Response to Reviewer #3

We thank the reviewer for their detailed comments on the manuscript. We have addressed these comments as described below. All reviewer comments are presented in italic font while the author responses are displayed in standard font. Specific text that was added to the updated manuscript is provided in blue text.

The manuscript submitted by Johnson and colleagues is a follow up of the work carried out by Mettig et al., 2022 using tropospheric ozone profiles reconstructed with the TOPAZ tool developed by the University of Bremen to exploit the synergy of UV (TROPOMI) and IR (CrIS) satellite observations. In this work the comparison is made over an 18-month period of TOPAZ retrieval and ground-based observations in North America (TOLNET lidar network and ECC ozonesondes). Mettig has already discussed the extent to which the synergy between UV and IR can improve the restitution of tropospheric ozone profiles but with a different validation data set based on NDACC observations in Europe and the USA. In the present work, the sensitivity study shown in Fig. 4 and the analysis of differences in several tropospheric layers are very useful, and was not present in that of Mettig et al. This work therefore deserves to be published in AMT, especially with the prospect of using TOLNET to validate the future GEO-TEMPO satellite mission.

My only minor concerns, which should be studied even if not taken into account, are the followings:

1) The discussion is sometimes based on the use of ground data convolved with AK of TOPAZ and sometimes based on the raw data interpolated vertically. It is better to use always the same criteria for the comparison of the three configurations. Use of the raw data should be made only for a better understanding of the results.

We thank the reviewer for this comment. For the satellite validation we compare observations (i.e., TOLNet and ozonesondes) convolved with retrieval AKs and a priori profiles (AK-convolved). Overall, we focus the validation of the satellites using the AK-convolved observations which we feel is well-described in the manuscript. However, it is also important to understand how satellite retrievals are able to replicate actual O_3 values, not just the capability of the spaceborne sensors, which was also done in other TROPOMI/CrIS validation studies (Mettig et al., 2022; Malina et al., 2022). To emphasize this, we have added the following text to the first section of the revised manuscript using TOLNet-raw data: "While observations convolved with the observation operator is the primary validation data source, comparing the three retrievals to TOLNet observations not convolved with the retrieval AKs (hereinafter TOLNet-raw) is also important to understand how the satellite retrievals reproduce actual atmospheric composition in the troposphere.".

2) The improvement when using the UV+IR configuration compared with IR-only is real for certain altitude ranges (boundary layer, UTLS) and for certain types of ozone profile (stratospheric intrusion), but does not significantly improve IR-only performance for other cases. This is not sufficiently recognized in the discussions of Fig. 5-6 and tables 2-3.

Sect. 3.3.3 of the revised manuscript, which focuses on the comparison of the 3 retrievals at multiple tropospheric layers, has been updated significantly to better describe the performance of IR-only retrievals in comparison to the two other retrievals. We also provide more quantitative information about the 3 retrievals evaluation at each vertical layer. At many points in this section, we now show how IR-only retrievals actually perform better compared to UV-only and UV+IR retrievals. An example of how we demonstrate this point is as follows: "Overall, between 2-8 km asl, IR-only retrievals have the least bias and spread, along with best linear regression fits. UV+IR retrievals are similar to IR-only data with only slightly worse performance when compared to TOLNet-AK. This result demonstrates that while the combination of UV and IR wavelengths tends to improve the performance of TOPAS retrievals compared to UV-only, this is not always the case for IR-only.".

3) It's a pity that the ozonesonde measurements are not used in conjunction with those from TOLNET for the scatterplots shown in each altitude layers (Fig. 7) and for the analysis of the seasonal variability (Fig.8). This would increase the representativeness of the results, as ozone distributions from TOLNET and ozonesondes are clearly complementary. We are left with the impression that the ozonesonde data have been discarded in the second part of the paper because they do not show a decisive contribution from IR+UV compared with IR-only in Fig. 6 and Table 3.

