RC2 - answers

General comments

The authors have improved their manuscript since the previous submission and some of the open questions have been addressed. But the discussion in several of the sections remain confusing. I strongly encourage the authors to review the manuscript again for clarity to ensure that important points are stated in simple language at the appropriate locations in the text. In several cases it only became clear what the authors meant when I read discussions later in the manuscript. I have identified a few of these instances (see below), but my review in this respect is not complete. I feel that a few sentences early in this paper outlining the analysis approach will help the reader to connect the content in the various sections.

We thank the reviewer again for the attentive reading and constructive feedback, which have significantly improved the clarity and structure of our manuscript. We regard the review process as an essential step in enhancing the communication of our research, therefore we appreciate the encouragement to strengthen the logical flow and we have revised key sections accordingly. Below, we address each comment in detail.

1. Some of my comments about the earlier version of this paper related to conclusions about the stray light characteristics that I felt were inadequately substantiated. In this revised version it seems as though the authors have the evidence at their fingertips to demonstrate they are accurately describing the internally scattered stray light characteristics of TropOMI. But actual evidence in this latest version of the manuscript is still lacking. The observed MgII index and the stray light region contents are two independent pieces of information about internally scattered stray light. The instrument parameter that connects the two is the stray light kernel. There are many potential stray light kernels that could equally explain one or the other of the observed phenomena. The authors will have a persuasive argument regarding their chosen SL kernel changes if they can demonstrate it explains both the MgII index and stray light row changes simultaneously. Perhaps the authors believe they have accomplished this, but I do not see the clear "2+2=4" arguments in the text. It may be that only minor adjustments are needed to achieve this objective.

   The straylight kernel changes do not explain the changes in Mg II index and straylight light rows. The straylight kernel is static, and any change will not account for the growth in signal in the straylight rows because the irradiance signal itself stays the same, or even decreases, and it does not increase enough around 317nm.

   In our manuscript, we describe a new pragmatic approach of trying to use the detector straylight rows (at the edges of the detector) as input for a dynamic – delta – correction, instead of changing the static kernel itself to account for the observed straylight growth in the straylight rows. The temporal change of the static kernel is theoretically possible; however, as there are no in-flight measurements for this purpose, this option has been discarded because it would not have enough foundation (this is also emphasized in the revised manuscript, lines 177-179).

The whole Sect. 2 has been updated, and we have also separated the updates of the 'static' straylight kernel from the description of the dynamic straylight corrections to enhance clarity (Sect 2.3 and 2.4, in the revised manuscript).

2. There should also be a discussion of the problems remaining in this new product. If the authors believe the resulting Level 1 product is not suitable for ozone trend analysis, they should state so and explain why. Don't leave it to the more careful readers to figure this out for themselves.

   We agree that this should be part of the manuscript. On the L1 side, in the 'Summary' section (2.6, revised manuscript), we mention that other choices can be made on the shape of the static straylight kernel and on the construction of the dynamic straylight correction (e.g. interpolation method). These choices were motivated by the specific goals of minimizing the over-correction of the straylight kernel and the dynamic straylight correction term. However, we do note that even with the updates of the reprocessed L1 version 3.0 data, the temporal bias is still not completely removed (Figure 12i-j or Figure 13, revised manuscript lines 398-399), therefore the soft calibration correction is still necessary (revised manuscript, lines 408-409).

   On L2 side, it is not possible for now to make a definitive statement on the effect on the ozone trends. We have not performed a trend study for our manuscript. Only a full reprocessing of the L2 ozone profile product, based on the L1 version 3.0 data, can confirm whether the latter results in reduced height-resolved drifts in comparison with the processing using L1 version 2.1 data, which has been found to show significant yet largely compensating positive tropospheric drifts and negative stratospheric drifts (Keppens et al., 2024).

   A preliminary drift analysis based on 15 days of new L2 data (L1 version 3.0, L2 version 2.9) pointed at a reduced – up to insignificant – tropospheric drift, which requires future confirmation based on the full dataset. This comment has been added to the Conclusions section (revised manuscript, lines 493-495).

