. The region below 3 km has a lot of local ozone sources and sinks (cities, airports, rural environment, ...), while the region above 8 km is influenced quite significantly be stratosphere-tropopsphere exchange, jet-streams, tropopause folds, .. Separating results for these regions would provide a clearer picture of ozone-sonde and IAGOS differences, as well as the limitations of the current comparison.

Response: Thanks for your suggestions. We added: " **Figure S1**. Bias, correlation coefficient (R), and RMSE for four types of ozonesonde and aircraft observations in different altitudes."

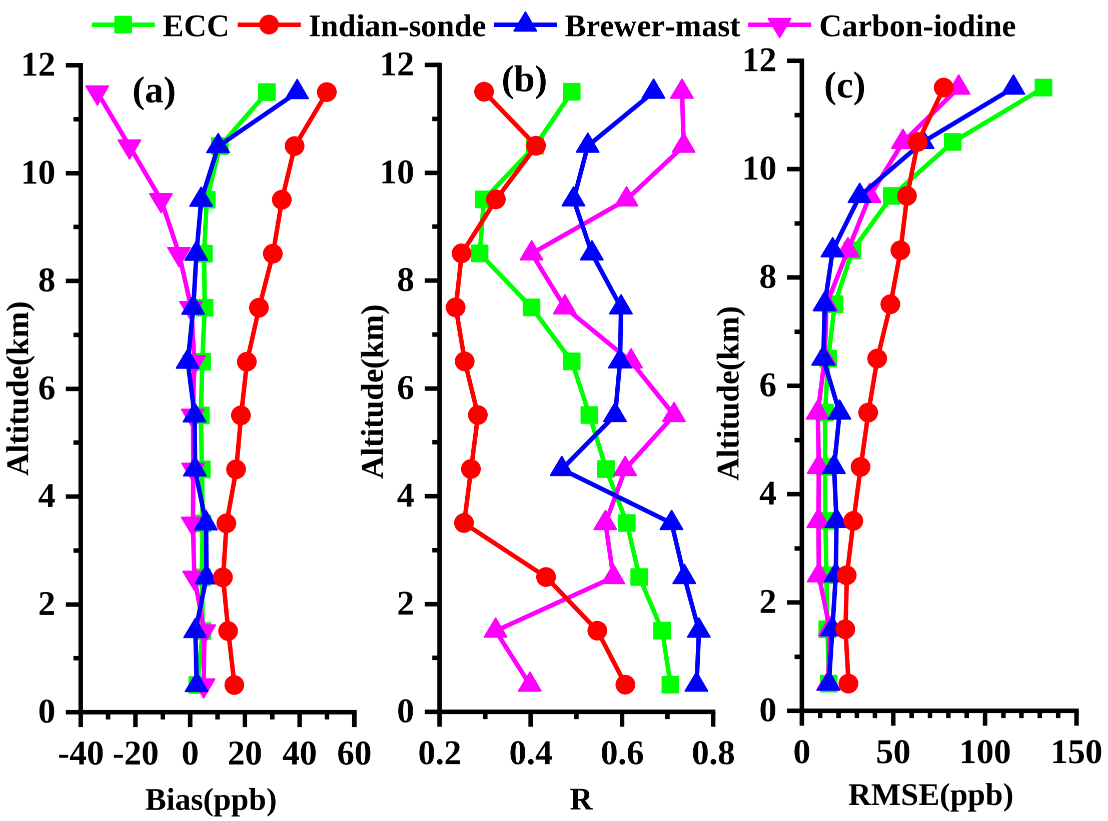


Figure S1. Bias, correlation coefficient (R), and RMSE for four types of ozonesonde and aircraft observations in different altitudes.

And added in the lines 201-224:

The region below 3 km has many local ozone sources and sinks (cities, airports, rural environment, etc). In comparison, the region above 8 km is significantly influenced by stratosphere-troposphere exchange, jet streams, and tropopause folds. Fig. S1 shows that the R between ozonesondes and aircraft observations is higher near the ground (< 2 km) and at high altitudes (> 10 km). This shows that although the influencing factors of O_3 near the ground and at high altitudes are more complex, their long-term temporal variation characteristics are similar. The influences of cities, airports, rural environment, stratosphere-troposphere exchange, jet streams, tropopause folds, etc., have a more significant impact on the concentration of O_3 in the short term.