We appreciate this comment. However, the focus of this study is to demonstrate the network-wide TOLNet capability to validate satellite O_3 retrievals. In order to show that TOLNet was sufficient for validating satellite O_3 profile retrievals in the troposphere, besides the fact these lidar data have been evaluated in past research and are shown to be highly accurate as discussed in the manuscript, it is important to see whether TOLNet results in similar validation statistics compared to the well-known satellite validation data source from ozonesondes. The final sentence of the abstract has been updated to read: "TOLNet was shown to result in similar validation statistics compared to ozonesonde data, which are a commonly-used satellite evaluation data source, demonstrating that TOLNet is a sufficient source of satellite O_3 profile validation data in the troposphere which is critical as this data source is the primary product identified for the tropospheric O_3 validation of the recently-launched Tropospheric Emissions: Monitoring of Pollution (TEMPO) mission.". If we combine the two validation data sources (i.e., TOLNet and ozonesondes) it is not possible to determine the similarities and differences between the validation using TOLNet and the well-known validation data source from ozonesones. The similarities in validation results determined in this study, compared to other past TROPOMI/CrIS validation studies (e.g., Mettig et al., 2022; Malina et al., 2022), which primarily used ozonesondes, are also important to demonstrate the capabilities of TOLNet to validate a satellite product. As noted above, we attempted to emphasize this objective in the original manuscript but have added additional text to help highlight this point such as that implemented into the results section of the updated manuscript: "The agreement in the validation statistics of TROPOMI UV, CrIS IR, and TROPOMI/CrIS UV+IR retrievals determined in this study when using TOLNet-AK and those using primarily ozonesonde data (Mettig et al., 2022; Malina et al., 2022) demonstrates that TOLNet is a sufficient validation source for satellite O_3 profile retrievals in the troposphere." and "It is important to note that TOLNet and ozonesonde validation statistics are generally consistent given the fact that ozonesondes are a highly-accurate and commonly-applied satellite validation data source. This suggests that TOLNet is a sufficient validation data source of tropospheric O_3 profile retrievals from satellites.".

Detailed questions or suggestions

Abstract line 27: Since contrary to Mettig, 2022 data in Europe are very limited in this work (10 % of the data base in September 2019), it is better to replace « Europe » by « Netherland in September 2019 »

We agree with the reviewer, and this has been corrected in the abstract of the revised manuscript.

Abstract line 51: TOLNET data are certainly consistent for a seasonal analysis, is it also true for the altitude range analysis?

We apologize for the confusion this text caused. We have removed "Consistent daily" from the beginning of the sentence to remove any potential confusion. We did not want to suggest that the O_3 profiles from TOLNet were consistent within seasons or altitude ranges. We were attempting to state that the lidars can consistently provide data; however, this is not important to the results of this study, so we revised the sentence.

Line 104: The introduction provides a very nice review of the different satellite missions including their horizontal and vertical resolution. A table to summarize these resolutions would be useful.

This information has been implemented in the updated manuscript as Table 1.

Line 112: Mettig et al. study also includes NDACC and SHADOZ observations in Europe and the Tropics (ozonesonde and lidar) in addition to the TOLNET lidar in California and Huntsville. The sentence should be changed to mention it.

The sentence has been modified in the revised manuscript to read: "The combined UV+IR TROPOMI/CrIS O_3 profile retrievals from Mettig et al. (2022) were evaluated in the troposphere for a full-year between 2018-2019 using a small sample (2 lidar systems which are also part of the Tropospheric Ozone Lidar Network (TOLNet)) of ground-based lidar remote-sensing observations from the Network for the Detection of Atmospheric Composition Change (NDACC) and ozonesondes (i.e., World Ozone and Ultraviolet Radiation Data Center (WOUDC) and the Southern Hemisphere Additional Ozonesondes (SHADOZ)) and demonstrated that the combined UV+IR retrievals were more consistent with observations compared to the UV-only product.".

Line 125: In order to clarify the contribution of this new study in relation to the work of Mettig et al., might be good to add « with an emphasis on North America and many lidar instruments» after « O3 profile retrieval ». It might be good to specify here that a detailed statistical analysis at different altitude ranges is conducted in this work while this point was not developed in Mettig et al.

