

Review: Validation of ATLID stratospheric aerosol retrievals using SAGE III/ISS and ground-based lidar observations

The purpose of this paper is to validate the stratospheric aerosol observations from ATLID on EarthCARE through comparison with observations from SAGE III/ISS and two ground-based lidars. Overall, the paper is well written and an important contribution for any users of the ATLID data. I recommend publication after minor revisions.

We sincerely thank the referee for the positive assessment of our manuscript and for the very constructive suggestions, which helped us to improve the paper. We have addressed all comments and detailed our revisions below.

I have two main suggestions:

1. Choice of lidar ratio: the lidar ratio directly scales the backscatter coefficient when calculating the extinction, so using an accurate value is important. The manuscript mentions that a full analysis of the lidar ratio choice is a topic of future work. However, the method this paper proposes to anyone who wishes to use stratospheric ATLID data is highly dependent on the lidar ratio, so I think the impact of the lidar ratio on the biases with SAGE should be explored further here. At the very least, it would be helpful to see the impact of varying the lidar ratio within a reasonable range on the bias between the ATLID and SAGE extinctions for some example profiles, like those in Figure 3. This would also help with verifying the statement on lines 352-354, which suggests that the larger bias in the NH is due to smoke particles that have a lidar ratio other than 50 sr

We agree that the effect of the lidar ratio deserves a more quantitative treatment. First, we would like to clarify that our methodology does not simply assume a constant 50 sr lidar ratio across all profiles. As detailed in Section 2.2, we utilize the ATLID L2-derived lidar ratio wherever it is available and falls within physically realistic bounds (30-90 sr), relying on the 50 sr fallback assumption only when the L2 retrieval fails or returns unphysical values.

Following the referee's suggestion, we have updated Figure 3 to make this approach perfectly transparent and to illustrate the impact of the lidar ratio. We have added a secondary axis (green curve) that displays the actual L2-derived lidar ratio used in our calculations. Furthermore, we added a sensitivity envelope (40-60 sr for background/volcanic profiles, and 40-70 sr for boreal smoke cases) strictly limited to the altitudes where the L2 processor does not provide a valid lidar ratio, meaning the 50 sr fallback was applied.

Regarding the statement on lines 352-354, the revised Fig. 3 gives a more nuanced picture. In the smoke case of Fig. 3c, the L2 provides a lidar ratio of about 60 sr within the smoke layer, and this value is used in the L1-derived extinction; the L1 sAOD (0.036) is then higher than the SAGE III/ISS sAOD (0.024). In the smoke case of Fig. 3d, the L2 provides no valid lidar ratio within the smoke layer (13-20 km), where the 50 sr fallback is used; the L1 sAOD (0.012, or 0.009-0.016 for 40-70 sr) is close to the SAGE III/ISS value (0.013). These examples show that the lidar ratio of smoke

particles alone cannot explain the larger negative bias found in the Northern Hemisphere. Section 4.1 has been revised accordingly: the lidar ratio of smoke is mentioned as one possible contribution, and a dedicated analysis comparing the summer 2025 smoke period with the rest of the record is left for future work.