The correlation between four types of ozonesondes and aircraft observations also varies with altitude (Fig. S1). From 0-8 km, the correlation between ECC and aircraft observations decreases with altitude, with R being 0.71 at 0-1 km and reaching a minimum of 0.29 at 8-9 km; from 8-12 km, R increases with altitude, reaching 0.49 at 11-12 km. The correlation between the other three ozonesondes (Brewer-mast, Indian-sonde and Carbon-iodine) and the aircraft observations all vary with altitude, with different inflection points. The number of stations for these three types of ozonesondes is small (Table 1). Therefore, local variable influences on O_3 are more important, so R varies more with altitude.

The bias and RMSE with respect to the aircraft observations of the four types of ozonesondes at 8-12 km are higher than that at other altitudes. In contrast, the bias and RMSE values below 8 km are smaller and vary less with altitude, consistent with the vertical distribution characteristics of O_3 concentration in Fig. 2. This is likely due to the higher concentration of O_3 and the typically larger difference in spatial distance between ozonesonde and aircraft observations at 8-12km.

In addition, the bias and RMSE relative to the aircraft observations at different altitudes for ECC, Carbon-iodine and Brewer-mast sondes are lower than those for the Indian-sonde, which is similar to the results of the above analysis of O_3 concentration.

And in conclusions, we added "Ozonesondes and aircraft observations have smaller R in the middle troposphere, but larger bias and RMSE in the upper troposphere. The bias and RMSE relative to the aircraft observations at different altitudes for ECC, Carbon-iodine and Brewer-mast sondes are lower than those for the Indian-sonde."

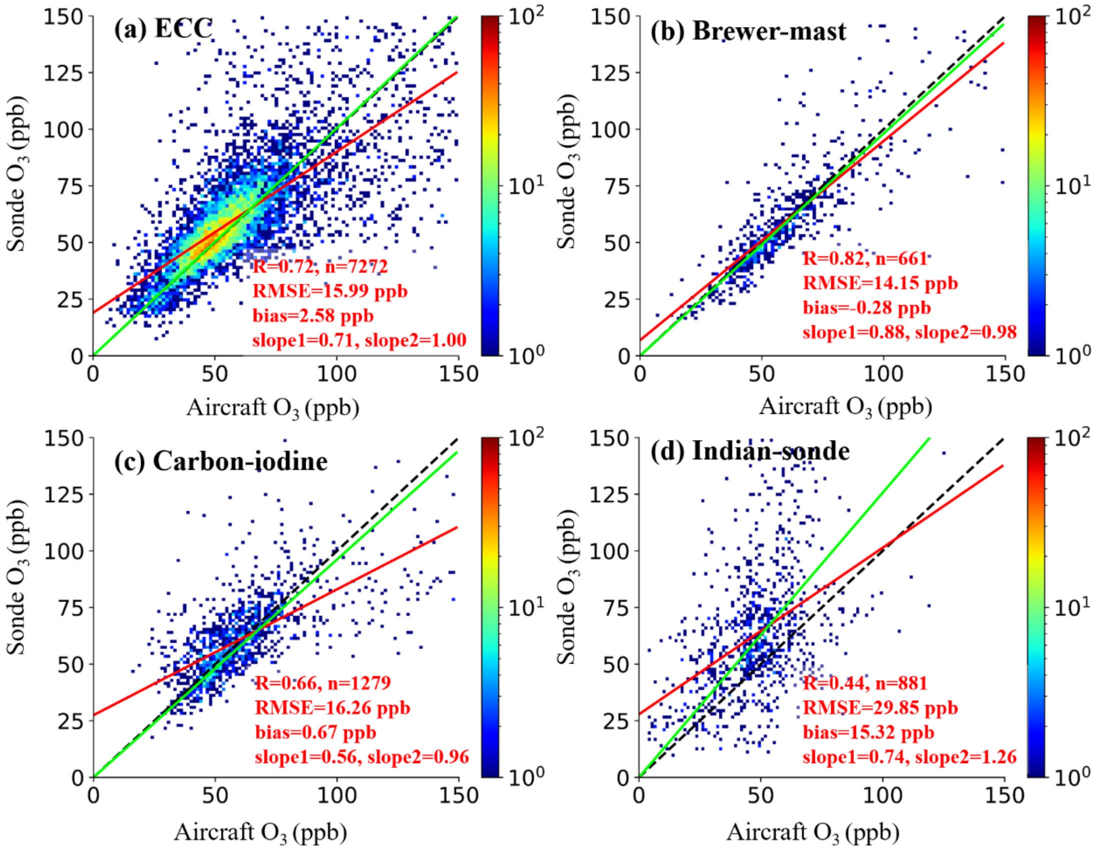
Table 1: station - airport distance is missing in most cases. Should be given, preferably also in kilometers. Also: minus sign before longitudes in the western hemisphere ended up in a different line. Please fix.