We agree with the reviewer that a more direct statement here would help separate this work from Mettig et al. (2022) and the previously uncited work by Malina et al. (2022). Besides using all the available TOLNet systems, which was not done in either Mettig et al. (2022) and Malina et al. (2022), we focus on chemical environments which are critical for air quality and tropospheric composition which can be challenging to retrieve from space (i.e., stratospheric intrusions, PBL O_3 enhancements) which was not done in Mettig et al. (2022) and Malina et al. (2022). This is emphasized in the updated manuscript: "This analysis of complex atmospheric environments important for air quality using idealized retrievals, produced with known O_3 profiles convolved separately with different retrieval AKs, in this study expands on past studies that have evaluated TROPOMI/CrIS retrievals (Mettig et al., 2022; Malina et al., 2022). It is important to understand the extent to which TROPOMI, CrIS, and TROPOMI/CrIS joint satellite retrievals, which rely on different wavelengths, can accurately retrieve typical and anomalous structures of O_3 in the troposphere.". In addition to this, our study conducts a very detailed validation of satellite O_3 profiles at multiple vertical levels of the troposphere which was not done in Mettig et al. (2022) and Malina et al. (2022). This is now emphasized in the updated manuscript: "This TROPOMI/CrIS validation at multiple layers in the troposphere allows for more detailed interpretation of the capability of these satellite vertical profiles to retrieve middle- to lower-tropospheric O_3 in comparison to other recent TROPOMI/CrIS validation studies (Malina et al., 2022; Mettig et al., 2022).". Furthermore, numerous sections in the updated manuscript have been revised to compare the results from our study to the two other TROPOMI/CrIS O_3 profile validation studies from Mettig et al. (2022) and Malina et al. (2022).

Line 145: The number of lidar observations considered (185) is different from the maximum value in table 2 (176). They should be consistent. The Mettig et al. study is finally not so different (170 lidar data and 200 ozonesondes for the same time period 2018/2019)

Table 2 in the revised manuscript presents the total number of days with observations for each lidar system and location. This sums to 185 which is what is stated in Line 145 of the original manuscript. Due to limitations of lidar retrievals due to inability to retrieve O_3 values accurately in high cloud and/or aerosol conditions, and saturation due to solar background becoming too large and saturating the lidar signal, the numbers in Table 3 of the revised manuscript will differ between vertical levels. Also, the TOPAS retrieval provides data at 1 km, thus more than one satellite/TOLNet co-location for the 2 km bins for each of the 89 total profile co-locations (N = 89 in Fig. 5) is possible.

The reason that the number of lidar profiles used in our study is similar to that in Mettig et al. (2022), even though we use more lidar systems, is that we use stricter colocation criteria. This is explained in the original manuscript starting in Line 238. We have updated this text in the revised manuscript to be more specific: "Statistical comparisons between co-located satellite retrievals and observations were conducted using spatiotemporal thresholds of 2.5 hours and 30 km. Sensitivity studies were conducted using coarser co-location spatiotemporal thresholds of 5 hours and 100 km to maximize the number of co-locations for statistical evaluation and to be more consistent with recent TROPOMI/CrIS O_3 profile validation studies which use looser colocation thresholds (Mettig et al., 2021, 2022). As this study focuses on tropospheric O_3 which has large spatiotemporal variability, we feel the stricter spatiotemporal thresholds are most appropriate.".

Line 165: give here the seasonal distribution of the TOLNET observations given line 473

The following sentence has been added to this section of the revised manuscript: "This study includes 13, 28, 78, and 66 TOLNet observations for the winter (DJF), spring (MAM), summer (JJA), and fall (SON) months, respectively.". Keep in mind this number will not match that given on Line 473 in the original manuscript as not all TOLNet observations pass co-location spatiotemporal thresholds.

Line 176: Add the positions of the ozonesonde stations on the TOLNET map (Fig.1). What is the seasonal distribution of the soundings?

Figure 1 was provided to inform the readers about the home stations for each of the lidar systems in TOLNet. The spatial locations of the lidar systems, and ozonesondes, are provided in Table 1 of the original manuscript (now Table 2 of the revised manuscript). In order to remind the readers to find this location information the following text was added to the manuscript: "In order to have a direct comparison of the validation using ozonesonde and TOLNet, we use ozonesondes which were nearly directly spatially and temporally co-located with lidar systems as shown in the location information provided in Table 2.".

Text was added to the revised manuscript to describe the seasonal distribution of the ozonesonde data: "The seasonal distribution of these ozonesondes were: 2, 2, 39, and 8 for the winter, spring, summer, and fall months, respectively.".

Line 236: In equation 3, I guess Xc is the convolved observations using the satellite AK.

That is correct. To clarify this, we have updated this sentence to read: "The satellite retrievals were compared to raw observations and when convolved (X_c) with the averaging kernel ($\boldsymbol{AK}$) and a priori information from each retrieval using Eq. (3)".