3. **Line 125-129.** The flow of the discussion here is awkward. I find myself reading a rereading the sentences trying understand the authors' points. A clearer approach would be to first discuss all the potential explanations for the behavior of the MgII line depth, including solar irradiance change, then eliminate the ones that are not plausible. Walk the reader through the logic rather than hiding the key points in clauses at the end of sentences.

   We agree that the description of the approach followed for the re-analysis of the calibration measurements can be improved. To address this question and the following until number 9, we have significantly updated Sect. 2. The paragraphs have been restructured and renamed: the L1 updates are separately introduced in Sect. 2.3, 2.4, 2.5 in the revised manuscript. We also moved the figure in Appendix A to the main text

(Figure 3, revised manuscript) as it is an important connector image in this part of the manuscript.

4. **Line 143.** What are the authors intending to say in this sentence? Addressed by whom or by what? The correction algorithm? This paper? The detector? It is unclear what this sentence means in the context of the preceding discussion.
   When revising the whole Sect. 2, this phrase has been replaced with: "*In the UV detector, the correction is limited to in-band straylight: only the signal measured on the detector is used as input and the straylight correction is merely a redistribution of the measured signal (Ludewig et al., 2020).* " (line 152, revised manuscript)
5. **Line 165.** I recommend that the authors find a term other than "far-field" to describe Band 1 stray light originating from Band 2. Far-field or out-of-field refers to the field angle, which is a spatial dimension. Its use in describing the spectral dimension of stray light is confusing. Please use "out-of-band stray light" or more generically "spectral stray light" to describe contributions in the spectral dimension (also in Line 330).
   By far-field and near-field straylight we mean the distance between the origin and the destination of the straylight signal, both in spatial and spectral dimension (as the kernel is 2-dimensional). However, since the definition might not be very clear in other parts of the manuscript, in the revised manuscript we added a more specific description of 'far-field' straylight: "*We emphasize here that the far-field straylight is intended in the sense of straylight signal coming from pixels on the same detector but that are in a different position, both in the spectral and spatial direction, and associated with the outer parts of the 2D kernel*". (line 206-208, revised manuscript).
6. **Lines 170-175.** The authors discuss the motivation for altering the stray light kernels: non-zero signal content in the stray light region after SL correction, and over-correction in Band 1. The authors provide detailed explanations for the stray light kernel assessment in Appendix B, but only for the spatial kernel adjustment. The stray light region contents are far more sensitive to the shape of the spatial kernel than to the shape of the spectral kernel, and therefore the latter cannot be tuned on the basis of the stray light region data. Since spectral stray light should be the main focus of this paper (it is the primary cause of their ozone profile retrieval errors), it would be best if the authors clearly and explicitly present the methodology for modifying the spectral axis shape of the kernel.
   There is only one 2D straylight kernel with a spatial and spectral dimension. The update of the ellipticity in the straylight kernel shape gives greater weight to the spatial dimension of the kernel at the cost of the spectral dimension, while keeping the total kernel mass the same. As shown in Appendix B (Figure 7, in the revised manuscript), this choice prevents overcorrection. Moreover, since this action does not affect the results of the on-ground measurements analysis because the specific choice of the kernel has a free parameter (the eccentricity), the new kernel shape is still valid.
   We emphasize again that the change of the kernel shape does not account for the straylight growth in time. This has been accepted as a 'problem' and solved by using the

actual straylight measurements from the straylight regions to implement a dynamic correction.

In the revised manuscript, we moved the Appendix B figure to the main text (to Figure 7) as we realized that they provide useful information for the update of the straylight kernel. The text in Sect. 2.4 (on the static kernel) has also been updated to present the methodology in a clearer way.

7. **Section 2.2.** The authors fail to describe how the dynamic SL correction is actually implemented. This is not an irrelevant detail because there are several possible approaches and not all will equally explain the observed measurements. Is the time-dependent behavior based on the MgII lines or is it based on the stray light region contents? Is the kernel modified by broadening its width or are the tails merely increased by adding a uniform background to the kernel? Are the spectral and spatial kernels adjusted identically or independently. These choices will significantly affect the residual signals.