Response: Thanks for your suggestions. We have modified it and added the station-airport distance (km) in Table 1.

Figure 3: please state in the caption that this is based on monthly means for both sondes and IAGOS. Also: It would be better to see the relative frequency of the data points, e.g. using a false color representation. As it is now, the plot tends to emphasize the more outlying data points, and one cannot see where most of the datapoints lie. I assume close to the fitted red lines. Also: I find it confusing that the fitted lines are in red, while the text describing the fits is in blue and in the same color as the 1:1 line. Please make these colors consistent. Finally: In Fig. 2d, the Indian sonde data are clearly higher than the IAGOS data. In contrast, in Fig. 3d, the fitted line is below the 1:1 line, which would indicate that the sondes give lower ozone, in contradiction to Fig. 2d. I think it would

be better to not force the fitted lines through zero, but fit slope and offset. An offset could indicate potential causes for systematic differences, e.g., high background current in the sonde data.

Response: Thanks for your suggestions. We have redrawn Figure 3. We found that the force the intercept to zero for the regressions and not force the intercept to zero for the regressions are quite different, just as the reviewer said, “not force the fitted lines through zero, but fit slope and offset. An offset could indicate potential causes for systematic differences, e.g., high background current in the sonde data.” But in generally, when O_3 is zero both the ozonesondes and the aircraft will measure zero. Therefore, we only give the force the intercept to zero for the regressions in the article, but we give the slope and offset of the not force the intercept to zero for the regressions in the main text analysis.



We added in Line 178-185: " After calculation, we obtained the slopes and offsets of ECC, Brewer-mast, Carbon-iodine and Indian-sonde without forcing the fitted lines through zero, the slope is 0.71, 0.88, 0.56 and 0.74, respectively, and the offset is 18.94 ppb, 6.89ppb, 27.48ppb and 27.84ppb. When we force the intercept to zero for the regressions, the slope is larger than the slope without forcing the fitted lines through zero (fig. 3). In generally, when O_3 is zero both the ozonesondes and the aircraft will measure zero. However, there is an offset in the fit of the two data sets due to potential causes for systematic differences during the observation measurement process, e.g., high background current in the sonde data. "

