

## Response to Referee #2

We would like to thank the reviewer for carefully reading the manuscript and for providing helpful comments, remarks and suggestions. You can find below our responses in red after each individual comment:

**Manuscript:** *Ground-based total ozone column measurements in the Huggins and Chappuis bands using Direct-Sun DOAS observations* **ID:** egosphere-2025-5627

This manuscript presents direct-sun DOAS total ozone column (TOC) retrievals in both the UV (Huggins) and visible (Chappuis) spectral regions using the Delta UV-VIS DOAS system in Thessaloniki. The work is relevant for the AMT community, particularly as it explores the feasibility and performance of visible-band direct-sun ozone retrievals, which remain less established than traditional UV techniques. The manuscript is generally well written, and the reported agreement with collocated Brewer and Pandora observations is good. The study has the potential to make a valuable contribution to atmospheric measurement techniques, especially if methodological choices, calibration assumptions, and statistical aspects are documented more transparently. The comments below are therefore mainly requests for clarification, additional documentation, and improved transparency rather than indications of major methodological errors.

For transparency, I wish to inform the editor and authors that I am actively involved in ozone calibration and intercomparison activities related to Brewer. I believe this background is relevant to the comments provided.

In Section 3.4, the manuscript applies the Bootstrap Estimation (BE) method to derive the ozone reference slant column, representing a novel extension of an approach that has primarily been developed and applied for NO<sub>2</sub> retrievals (e.g. Cede et al., 2006; Herman et al., 2009). Given the fundamentally different vertical distribution, variability, and tropospheric contribution of total ozone compared to NO<sub>2</sub>, this application would benefit from more explicit justification and validation. At present, BE results are shown mainly for the VIS channel (Fig. 4b), while the corresponding UV results are not presented (line 349), and no sensitivity analysis is provided regarding the choice of percentile or AMF binning. Including the UV BE results and a brief sensitivity assessment would strengthen confidence in the robustness of the BE approach when applied to ozone.

We thank the reviewer for this important comment. Sect. 3.4 has been revised to clarify both the methodology and the terminology used for the determination of the O<sub>3</sub> SCD of the FRS. We would like to clarify that the term "Bootstrap Estimation (BE)" was not used accurately in the original manuscript. The analysis in this study employs the Minimum-Amount Langley-Extrapolation (MLE) method (Herman et al., 2009), rather than a bootstrap estimation approach. The manuscript has been revised accordingly to use the correct terminology and to more clearly describe the implementation of this method.

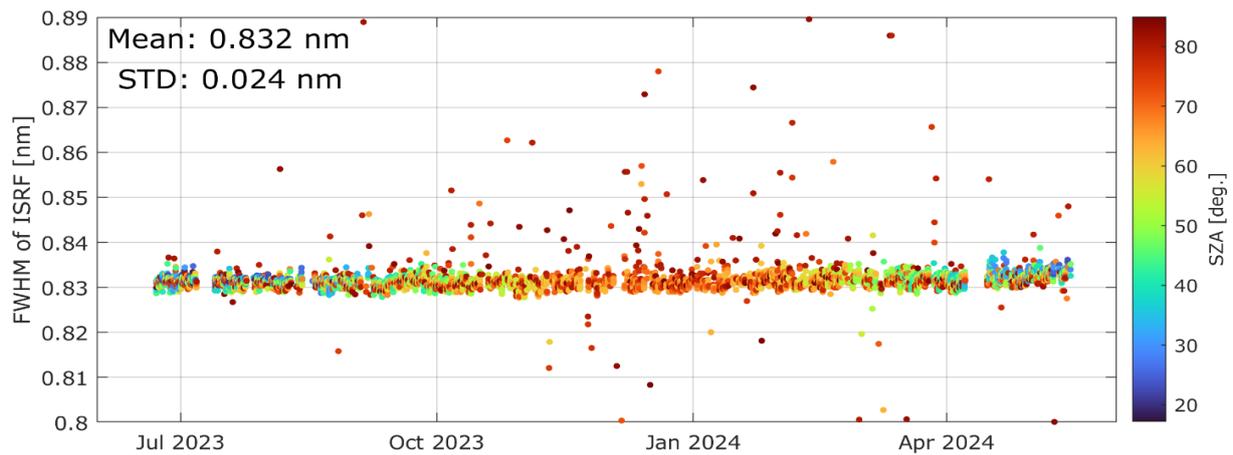
Sect. 3.4 has further been revised to clarify that MLE relies on the assumption that, if TOC was constant throughout all days in the period of study, dSCDs would exhibit a linear dependence on AMF (Eq. 3), and the extrapolated dSCD at AMF=0 would equal the SCD of the FRS. If the minimum vertical column amount of an atmospheric species is constant over part of the dataset, then this will be independent of AMF. Over Thessaloniki, the lowest TOC values typically occur during the summer and autumn months, hence cover the entire range of the AMF annual variability. Given that a sufficient number of