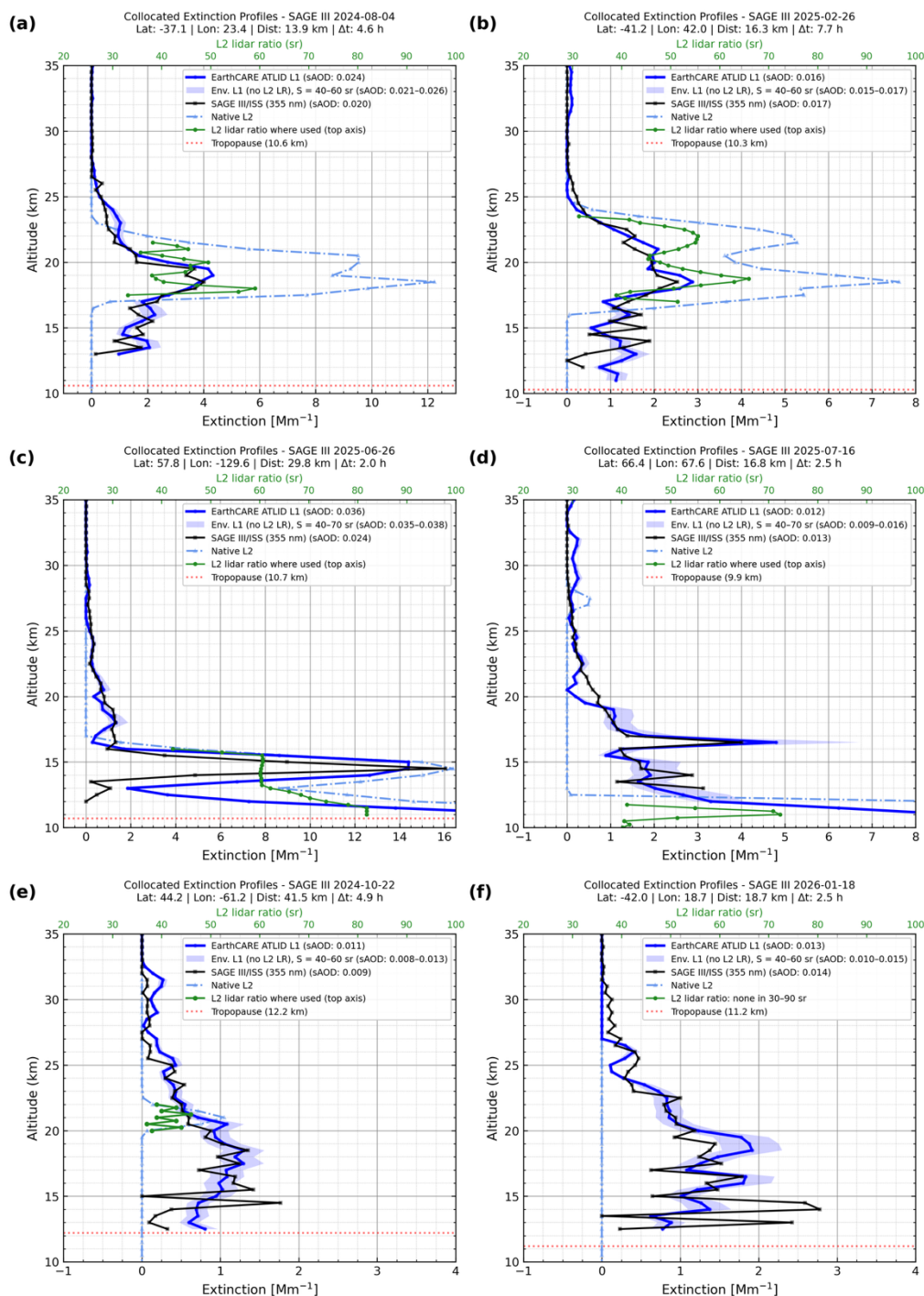


Figure 3. Collocated extinction profiles from SAGE III/ISS, ATLID L1, and native ATLID L2 for selected case studies: (a, b) Southern Hemisphere Ruang volcanic plume, (c, d) boreal wildfire smoke, and (e, f) background aerosol conditions. The green curve (top axis) shows the actual L2-derived lidar ratio applied in the L1 calculation where valid (30-90 sr). The shaded blue envelopes represent the L1 extinction sensitivity to the assumed lidar ratio, using a range of 40-60 sr for the volcanic and background cases (panels a, b, e, f) and 40-70 sr for the smoke cases (panels c, d), and are shown exclusively at altitudes where a valid L2 lidar ratio is unavailable. The horizontal dotted pink line indicates the local tropopause altitude. Dist is the great-circle distance between the SAGE III/ISS tangent point and the closest of the averaged ATLID profiles (all within 50 km), and Δt is the time difference between the SAGE III/ISS event and each collocated ATLID overpass.

2. Organization of results (Section 3): As a user of the ATLID data, my initial thought would be to use the L2 product, and I believe this paper is the first to clearly demonstrate the problems with the L2 aerosol retrieval in the stratosphere. Therefore, I think it would be helpful to show the L1/L2 comparison at the start of the paper, which would then motivate the detailed validation of the L1-derived extinction. This could be done by moving Figures 3 and 7 and the surrounding discussion to the start of Section 3, so that the issues with the L2 product are immediately clear.

We considered this suggestion carefully, but we prefer to keep the current order of Section 3. The L2 sAOD comparison (Fig. 7) uses the same matchups as the L1 comparison (Fig. 4) and is discussed relative to it, so it is easier to follow once the L1 results have been presented. In addition, the limitations of the L2 product are already stated in Sect. 2.2 (lines 65-68, rewritten following Referee #3), where they motivate the L1-based retrieval. To make them clear from the start of the results, we have added a short paragraph at the beginning of Section 3 that summarizes them, with pointers to Sects. 3.2.2 and 3.6.

Specific comments:

- Line 62: An explanation of the baselines and the difference between them would be helpful.

BA, BB and BC refer to the EarthCARE processing baselines, i.e. successive versions of the ESA ground-processing chain used to generate the ATLID products. EarthCARE processing baselines are identified by two capital letters: the first letter denotes a major processing/reprocessing generation, while changes in the second letter indicate subsequent incremental updates. Baseline BA corresponds to the first homogeneous reprocessing of the EarthCARE data record using the B-generation processors, while BB and BC contain subsequent updates. We have clarified this terminology in the revised manuscript. Baars et al. (2026) gives a particularly clear definition and was cited here.