   We agree that indeed there are several approaches to address the straylight growth in time. In the revised version of the manuscript, we try to improve the explanation of this additional correction algorithm introduced in the L1 processor. The time-dependent behaviour of the dynamic straylight correction is based on the straylight rows of the UV detector which contain in-flight straylight information. The Mg II lines do not play any special role in the dynamic correction algorithm. The correction is implemented as a linear interpolation in the spatial direction, between the upper and lower straylight rows, after smoothing and quality controls. The static kernel is part of the straylight correction algorithm, but it is a separate algorithm, implemented before the dynamic correction. The kernel update to elliptical shape affects only this static straylight algorithm, not the dynamic one. To enhance clarity for the reader, we separated the description of the L1 straylight algorithm updates in Sect 2.3 and 2.4 (revised manuscript).

8. **Section 2.3.** The authors explain that there are signal errors remaining even after the standard signal corrections. They describe the correction as a CKD image, which requires a bit more explanation for readers (or eliminate it if it provides no useful information). It is difficult to understand from the discussion exactly what artifacts are being corrected; how are they identified and how are they corrected. What is the magnitude of the correction? It is only in the discussion surrounding Figure 9 later in the paper that the reader discovers these additional signal corrections have little effect on the results. The authors should say so in this section. As currently written, the reader is left wondering what these errors are and the role they play in the observed TropOMI radiometric behavior.

   We agree that the structure of this paragraph can be improved, so we made two updates in the manuscript: first, we introduced the residual correction already in Sect. 2.2 (revised manuscript lines 141-147), then we also updated the relative Sect. 2.5 to discuss the correction itself. Since the residual signal cannot be modelled, it is obtained from monthly aggregates of background measurements of the night side of the orbit.

Because the construction described above requires a significant effort and because the residual signal is a very small fraction of the radiance signal itself, it was initially (after launch) decided to have this correction off. However, this argument is less valid for band 1 as we explain in the manuscript, hence the decision to implement it anyway. The simultaneous observation of the residual binning artifacts in the soft calibration spectra (Figure 12c-d, revised manuscript) gave an additional incentive to look again into the background measurements.

9. **Lines 201-204.** The before/after comparison of the MgII index and Ca K line index shown in Figure A2 suggests the post-correction signal variation at 280 nm may be close to true solar variation, but it is hardly convincing evidence. It would be better if the authors could directly compare the post-correction MgII index observed by TropOMI with a predicted index (adjusted for the TropOMI bandwidth) based on an external source. Furthermore, this evidence and its discussion should not be relegated to an appendix. In my opinion, it is central to the stray light discussion and the authors' assertion that TropOMI stray light characteristics are changing in orbit.

   We think that the correlation between two different detectors in TROPOMI is equally convincing, at least on our weak claim that it is an improvement in the right direction. However, we compared the TROPOMI Mg II and Ca K lines also with an external data set (the Bremen Composite Mg II index). This comparison is shown in a second panel in Figure 5 (revised manuscript), together with the correlation of the TROPOMI Mg II and Ca K (previously in Appendix A). We agree that they should be in the main text as they are part of the validation of the newly introduced dynamic straylight correction algorithm in the science region.

10. **Section 3.** In this section the authors describe an approach to radiometric corrections that is rather difficult to follow. It lacks a high-level discussion of the authors' motivation for choosing this soft calibration approach. For example, it appears that the soft calibration of TropOMI is based on ozone climatology, not just initially but also throughout the mission. I presume this means the authors have given up on producing radiances suitable for trend-quality ozone retrievals. If this is the case, the authors should say so clearly and unequivocally here and in the abstract. If it is not the case, the authors need a clearer discussion in this section to explain how radiance trends are preserved.

    To improve clarity regarding the soft calibration procedure, we have substantially revised this section. The introduction to this section (revised manuscript) contains the general description of the TROPOMI UV soft calibration, which before was in the Method section 3.3, creating perhaps confusion in the flow of the manuscript.