Line 245: I guess the " known TOLNet O3 profile" is the black curve in Fig. 4. Please be more specific here.

This is now described in the revised manuscript as: "…replaced with a known TOLNet O_3 profile (black lines in Fig. 4).".

Line 285: Once it has been stated that UV-only has limited information below 15 km, I suggest changing the way the end of this sentence is written to focus on the comparison with IR-only:

«are much improved (8-10 km) compared to UV-only profiles below 15 km asl. » by

«are improved (8-10 km) compared to IR-only above 12 km and below 8 km ».

We agree with the review and this sentence now reads as: "When combining UV and IR information vertical resolutions of the retrievals are improved (8-10 km) compared to IR-only above 12 km and below 8 km.".

Line 306: Fig. 4 is a very nice figure and a useful addition to Mettig et al. study. I disagree with the statement « demonstrate the capability of the UV-only, IR-only, and UV+IR retrievals to replicate tropospheric and lowermost tropospheric O3 during a PBL pollution event ». None of the configurations is able to reproduce the ozone enhancement in the lowermost troposphere. It is not so surprising considering the low value of the AK below 2km. It is better to emphasize the very good results obtained for the stratospheric intrusion case for UV+IR, where TOPAZ avoids the downward propagation of the upper tropospheric enhancement compare to IR only.

We apologize for the confusion the wording of this sentence caused. The purpose of this sentence was simply to introduce/describe Fig. 4. However, as written in the original manuscript, it reads more as the TOPAS retrievals using all wavelengths were able to replicate O_3 profiles during pollution and stratospheric intrusions cases. As stated by the reviewer, this is not the case for the PBL pollution event. We have edited this sentence to read: "Figure 4b and 4c demonstrate whether the UV-only, IR-only, and UV+IR retrievals were able to replicate tropospheric and lowermost tropospheric O_3 during a PBL pollution event and a stratospheric intrusion, respectively.". The rest of this section in the original manuscript describes the inability of the retrievals to capture the enhanced O_3 values in the PBL and ability of the UV+IR products to capture tropospheric O_3 during a stratospheric intrusion event.

Fig. 4. Considering the very high value of this figure, I suggest to add the 6-12 km NMB in panel c to discuss the ability of the 3 configurations to reproduce the upper tropospheric enhancement.

This is a good point. We have added the NMB values for the mid- to upper-troposphere (4-12 km) for the stratospheric intrusion case study shown in Fig. 4c. We also added text to the revised manuscript to describe this evaluation: "In the mid- to upper-troposphere (4-12 km), UV+IR retrievals had the least high bias (NMB) of 11.3% while IR-only (12.8%) and UV-only (15.9%) retrievals had larger high biases. Compared to the a priori, true lidar profiles convolved with all three retrieval AKs compared much more accurately emphasizing the ability of these retrievals to capture enhanced mid- to -upper tropospheric O_3 enhancements.".

Line 314: Yes I agree with this last statement. Why is this result different from the Cuesta et al. comparison between chemical transport model and combined analysis of GOME-2 and IASI showing a reasonable agreement for ozone enhancement below 3 km? This is worth to be discussed in section 4.

Thank you for pointing out these differences. The following text has been added to Sect. 4 of the revised manuscript: "Applying different combinations of UV+IR joint wavelength retrievals (e.g., GOME-2+IASI) also displays improvements compared to UV-only products in the troposphere similar to that determined in this study (e.g., Cuesta et al., 2013, 2018). Cuesta et al. (2013, 2018) demonstrated how GOME-2+IASI retrievals show high accuracy compared to ozonesondes in the lowermost troposphere and displays a clear capability to capture PBL O_3 enhancements. This differs from the results of this study which suggest that TROPOMI+CrIS UV+IR joint wavelength retrievals still struggle to reproduce large PBL O_3 enhancements due to limited lowermost tropospheric sensitivity. The reasons why GOME-2+IASI displays the remarkable capability to retrieve lowermost tropospheric enhancements compared to the results from TROPOMI+CrIS is not immediately apparent. There are differences in the retrieval algorithms, a priori input data sets, and the spectral resolutions of the UV and IR sensors applied. Comparing our results to Cuesta et al. (2013) shows that DOFs are higher in the troposphere and in the 0-2 km agl column (>33% higher) in GOME-2+IASI retrievals compared to TROPOMI+CrIS which would explain some of the differences in capabilities to retrieve lowermost tropospheric O_3 enhancements.".