measurements is available for each AMF bin during these periods, the minimum dSCDs are expected to be similar across bins, thereby satisfying the assumption of the MLE method. A sensitivity analysis was performed to assess the robustness of the MLE method with respect to the choice of percentile. Percentiles ranging from the 2<sup>nd</sup> to the 5<sup>th</sup>, in steps of one, were tested. The resulting differences in the estimated SCD of the FRS were found to be up to 0.6% in the VIS and up to 2% in the UV. These differences reflect the statistical nature of the MLE method, which requires a sufficiently large number of measurements to yield stable results. The larger variability observed in the UV retrievals is attributed to the substantially lower number of available data in this spectral range (see Sect. 2.2). Additionally, care must be taken in the selection of the AMF bin width as overly broad AMF bins reduce the number of points used for the extrapolation, while excessively narrow bins may contain too few measurements for a reliable estimation of the selected percentile. UV data have now been included in Figure 4 as panel c. When the same MLE methodology is applied to data in the UV range retrievals, a comparable value of SCD\_FRS=408.54 DU is obtained. However, the UV dataset covers a shorter time period and does not include the summer and autumn months, when TOC typically reaches its annual minimum. Furthermore, at larger SZAs the RMS filtering that is applied (see Sect. 4) removes a substantial number of measurements. As a result, the statistical assumptions of the MLE method are less optimally satisfied for the UV, which leads to a larger sensitivity to the choice of the percentile and the AMF range. In this case, to ensure statistical robustness, the AMF range was restricted to values below 4. Despite these limitations, the close agreement between the UV- and VIS-derived values of SCD\_FRS indicates that the calibration of the reference spectrum is robust.

It is difficult to follow the different dataset lengths of the UV and VIS channels described in Sections 2 and 3. The manuscript would benefit from a summary table indicating the operating periods of each channel together with the main instrumental modifications. While the authors assume that the instrument remained stable over the analysis period, no supporting evidence is provided. Although it is stated that lamp spectra are used for accurate wavelength calibration (line 140), no results are shown demonstrating regular monitoring or otherwise documenting the long-term stability of the spectrograph.

We thank the reviewer for this comment. The manuscript has been revised to clarify the operating periods of the VIS and UV channels and to better document the stability of the instrument over the analysis period. We now explicitly state at the end of Sect. 2.2 that the VIS dataset spans from June 2023 to May 2024 (~11 months), while the UV dataset covers the period from January 2024 to May 2024 (~6 months). This clarification helps the reader to more easily follow the temporal coverage of the two datasets.

In addition, a description of the instrument stability monitoring has been added in Sect. 2.2. The stability of the instrument's slit function was regularly monitored by comparison of measured spectra with a high-resolution solar spectrum, convoluted with a Gaussian function (Fayt and Van Roozendaal, 2001; Aliwell et al., 2002; Danckaert et al., 2017). As shown in the figure below, the slit function remained stable throughout the study period, exhibiting a mean FWHM of approximately 0.83 nm ( $\pm 0.02$  nm). This aligns closely within the 0.85 nm value derived from mercury lamp measurements. Most of the outliers are found at high SZAs, where reduced signal-to-noise ratios affect the slit function retrieval accuracy. The corresponding discussion has been added to Sect. 2.2 of the revised manuscript, while the figure itself was not included.



Although EuBrewNet is cited in Section 2, EuBrewNet ozone data are not used in the analysis presented in Section 4. Including EuBrewNet data—particularly the Version 2 total ozone product—would be a valuable addition, as Version 2 (Rimmer et al 2018) employs updated ozone absorption cross sections consistent with those used in the Pandora and Delta retrievals and applies a similar treatment of effective ozone height and effective temperature. In addition, it should be clarified whether the Brewer data used in the study include the straylight correction introduced during the 2021 calibration (WMO 2024, GAW Report No. 301, p. 43) and implemented on EuBrewNet processing.