    Regarding the climatology, we do not use any ozone climatology in the soft calibration calculation but the CAMS operational forecast profiles, adjusted with the TROPOMI total ozone and MLS profiles. This is described in the forward model section (3.2, revised manuscript), outlining the framework used for the modelled radiances. The ozone climatology (Labow et al, 2015) is used as a-priori input for the ozone profile retrieval algorithm which is not discussed in this manuscript.

Regarding the question over the trend-quality ozone retrievals, we refer to our response in question 2.

11. The soft calibration approach also addresses an apparent non-linear response of the instrument. Is this an existing operational correction that you are merely updating to be more accurate, or is this an entirely new correction? Though this may have been previously described in another publication, these unusual corrections need some high-level description and justification. Figures 7 and D2 indicate there is a residual non-linearity in the instrument system. What is the cause of this non-linearity? Detector? Stray light overcorrection? Incorrect wavelength registration? Something else? How this non-linearity is addressed depends a lot on its cause. For example, a detector non-linearity will not have a wavelength dependence, nor does it typically have much time-dependence. Stray light errors are not addressed well by broad addition or subtraction of signal because they depend on neighboring pixel signals. Indeed, a comparison of Figures 9 and 10 suggests that some of the non-linearity was caused by stray light, but much of the non-linearity remains between versions. Even if the authors do not know the exact cause, they should be able to speculate how these errors might arise. The statements by the authors in lines 309-312 that such discussion is outside the scope of this study do not absolve the authors of their responsibility to convince the readers they are addressing the data problems in an appropriate manner.

    We thank the reviewer for this careful observation and the opportunity to improve our analysis. It is important to note that because of the empirical nature of the soft calibration correction, instrument (L1) and forward model (modelled radiance) errors are combined, making it difficult to unambiguously separate these contributions.

    In the revised Sect. 3, we have more clearly emphasized two aspects of the correction that may not have been sufficiently described previously:

    i. The soft-calibration correction is computed as a function of the radiance signal as well. This formulation is important to capture both additive and multiplicative effects in the instrument response.
    ii. The forward model of the modelled radiances implements a surface albedo fit in band 2 (328-330 nm), prior the residuals' calculations. This provides an anchoring point in band 2 that effectively sets the residuals to zero in that range for most cases. However, this approach has also limitations, as it cannot account for all non-clear-sky scene variability.

    In addition to significantly revising Sect. 3, we added a new Discussion section (Sect. 3.5) in the revised manuscript to further discuss the soft calibration spectra obtained with L1 version 2.1.

12. **Section 3.4 and Figure 8.** It is clear, upon reading Section 4, that the soft calibration presented here applies to the current operational processing rather than the proposed future operation processing. This point should be stated more clearly. Perhaps the

authors should refer to the actual product version numbers rather than "operational" and "updated" to indicate old and new.

Yes, Sect. 3.4 and Figure 8 referred to the soft calibration correction computed using the L1 version 2.1. We agree that it would be clearer to use the L1 processing version numbers, so we updated text and figure captions accordingly in the revised manuscript.

13. **Lines 314-317.** This is a confusing paragraph. Exactly which corrections were applied in the reprocessing? Based on the subsequent text in this section and the contents of Figure 9, it seems that the authors are trying to say that they are recomputing the soft cal. correction for different combinations of instrument corrections applied in the L1B processing. If this is what the authors mean, just state so in plain and simple language. What does the sentence "The soft calibration procedure is instead kept the same ..." ? Instead of what?

    Yes, we recomputed the soft calibration correction using different L1 processings with corrections switched off and on in a particular combination. The procedure of the soft calibration correction computation is left the same, so the only difference is in the L1 data. We realized that this paragraph could be confusing, so we rephrased it (Sect. 4, revised manuscript).

14. **Lines 330.** The authors refer to a spectral cut off of the stray light convolution kernel. If such a cutoff was described in Section 2.2.1 it was not done so in a clear manner that allows the reader to identify what feature of the kernel the authors are referring to here. Are the authors saying the narrower kernel reduces the contributions to Band 1 from Band 2? If so, please say that instead of calling it a cut off.