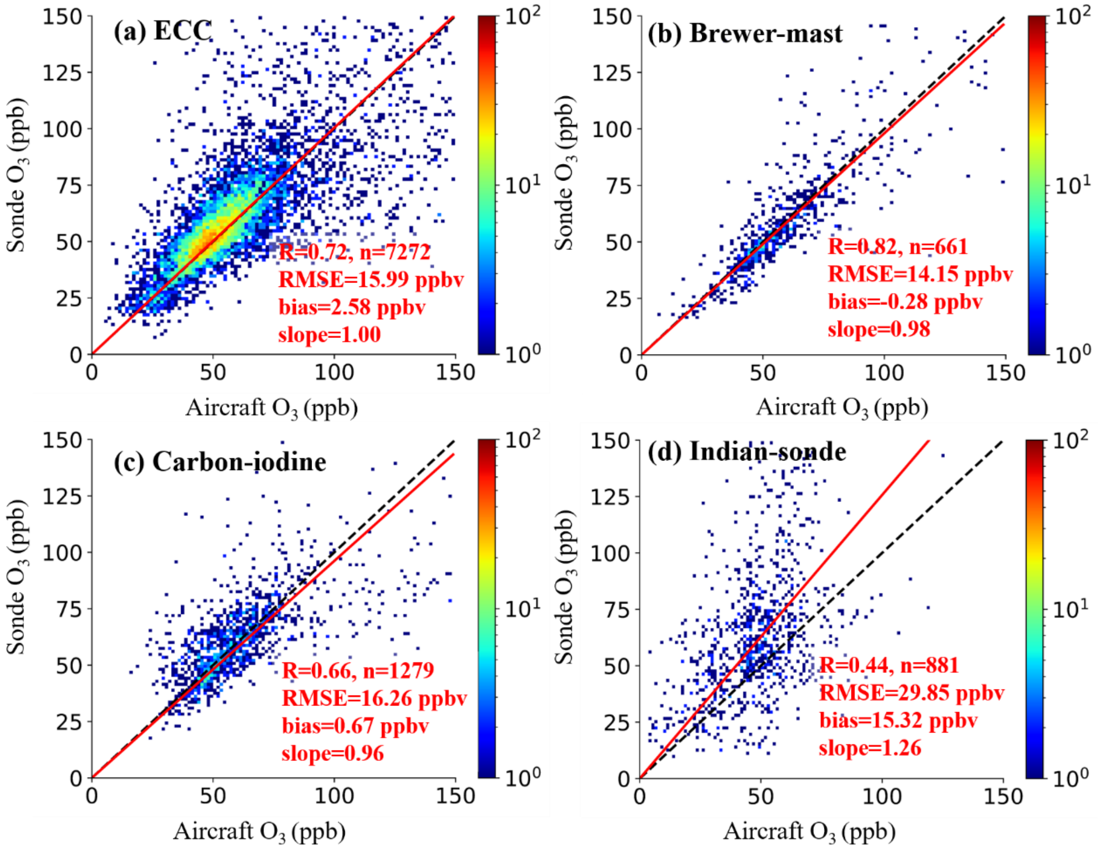


Figure 3. Correlation (*R*) of monthly mean ozone mixing ratios between ozonesonde and aircraft measurements. While IAGOS does measure in the lower stratosphere these values are usually far from the airport, so the sonde-aircraft distance will be large, we only plots data below 150 ppb. The black dashed line shows the 1:1 axis, the red line shows the linear fit (with the intercept set to 0), the color bar shows the data counts. Correlations are significant at the 99% level (p <0.01). N denotes the number of data points, R is the correlation coefficient, Bias is the overall average difference in monthly mean values [Ozonesonde ozone – Aircraft ozone, in ppb], RMSE is the root mean square error, slope is the slope of the linear fit line. All data points are based on the monthly mean.

Line 166: I would not call a ~45ppb RMSE "small". Tropospheric ozone mixing ratios are around 50 ppb, so 45 ppb RMSE corresponds to around 100% uncertainty. That is hardly "small". Please correct.

Response: Thanks for your suggestions. We modified into " The RMSE of O_3 observed with the four types of ozonesondes (ECC, Brewer-Mast, Carbon-Iodine and Indian-sonde) and the aircraft is 15.99 ppb, 14.15 ppb, 16.26 ppb and 29.85 ppb, respectively."

Figure 4: Please add zero line for easier reference. Also, Fig 3 and Fig 4 bring up the question how R, slope, offset and RMSE behave for different altitudes. I think this should be considered, and additional plots should be shown and discussed. I have a feeling that agreement would be best near 5 km, and would deteriorate significantly below 3 km and above 9 km.