We thank the reviewer for this thoughtful suggestion. We agree that the use of EuBrewNet Version 2 TOC product would potentially provide results more consistent with those of Pandora and Delta, as it employs updated ozone absorption cross sections and applies a similar treatment of effective ozone height and effective temperature. As shown in Figure 11 of the manuscript, the largest differences between Delta and Brewer occur at high SZAs ( $> \sim 75^\circ$ ), where stray light effects become increasingly important, especially for the Brewer. Under these conditions, for our single-monochromator Brewer, stray light leads to an underestimation of TOC, as documented in the RBCC-E campaign reports. The positive drift observed in the Delta-Brewer differences at large SZAs is consistent with this known behavior. This indicates that the Brewer measurements are more strongly affected by stray light than the Delta retrievals, even though the measured spectra of Delta are not directly corrected for it. However, version 2 of the EuBrewNet TOC product excludes values for  $AMF > 3$ . This limits the AMF range available for comparison with Delta, which, as previously discussed, is important for our assessment. Nevertheless, we did compare Delta with the version 2 EuBrewNet dataset, and the statistics are practically the same, with a change in the slope from 0.98 to 0.97 and an increase in the mean difference from  $-0.08\%$  to  $-0.23\%$ . Therefore, we decided to keep our in-house processing rather than the EuBrewNet Version 2 data.

It is noted that no stray light correction was applied to the Brewer data used in this study. The stray light correction introduced during the 2021 calibration campaign was calculated as a power function of the  $O_3$  SCD and for our instrument it was derived for SCD values below  $\sim 1400$  DU. At larger SCDs (corresponding typically to larger AMFs), extrapolation of the power fit correction introduces additional uncertainties leading to over-correction for  $SZA > 75^\circ$ . Therefore, to avoid introducing additional assumptions, no stray light correction was applied to the in-house Brewer TOC processing.

The manuscript refers to Pandora “version 1.8.49” in Section 2; however, this designation corresponds to the processor version rather than to a specific ozone product definition.

For clarity and reproducibility, it would be helpful to explicitly state which Pandora ozone product is used (e.g. OUT2) and to briefly summarize the key retrieval assumptions relevant for comparison, such as the spectral window, ozone cross sections, and treatment of effective ozone temperature, with reference to the appropriate PGN documentation. The calibration of Pandora instrument is not commented.

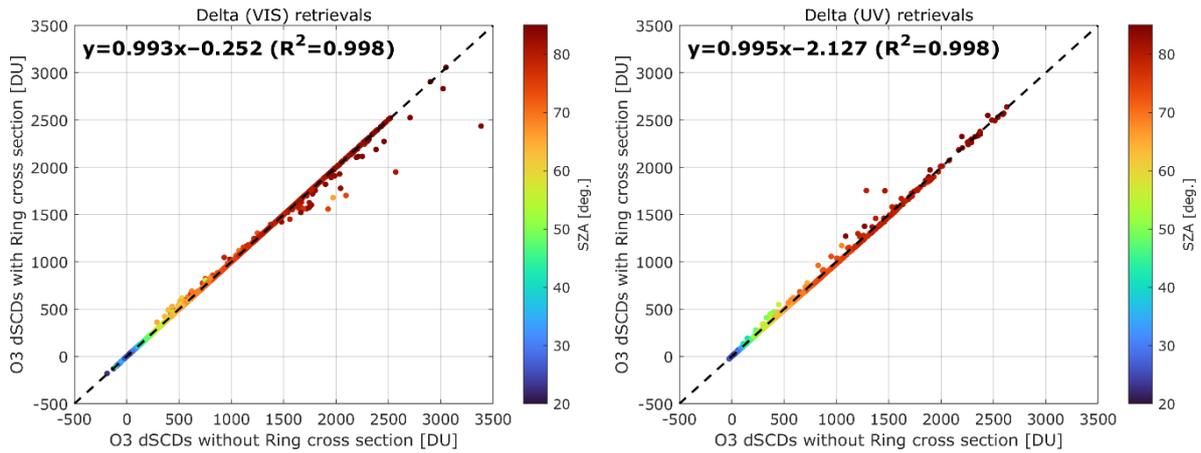
We thank the reviewer for this helpful comment. We agree that "1.8.49" refers to the BlinkP processor version rather than to a specific ozone product definition. Sect. 2.4 has been revised to specify that the PGN TOC product used in this study corresponds to retrieval version "rout2p1-8", which is the current operational PGN TOC product. In addition, we now briefly summarize the key retrieval settings, including the spectral fitting window, the use of a high-resolution extraterrestrial reference spectrum, the O<sub>3</sub> absorption cross sections used in the analysis and the treatment of effective O<sub>3</sub> temperature. We also provide a reference to the corresponding PGN Data Products Readme document for additional retrieval details. Furthermore, the discussion of Sect. 2.4 has been slightly revised to clarify that the calibration procedures applied within the PGN framework are described in the official PGN manuals.