Table 2. Why are the numbers of observations different in the different vertical layers? Altitude range of the lidar profiles? Clouds?

We provided this explanation in response to a comment above. Due to limitations of lidar retrievals to accurately retrieve O_3 values in high cloud and/or aerosol conditions, and when mid-day solar background becomes too large and saturates the lidar signal, the numbers in Table 2 will differ for each vertical level.

Line 370: I would suggest discussing all the results in Table 2, including RMSE and bias, in this paragraph instead of mixing them with other topics of Section 4, which should be limited to a general discussion and comparisons with previous works.

The paragraph discussing RMSE in Sect. 4 in the original manuscript has now been moved into Sect. 3.3.1 of the revised manuscript where Table 2 (now Table 3 in the updated manuscript) statistics are discussed.

Fig. 5 and Table 2. I do not understand the 89-number of observations in Fig. 5 caption while Table 2 shows up to 172 colocations. Better to have IR-only and UV+IR on the same page in Table 2.

The total number of satellite/TOLNet profile co-locations using the spatiotemporal co-location criteria of 2.5 hours and 30 km resulted in 89 co-locations. The TOPAS retrieval provides data at 1 km, thus more than one satellite/TOLNet co-location for each of the 2 km bins is possible for each of the 89 total profile co-locations. This is why the numbers in Table 2 for each 2 km bin are larger than 89.

The revised manuscript has been updated so all tables are on the same page.

Fig.6 and Table 3. Again I do not understand the 26-number in the Fig. 6 caption while 50 soundings are considered in Table 3.

The answer for this comment is the same for ozonesondes as described above for satellite/TOLNet co-locations.

Line 375-385: It is a pity that the differences with the TOLneT comparison are not highlighted. This paragraph sounds very positive while the differences with the ozonesonde-raw are significant below 4 km. The improvement using IR+UV instead of IR-only is not obvious anymore for this subset (NMB in Table 3). Is it because the ozonesonde profiles include several cases with lowermost tropospheric enhancement as shown in the sensitivity study in Fig.4b ?

Text was added to the revised manuscript to better describe the comparison of TOPAS products with ozonesondes and TOLNet data. In the paragraph describing the evaluation with Ozonesonde-AK we added the following sentence: "The evaluation of TOPAS retrievals with Ozonesonde-raw differs from results using TOLNet-raw primarily below 4 km where ozonesonde observed large O_3 enhancements in the lowermost troposphere which were not evident in the TOLNet data.". The final paragraph of Sect. 3.3.1 discussed the similarities and differences between the validation of TOPAS retrievals with TOLNet-AK and Ozonesonde-AK. It has been updated slightly in the revised manuscript to better explain this comparison: "In the troposphere, UV-only retrievals were consistently biased high compared to Ozonesonde-AK data (see Fig. 6a, b; Table 4). This systematic high bias is consistent with the validation using TOLNet-AK observations. IR-only O_3 profiles compare very well to Ozonesonde-AK data with NMB values <3% throughout the troposphere. This outperforms the IR-only profiles when compared to TOLNet-AK data which displayed a low bias aloft. Finally, the UV+IR retrievals have minimal bias below 10 km asl with NMB values <10% when compared with TOLNet-AK observations; however, when compared with Ozonesonde-AK data the UV+IR retrievals had a noticeable high bias above 9 km. The overall validation of the three satellite O_3 profile retrievals using Ozonesonde-AK was generally consistent compared to when using TOLNet-AK. It is important to note that TOLNet and ozonesonde validation statistics are generally consistent given the fact that ozonesondes are a highly-accurate and commonly-applied satellite validation data source. This suggests that TOLNet is a sufficient validation data source of tropospheric O_3 profile retrievals from satellites. Given that TOLNet is able to accurately validate satellite-derived O_3 profiles, and the focus of this work is on the demonstration of TOLNet for validating satellite retrievals, the rest of this study focuses on the validation using the lidar network observations.".

Line 385. As mentioned for Table 2 it is good also to include the RMSE and bias analysis of Table 3 in this paragraph. By the way why is IR-only RMSE smaller than UV+IR RMSE? This should be discussed.