Response: Thanks for your suggestions. We have added the zero line in Figure 4. We have added "Figure S1 Bias, correlation coefficient (R), and RMSE for four types of ozonesonde and aircraft observations in different altitudes" and performed the analysis in the article. Please see the response to comment 1 for details.

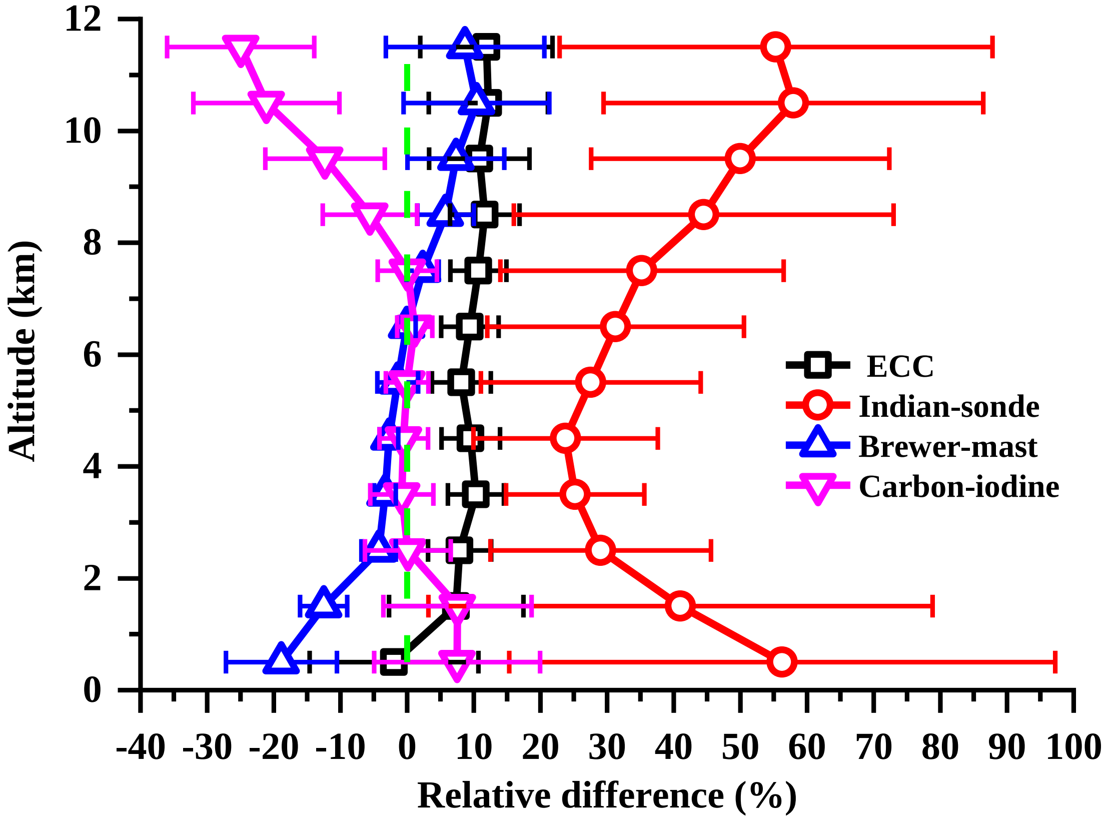


Figure 4. Mean relative difference (RD) between the ozonesonde O_3 and aircraft O_3 data. RD is calculated from ($O_{3\text{-ozonesonde}}$ - $O_{3\text{-aircraft}}$)/ $O_{3\text{-aircraft}}$×100%. The green dashed line is the zero line.

We also plotted the fitting figures for 0-3 km, 3-8 km and 8-11 km, but we found that the R value at 0-3 km is relatively high. This indicates that although near-ground O_3 is greatly affected by the underlying surface and anthropogenic sources, this influence exists for both ozonesondes and aircraft, that is, the influence on these two observation datasets is consistent, so when comparing their temporal correlation, the influence is not significant.