The UV channel inherits the absolute calibration from the VIS channel through the use of a common reference slant column, as described in Section 3.4. While this approach is acceptable, the manuscript would benefit from explicitly stating this assumption and from demonstrating that UV-based calibration estimates are consistent with the VIS-derived value, which is not shown in the current version. In this context, providing Figure 4c (UV Bootstrap Estimation), analogous to Fig. 4b, would significantly improve the work.

We thank the reviewer for this important comment. This point has been addressed in the response above (see first comment). Sect. 3.4 has been revised to clarify the calibration approach, and the corresponding UV results are now included in Figure 4c, demonstrating the consistency with the VIS-derived data.

The manuscript states in Section 3.3 that the Ring effect is not included in the DOAS fitting because direct-sun spectra are used. While this assumption is generally reasonable, a brief justification or sensitivity assessment would improve clarity, particularly in light of the discussion of thin cloud contamination and potential stray light effects in Section 4. Clarifying whether inclusion of a Ring pseudo-cross section has a negligible impact under the observed measurement conditions would strengthen the methodological description.

We thank the reviewer for this comment. To assess the potential impact of the Ring effect on the Delta retrievals, we performed a sensitivity test in which a Ring pseudo-cross section was included in the DOAS analysis. The comparison of O<sub>3</sub> dSCDs retrieved with and without the inclusion of the Ring cross section is shown in the figures below for the VIS (left) and UV (right). The data are colored by SZA.