We agree with the reviewer and the following paragraph has been added to the revised manuscript: "The RMSE values in Table 4 represent the random errors in the daily TOPAS O_3 profile retrievals when validated with Ozonesonde-AK observations. All three TOPAS retrievals had lower random errors compared to the a priori profiles; however, random errors still remained elevated in most instances except for the IR-only retrievals. UV-only retrievals had unresolved errors ~50% less compared to the a priori (13.9 ppb). IR-only retrievals displayed the least unresolved errors of all three retrievals with average RMSE values of 6.1 ppb which is ~80% less compared to the a priori. The combined UV+IR profiles had average RMSE values of 11.4 ppb, ~60% less compared to the a priori, throughout the troposphere. Given that unresolved errors of daily profiles on average still remain large (>10 ppb) for retrievals using UV wavelengths (UV, UV+IR), the accuracy of these satellite products still suffer due to the limited sensitivity of spaceborne sensors to tropospheric O_3. On the contrary, NMB and RMSE values for IR-only retrievals when compared to Ozonesonde-AK observations were low suggesting this product had some skill in capturing the daily vertical distributions of O_3 in the troposphere during this validation. This increased tropospheric sensitivity in IR-only profiles, and when combining UV and IR wavelengths, allows these retrievals to deviate from a biased a priori profiles which improves the ability of this retrieval to capture daily O_3 vertical profile distribution variability in the troposphere which is agreement with many recent studies (e.g., Landgraf and Hasekamp, 2007; Worden et al., 2007b; Cuesta et al., 2013, 2018; Costantino et al., 2017; Colombi et al., 2021; Malina et al., 2022; Mettig et al., 2022).".

Line 428: This sentence is relevant for the validation of the future TEMPO-GEO mission. It is not mandatory for the analysis of the satellite data of this paper where lidar and ozonesonde observations are equally relevant. It is a pity that the ozonesonde data are not included in Fig. 7. The latter ozone vertical distributions are indeed different and complementary from those corresponding to the TOLNET observations (see the comparison between Fig. 5 and 6).

We have updated this sentence to now read: "A major advantage of using TOLNet for validation of satellite O_3 profile retrievals is the ability to make accurate, high temporal and vertical resolution observations at different vertical levels of the troposphere.".

The main objective of this manuscript was to demonstrate the capability of using TOLNet to validate satellite O_3 profile retrievals. Therefore, Fig. 7 in the revised manuscript still only applies TOLNet observations. However, based on the reviewers comment below, we now add ozonesonde data with TOLNet to validate seasonal TOPAS retrievals (Fig. S2 in the revised manuscript) to increase the number of seasonal co-locations.

Line 441: The results of RMSE and slopes in the 4-6 km are not much better than those in the layers 0-4 km even for the IR-only and UV+IR while the DOF are significantly larger than below 4 km, e.g. the slopes in the layer 4-6 km in Fig. 7 are < 0.5 in the Table 2 and Fig. 7. The reason for this limited improvement of IR or IR+UV could be discussed in this section, instead of focusing again on the limitation of UV-only configuration. The latter is already very well demonstrated by the results presented in p. 13 to p.18.

We now provide more quantitative information for the comparison in these two layers: "Between 2-4 km the UV+IR (NMB of 5.8%, RMSE of 11.7 ppb, slope of 0.46) and especially IR-only (NMB of 4.9%, RMSE of 6.5 ppb, slope of 0.54) retrievals outperform UV-only retrievals (NMB of 18.0%, RMSE of 14.6 ppb, slope of 0.14) due to the enhanced sensitivity provided by the IR wavelengths. The UV+IR and IR-only retrievals have better linear regression slopes compared to the UV-only product (UV-only results have similar slopes as the a priori profile below 6 km) due to the ability to deviate further from the a priori profile shape. In the vertical layer between 4-6 km, similar to the layer between 2-4 km, the UV+IR (NMB of 6.2%, RMSE of 12.3 ppb, slope of 0.45) and in particular the IR-only (NMB of 2.4%, RMSE of 7.5 ppb, slope of 0.62) retrievals outperform UV-only (NMB of 20.1%, RMSE of 16.0 ppb, slope of 0.20) retrievals with less bias and RMSE and better linear regression slopes.". The following text has been added following this to discuss the differences in performance between 2-4 km and 4-6 km: "It should be noted that the retrievals without UV wavelengths (IR-only) was the only satellite product with improved statistics (lower NMB and higher slope) at 4-6 km compared to 2-4 km. The vertical level around 4-6 km is where IR-wavelengths have peak sensitivity to O_3 in the TOPAS CrIS retrieval, which contributes to this result. However, given that DOFS for O_3 profile retrievals are < 1.0 below 12 km agl, no individual 2 km layer evaluated in this study is completely independent from the retrieval performance throughout the troposphere.". It is important to note that each 2 km layer is not independent and is also driven by retrieval performance at all vertical levels throughout the troposphere.