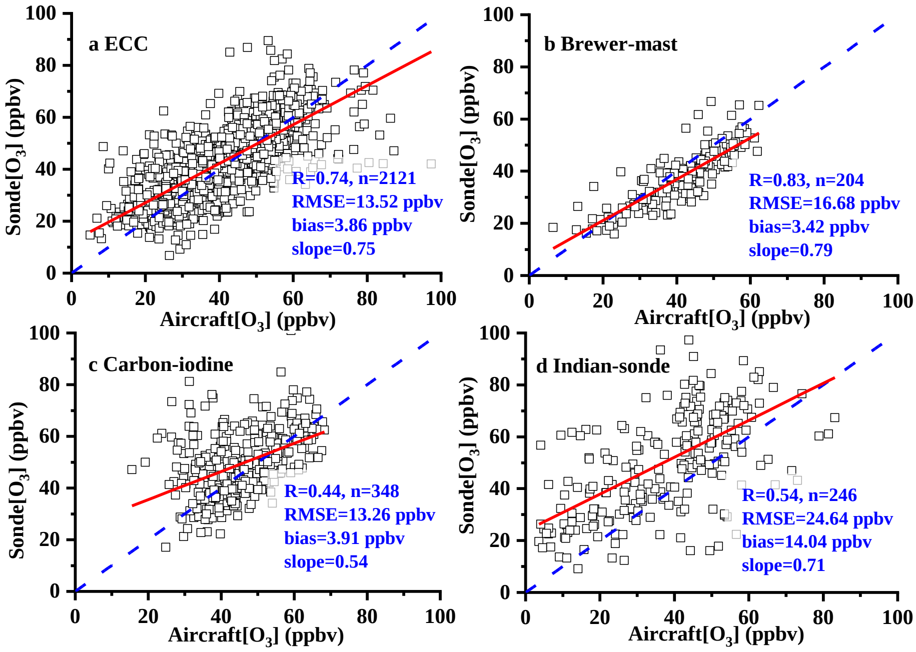


Figure 1 Correlation (*R*) of monthly mean ozone mixing ratios between ozonesonde and aircraft measurements at 0-3 km. The blue dashed line shows the 1:1 axis and the red line shows the linear fit. Correlations are significant at the 99% level (p <0.01). *N* denotes the number of data points, and RMSE is the root mean square error.