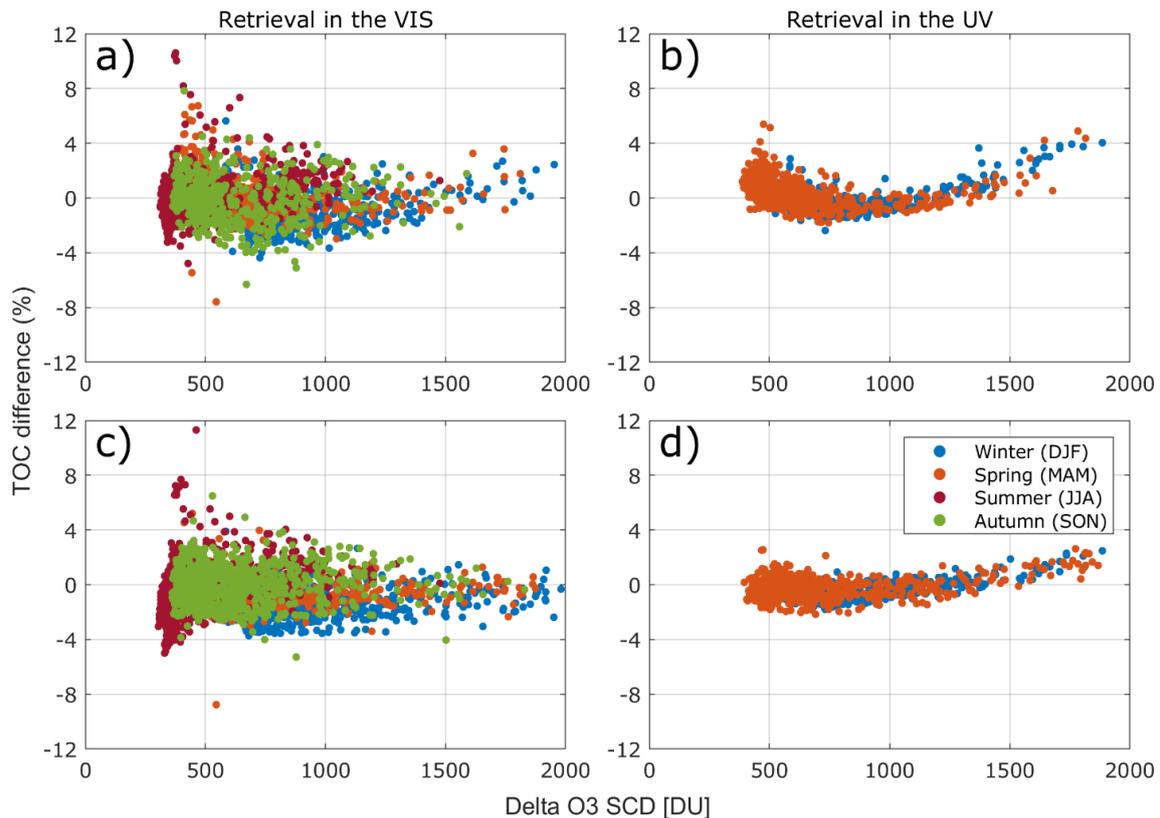


- For the VIS retrievals, the inclusion of the Ring cross section has a negligible effect on the retrieved O3 dSCDs, with a median difference of approximately  $-0.32\%$ , confirming that the Ring effect has a limited impact under direct-sun measurement conditions. Most outliers are found at high SZAs.
- For the UV retrievals, the effect is slightly more pronounced, as expected due to the stronger Ring contribution at shorter wavelengths, but it remains small, overall, with a median difference of approximately  $-1.5\%$ .

Nevertheless, the main focus of this study is on the VIS retrievals. We also note that the Brewer and Pandora instruments used in this study do not include a Ring correction in their operational TOC retrievals. For consistency with the reference datasets used in the intercomparison, we therefore chose not to include a Ring pseudo-cross section in the Delta retrievals.

For completeness, it may be helpful to further explore potential stray light effects by presenting the differences between Delta and the reference instruments as a function of the ozone slant column (SCD), used here as an indicator of high absorption conditions. Such a representation could provide additional insight into whether residual stray light effects contribute to the observed differences, particularly under high SZA or high aerosol conditions, and would complement the existing comparison plots shown in Figures 6–9.

We thank the reviewer for this helpful suggestion. We investigated the dependence of the differences between Delta and the reference instruments as a function of the ozone SCD, which can be used as an indicator of high absorption conditions. The corresponding analysis is provided in the figure below and is consistent with Figure 11 of the manuscript, except that the differences are shown as a function of SCD rather than SZA. Overall, the SCD-based representation yields results similar to those obtained as a function of SZA and essentially provides the same information. However, the SCD-based representation focuses on higher absorption conditions, while most of the observations correspond to typically lower SZAs (and hence lower SCDs). For clarity and to avoid redundancy, we prefer to show the SZA-based representation in the manuscript, as it provides a more direct interpretation of geometry-related effects.



The manuscript compares Delta with Brewer and Pandora separately in Section 4. Including a short Brewer–Pandora comparison, or referencing established agreement at the site, would provide useful context for interpreting the Delta validation and for assessing the consistency of the reference instruments themselves.

We thank the reviewer for this helpful suggestion. A brief comparison of collocated Brewer and Pandora TOC measurements has now been included in the revised Supplement (as we prefer to keep the focus of the paper to the Delta instrument) and a corresponding discussion has been added at the end of Sect. 4.2. The results show a high level of consistency between the two reference instruments at Thessaloniki, with a mean difference of less than 0.4% and a narrow distribution, with most differences confined within approximately  $\pm 1\%$ .

Throughout Section 4, intercomparison results are presented without reporting the number of collocated measurements (N), which is essential for assessing the statistical significance and robustness of the reported correlations and biases. Correlation coefficients and RMS values are given for several comparisons (e.g. Delta–Brewer, Delta–Pandora, UV–VIS) without specifying N, and qualitative statements about higher or lower numbers of collocations are not supported by quantitative values. Reporting N in the text and figure captions, and where appropriate providing confidence intervals for key metrics, would substantially improve transparency and interpretability.

We agree that showing and discussing the number of collocated measurements is essential for assessing the statistical significance and robustness of the reported correlations and biases and the number (N) was already included in the plots below the linear fit equation. All figure captions have been revised to explicitly state that N corresponds to the number of collocated measurements. Additionally, the number of collocations was already provided

in Sect. 4.2, but has now been further clarified, including a quantitative comparison of the number of VIS and UV collocations.