Line 471-475: The seasonal analysis is indeed a nice contribution of this paper. However the number of limited co-locations being a limitation of the interpretation of the results, once again the use of the ozonesonde data as well as the TOLNET observations would improve the value of such an analysis.

We agree with the reviewer that including both TOLNet and ozonesonde data to validate the satellite retrievals would increase the number of seasonal co-locations. In the updated manuscript we include a supplemental figure with this validation. The following text has been added to the revised manuscript: "The focus of this study was to demonstrate the capability of TOLNet data to validate satellite retrievals; however, to improve the number of seasonal co-locations we also used ozonesonde data (Ozonesonde-AK), in addition to lidar measurements, and these results are shown in Fig. S2. Given the performance of the validation was similar when using TOLNet-AK and the combination of TOLNet-AK and Ozonesonde-AK, the main text of the paper focuses on the seasonal validation of TOPAS retrievals using TOLNet-AK data only.".

Line 504: I disagree with this statement. The IR-only shows better results below 9 km and the UV+IR SON differences in Fig. 8 are often larger than 10%.

We thank the reviewer for catching this mistake and this paragraph has been updated in the revised manuscript as: "At all altitudes in the troposphere during the fall months the retrievals using IR wavelengths (IR-only and UV+IR) compared the best to observations with NMB values <15%. UV-only retrievals had consistent high biases typically >20%. IR-only profiles had the best overall performance with small biases (within ±10%) below 9 km and a larger negative bias aloft. All three retrievals had smaller RMSE values compared to the a priori of 16 ppb, 10 ppb, and 13 ppb for the UV-only, IR-only, and UV+IR retrievals, respectively. Similar to the summer months, during the fall all three retrievals had noticeably lower random errors compared to the a priori profiles.".

Line 526. The sentence is not complete

To address another reviewer's comment this sentence was removed. We have demonstrated this point throughout the manuscript and it is not needed here.

Line 527-537. The discussion of RMSE values of Table 2 and 3 would be understood if presented in section 3.3.1 where Fig. 5 and 6 and other statistical parameters of Table 2 and 3 are presented. Mixing this RMSE analysis with a general discussion of the value of the paper results and with a comparison with previous work make reading of this paragraph a little bit difficult.

As mentioned to a previous comment we have moved the RMSE discussion to Sect. 3.3.1 in the revised manuscript.

Line 567-568: Remove or change the part of the sentence saying "more capable of capturing conditions with air quality impacts such as pollution events " because this paper does not show this paper does not provide strong evidence for this. It is mainly shown that the stratospheric intrusions are better reproduced.

This sentence has been revised to read: "Retrievals using combinations of wavelengths proved to be more capable of capturing conditions with air quality impacts such as stratospheric intrusions.".

Line 574: Again remove the end of the sentence saying "during times of PBL-level O3 enhancements" as it is not clearly shown in this paper (see Fig. 4 or Fig. 6).

This sentence has been removed in the revised manuscript.

References:

Malina, E., Bowman, K. W., Kantchev, V., Kuai, L., Kurosu, T. P., Miyazaki, K., Natraj, V., Osterman, G. B., and Thill, M. D.: Joint spectral retrievals of ozone with Suomi NPP CrIS augmented by S5P/TROPOMI, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2022-774, 2022.

Mettig, N., Weber, M., Rozanov, A., Burrows, J. P., Veefkind, P., Thompson, A. M., Stauffer, R. M., Leblanc, T., Ancellet, G., Newchurch, M. J., Kuang, S., Kivi, R., Tully, M. B., Van Malderen, R., Piters, A., Kois, B., Stübi, R., and Skrivankova, P.: Combined UV and IR ozone profile retrieval from TROPOMI and CrIS measurements, Atmos. Meas. Tech., 15, 2955–2978, https://doi.org/10.5194/amt-15-2955-2022, 2022.