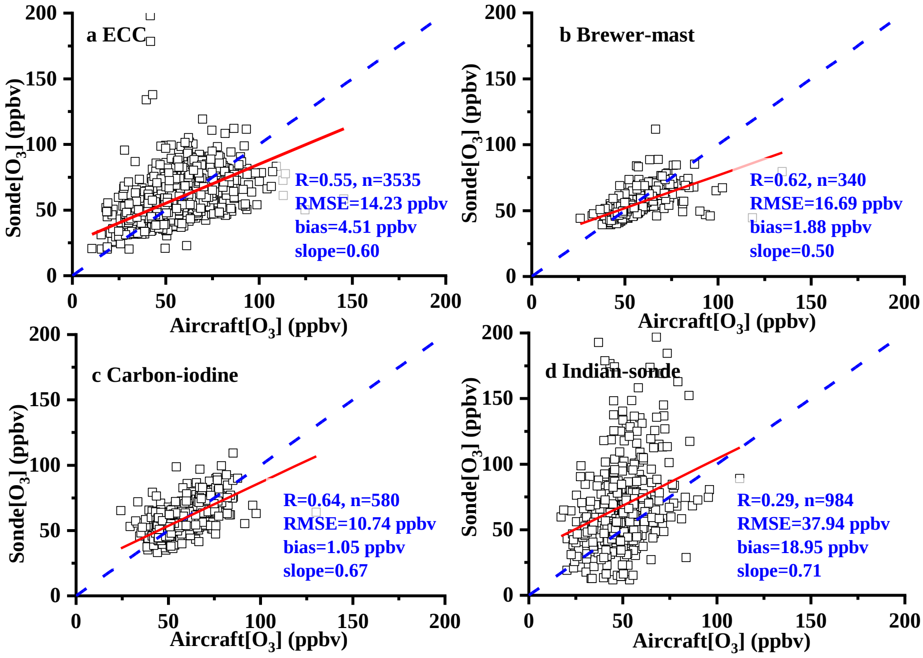


Figure 2 Correlation (*R*) of monthly mean ozone mixing ratios between ozonesonde and aircraft measurements at 3-8 km. The blue dashed line shows the 1:1 axis and the red line shows the linear fit. Correlations are significant at the 99% level (p <0.01). *N* denotes the number of data points, and RMSE is the root mean square error.

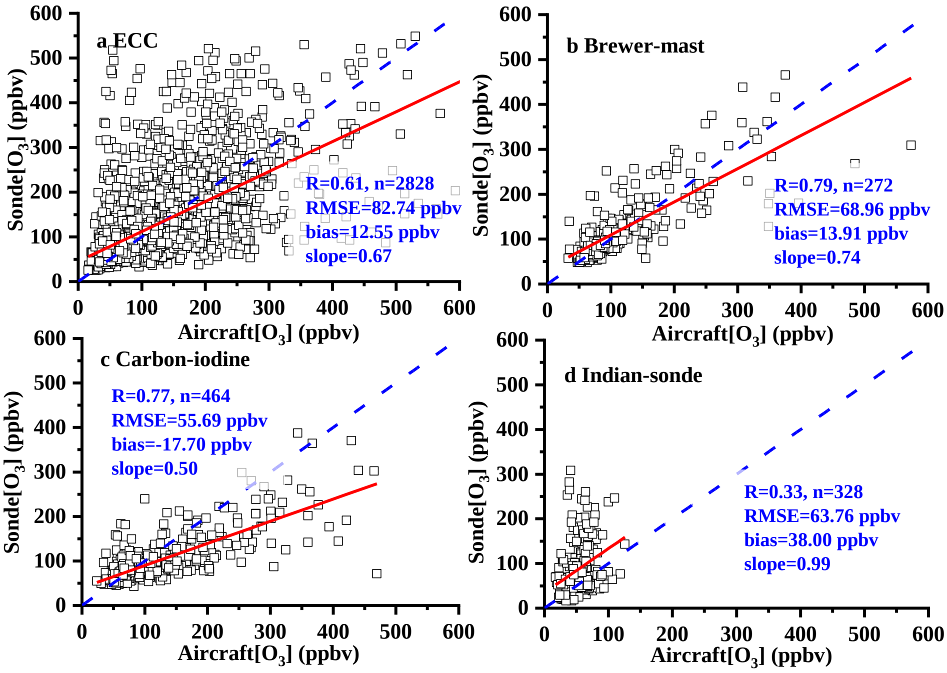


Figure 3 Correlation (*R*) of monthly mean ozone mixing ratios between ozonesonde and aircraft measurements at 8-11 km. The blue dashed line shows the 1:1 axis and the red line shows the linear fit. Correlations are significant at the 99% level (p <0.01). *N* denotes the number of data points, and RMSE is the root mean square error.

Along the same lines, a table similar to Table 2, but giving results not for the four season but for three (or more) altitude regions would be very helpful.

Response: Thanks for your suggestions. We have added "Figure S1 Bias, correlation coefficient (R), and RMSE for four types of ozonesonde and aircraft observations in different altitudes" and performed the analysis in the article. Please see the response to comment 1 for details.

Table 4: Please clarify what is shown: Ratio of three other sonde types to ECC sondes, using IAGOS as transfer standard. Instead of showing X/ECC, it might be better to show X/ECC-1 and difference values as percent. Also: are the given uncertainties / standard errors correct. For Brewer-Mast/ECC at 0~1km, the given ratio is 0.83+-0.96. That means the ratio could be anywhere between -0.13 and 1.79. Really that wide uncertainty range? Please check.