    We agree with the reviewer that 'cut off' can be a misleading term. We replaced the text " As expected, the spectral far-field cut off of the elliptical straylight convolution kernel mostly affects..." with "*this L1 processing with the elliptical straylight convolution kernel mostly decreases the systematic effects seen at the interface between bands 1-2 (285–300 nm), in both orbits.*" (line 390, revised manuscript)

15. **Lines 339.** Please identify the quantity being discussed. The time difference between two orbits is a constant and cannot become smaller, so the authors must be referring to something else. Correction magnitude?

    We indeed refer to the correction magnitude; in the updated text we use '*temporal increase of the soft calibration correction magnitude*' (e.g. line 384 of the revised manuscript).

16. **Line 370.** The sentence beginning "We do not observe ..." is awkward. I do not understand what the authors are trying to say. Perhaps removing the word "instead" will help. I've noticed several places throughout the manuscript where this word has been unnecessarily introduced.

    We thank the reviewer for the attentive comment. The revised manuscript has an updated text where we tried to avoid using the word "instead" where not necessary.

17. **Lines 372-373.** The authors here conclude on the basis of Figures 12b and 12d that there is a reduction in anomalies. This point is not at all clear from the figures. By some

criteria it could be argued the new version actually looks worse. I recommend revising the figures to more clearly support your conclusion. Or perhaps the authors just need to revise their tropospheric ozone conclusions.

To enhance the clarity of the retrieval conclusions, we edited Sect. 5 (revised manuscript) in the following way:

- Figure 16 (revised manuscript): we narrowed the latitude range to 30S-30N (compared to Figure 12 in the previous version) to avoid the gap in the South Hemisphere (visible in sub-figure (a)) thereby ensuring a more consistent comparison across retrieval versions. In addition, we reordered the sub-figures to emphasize the impact of the L1 v3.0 update on the retrieval, both without and with soft calibration.

  The revised ordering provides a clearer view of whether – and where -tropospheric ozone anomalies are reduced as a result of the L1 v3.0 update. Specifically, the anomalies are mainly reduced in the central part of the swath and at northern latitudes when using L1 v3.0 without soft calibration (panels (e), (f)). Panels (g), (h) show that the L1 v3.0 leads to a clear reduction in variability across latitude bands, with individual anomaly profiles clustering more closely together.

  We further observe that applying the soft calibration (for both L1 data version) introduces larger across-track structures in the anomalies. It is not yet clear whether these structures arise from the soft calibration procedure itself or reflect geophysical effects (e.g. the peaks in the west part of the swath that might be linked to sun-glint conditions).
- In Appendix C (revised manuscript), we present similar plots to those in Figure 16, applying the same data selection as in sub-figure (a), which corresponds to the retrieval with the highest number of non-convergences and data gaps. We note that the anomalies in Appendix C (f), (g), (h) are larger than in the corresponding plots of Figure 16.
- Figure 15 (revised manuscript) presents the plots previously included in our previous responses to reviewer #1. We believe that this figure more clearly illustrates the improvements achieved with L1 v3.0, in particular the enhanced global convergence and the reduced dependence of the retrieval on the soft calibration correction.

18. **Lines 385-387.** Please make it clear that this statement refers to the current operational product.

    Following also the suggestion in question 12, the manuscript (text and figures) has been revised to use the L1 version numbers.

Grammatical

- Line 127: "a part for the spectral region": not present in the revised version
- Line 134: "part of to the instrument": not present in the revised version
- Line 238: "and the ISRF also updated": not present in the revised version
- Line 264: "approach as described in (?)": fixed

- Line 331: "between the orbits is reduced of around": not present in the revised version
- Line 350: "slightly increases of around 0.2% .... it decreases of 1-2%" See also Line 365, 394.: fixed
- Line 374: "as it can be seen from": not present in the revised version
- Line 375: "they decrease of few percents": not present in the revised version
- Line 388: "updated with adjustments regarding": not present in the revised version