In relation to the calibration methodology described in Section 3.4 and the comparison results presented in Section 4, it would be useful to include uncertainty information for the Delta TOC retrievals. A concise discussion or table summarizing the main sources of uncertainty—such as DOAS fitting residuals, determination of the reference slant column (LE and BE), AMF/effective height assumptions, and sensitivity to atmospheric conditions (e.g. aerosols or thin clouds)—together with indicative uncertainty ranges, would improve the quantitative interpretation of the results shown in Figures 6–9.

We thank the reviewer for this suggestion. An error budget analysis has been included in the revised manuscript in a new section (3.5), where the main sources of uncertainty affecting the Delta TOC retrievals in both the UV and VIS ranges are discussed, along with a summary table.

Table 1 summarizes the DOAS fit settings used in the analysis; however, several key terms are not sufficiently defined in the manuscript. In particular, the meaning and implementation of the  $I_0$  correction, the intensity offset term, and whether the absorption cross sections are orthogonalized are not clearly explained. Providing brief clarifications of these elements—either in the table caption or in the accompanying text—would improve transparency and reproducibility.

Sect. 3.2 has been revised to include a clear explanation of the “ $I_0$ -correction” and “intensity offset” terms, as well as their physical meaning in the DOAS retrieval. We now explain that the  $I_0$  effect originates from inconsistencies between laboratory absorption cross sections measured with a smooth light source and the structured atmospheric absorption spectra recorded with lower spectral resolution, while the intensity offset compensates for potential stray light or dark signal that is not effectively removed from the measured spectrum. The use of orthogonalized absorption cross sections was already specified in Table 1.

Figure 4 would benefit from clearer labeling and explanation. It is not immediately evident whether the data shown correspond to the VIS or UV channel, particularly in panel (b). In addition, panel (a) refers to a specific clear-sky day used for the Langley analysis, but this day is not explicitly identified in the figure or clearly linked to the data shown in panel (b). The caption should explicitly state the spectral channel shown in each panel and identify the specific day used for the Langley extrapolation.

We thank the reviewer for this helpful comment. Figure 4 has been revised to improve clarity and help interpretation. In panel b, the measurements corresponding to the clear-sky day used for the Langley extrapolation in panel a are now explicitly highlighted. In addition, text has been included in the lower right corner of each panel to clearly indicate the dataset that is used (VIS or UV).

Minor points:

the use of MLS–GMI climatology for the effective height in lines 305–320 could be briefly justified, as Gröbner et al. (2021) uses local ozonesonde profiles.

Sect. 3.3 has been revised to clarify the use of the MLS–GMI climatology for the effective height. This approach was adopted because routine ozone soundings are not available at Thessaloniki, and therefore local ozonesonde profile measurements cannot be used.

Figure 7, please clarify how the Brewer slant column is calculated (e.g.  $SCD = TOC \times AMF$ , and which AMF is used);

Sect. 4.2 has been revised to clarify that for the Brewer and Pandora measurements, the AMFs provided in their respective operational products are used. The calculation of the SCD using the formula  $SCD = TOC \times AMF$  (Eq. 2) is described at the beginning of Sect 4.2.

Figure 9 appears largely redundant with Figure 8 and could potentially be merged;

Even though Figs. 8 and 9 show very similar results, they present comparisons with the two independent reference instruments (Brewer and Pandora). Presenting these comparisons separately allows for a clearer evaluation of the Delta retrieval performance against each reference dataset. Since the intercomparison with both instruments constitutes a core result of this study, we believe that retaining the figures separately improves the clarity of the presentation.

Figure 10; Include UV data, is not shown on the paper and can illustrate the different periods used on the paper.

We thank the reviewer for this suggestion. Figure 10 has been revised to include the UV TOC retrievals in addition to the VIS data. This addition allows for a clearer illustration of the different time periods covered by the UV and VIS measurements and improves the overall completeness of the dataset presentation. The corresponding discussion in Sect. 4.2 has been updated accordingly.

line 182 (WMO 2024), please confirm whether updated stray light corrections reported in recent Brewer calibration reports were applied to the Brewer 005 data used in this study.

Please see our response to the comment on the usage of EuBrewNet data.