Response: Thanks for your suggestions. Table 4 is the "Ratio of three other sonde types to ECC sondes, using IAGOS as a transfer standard." We intend to give such ratios so that we can use these ratios to convert different types of ozonesonde data. In addition, the values in Table 4 are average ± 2 times the standard error (SE), that is, the 95% confidence limts. Unfortunately, they are that large. not the standard deviation.

Discussion at the end, around lines 310 to 320: Could the high ozone observed by sondes have something to do with insufficient background subtraction? Certainly might be a problem for the Indian sondes who are also flying in a region with low tropospheric ozone. What would be the implication of the new improved background estimation methods outlined by Vömel et al. 2020 and Smit et al. 2023? Please discuss.

Response: Thanks for your suggestions. Yes, and no; background subytaction is not much of an issue at northern midlatitudes (where most of our comparison data originate). We added instead" However, as noted by Saltzman and Gilbert (1959), the differences in stoichiometry found at different pH values imply that the chemistry of reaction of ozone with KI is complex, involving reactions that cause loss of iodine, as well as reactions other than the principal one that produce additional iodine. Several authors have noted the existence of slow side reactions involving the phosphate buffer, with a time constant of about 20 minutes, that may also increase the stoichiometry from 1.0 (Tarasick et al., 2021, Smit et al., 2024). Furthermore, evaporation causes the concentration of the sensing solution to increase, which can further enhance the stoichiometry, by concentrating the phosphate buffer, and to a lesser degree, by increasing the concentration of the KI itself (Johnson et al., 2002). These factors could contribute to the observed average relative bias between sondes and IAGOS found in this study."

Since we do not know the preparation and sampling process or data reduction scheme of the Indian-sonde, and there are no recent intercomparison results, we cannot speculate on the impact of stoichiometry or background subtraction on the Indian-sonde results.

Also: Is there no new information since Thouret et al. 1998 checking for the correctness of the MOZAIC / IAGOS inlet system? Is it possible that enhanced NO_x from aircraft exhaust gases results in local reductions (titration) of ozone in aircraft flight corridors? Local differences in the photo-chemical ozone regime (NO_x limited or VOC limited) could very well play a role for differences between sonde stations and airports in the lowest 1 to 3 kilometers of the atmosphere.

Response: Thanks for your suggestions. We have now included unpublished results of a recent intercomparison of IAGOS instruments at the World Calibration Centre for Ozone Sondes (WCCOS) in Julich in June 2023. According to the publicly available literature, no one has systematically compared the consistency of MOZAIC/IAGOS observations since Thouret et al. 1998. Enhanced NOx from aircraft exhaust gases may result in local ozone reductions (titration) in aircraft flight corridors. Unfortunately, we do not have relevant NOx data. Although IAGOS has NOx observation data, the observation data is relatively short-term, and the amount of data is small. When we compared the two data sets in our article, we first processed the two observation data sets into monthly average data. Therefore, what we compared was the consistency of the monthly average data of the two data sets. For the monthly average O_3 data, the impact of NO titration and local differences in the photo-chemical ozone regime (NOx limited or VOC limited) is relatively small, which is not the main reason for the difference between the two data sets.

References:

Smit, H. G. J., Poyraz, D., Van Malderen, R., Thompson, A. M., Tarasick, D. W., Stauffer, R. M., Johnson, B. J., and Kollonige, D. E.: New insights from the Jülich Ozone Sonde Intercomparison Experiment: calibration functions traceable to one ozone reference instrument, Atmos. Meas. Tech., 17, 73–112, https://doi.org/10.5194/amt-17-73-2024, 2024.

Vömel, H., Smit, H. G. J., Tarasick, D., Johnson, B., Oltmans, S. J., Selkirk, H., Thompson, A. M., Stauffer, R. M., Witte, J. C., Davies, J., van Malderen, R., Morris, G. A., Nakano, T., and Stübi, R.: A new method to correct the electrochemical concentration cell (ECC) ozonesonde time response and its implications for "background current" and pump efficiency, Atmos. Meas. Tech., 13, 5667–5680, https://doi.org/10.5194/amt-13-5667-2020, 2020.