- Section 2.2: I think it is worth providing more details on the method used to calculate the extinction coefficient, even if it is a repeat of Khaykin et al (2026), given that one of the main recommendations in this paper is to derive extinctions from the L1 observations instead of using the official ATLID product.

We agree. Given that the L1-based derivation is the core recommendation of this paper, we have expanded Section 2.2 to provide a more detailed, step-by-step description of the unattenuated particulate backscatter retrieval from the molecular reference profile, rather than relying solely on the citation.

- Comparison with ground-based observations: What is the impact of using a fixed lidar ratio of 50 sr to convert these observations to extinction/SAOD, when the ATLID observations are (at least sometimes) converted to extinction using the retrieved L2 lidar ratio?

For the sAOD comparison against ground-based lidars, the ATLID L2 lidar ratio is used only where it lies within 30-90 sr; elsewhere, ATLID uses the same 50 sr value as the conversion of the ground-based profiles, so the two conversions differ only at these levels (see also our response to Referee #3, comment 2). More importantly, to rigorously isolate the performance of the ATLID retrieval from any lidar-ratio assumptions, our profile-level comparison against OHP and TMF (Section 3.5) is performed directly on the backscatter coefficients. This independent backscatter comparison reproduces the same bias structure, confirming that the bias originates in the Level-1 scattering-ratio retrieval itself rather than in the lidar ratio conversion. We have clarified this distinction in the text.

- Are data points below the tropopause removed before calculating the relative difference in extinction? (Figures 1 and 2)

Yes. Following suggestions from Referees #1 and #3, we have updated Figs. 1b and 2 to display the relative difference as a function of altitude *relative to the local tropopause*, strictly retaining only levels at least 1 km above the local tropopause (consistent with the sAOD integration limits).

- Figure 2 caption: the meaning of daytime and nighttime SAGE III occultations is not clear, given that the measurements are at sunrise and sunset. Also, the text on line 189 says that the split is based on the EarthCARE overpass time, not the SAGE III occultation time. Please clarify this.

The referee is correct. The day/night split refers exclusively to the illumination conditions of the ATLID overpass (based on the solar elevation angle), not the SAGE III/ISS occultations. SAGE III/ISS events are sunrise or sunset occultations with the same geometry. The caption of Fig. 2, which incorrectly referred to "SAGE III/ISS occultations", has been corrected to "daytime and nighttime ATLID overpasses", and the text in Section 3.2.1 has been revised to remove this ambiguity.

- Line 194: Was it unexpected for the day and night bias to have a similar vertical structure? If so, this should be explained, and if not, I would remove the word "remarkably".

It was not unexpected: the vertical structure of the bias is mainly governed by factors common to both subsets, such as the vertical distribution of the aerosol loading and the altitude relative

to the tropopause. The word "remarkably" has been removed, and the paragraph has been rewritten for the new tropopause-relative version of Fig. 2 (see our response to Referee #1, minor comment 5).

- Figure 3 caption: this figure is very insightful, but discussions of the bias between L1/L2 and SAGE should be in the main text, not the figure caption.

Done. We have streamlined the caption and moved the detailed profile-by-profile analysis into the main text (Sect. 3.2.2).

- Section 3.3: Do you have any theories as to why ATLID overestimates the SAOD in cases with more aerosol?

We thank the referee for this question. We see two possible contributions, which are now discussed in Sect. 3.3.

(i) Coverage of the lowermost stratosphere. Both sAODs are integrated from the tropopause + 1 km to 30 km, but in dense layers the lowest SAGE III/ISS levels are more likely to be removed by the quality screening (e.g. aerosol–cloud mixture flag) or not retrieved at all, as the transmission along the line of sight decreases. The ATLID L1 profile retains these levels. When a dense layer lies just above the tropopause, as in Fig. 3c, where ATLID shows large extinction values between the tropopause and 12.5 km that are absent from the SAGE III/ISS profile, the ATLID sAOD therefore includes a contribution that is missing from the SAGE III/ISS sAOD.

(ii) Sampling geometry. SAGE III/ISS integrates along a limb path of a few hundred kilometres, whereas the ATLID extinction is an average of nadir profiles within 50 km. For compact plumes, ATLID can sample the core of the plume while SAGE III/ISS sees it diluted along its line of sight, which produces large positive differences; in the opposite case, the negative difference is limited by this dilution. This favours a positive tail at high aerosol loading.

- Figures 4 and 7: displaying the comparison with SAGE III in a histogram form, as was done for Figure 1a, would be helpful for understanding these figures. Currently, it is difficult to determine the range of SAOD values where the agreement is best.

We followed this suggestion. In the revised Figs. 4 and 7, the SAGE III/ISS matchups are shown as 2D histograms (number of matchups per 0.001×0.001 bin, logarithmic colour scale), as in Fig. 1a. The few TMF and OHP matchups are still shown as individual symbols. To show directly where the agreement is best, we also added the median ATLID sAOD in 0.002-wide classes of SAGE III/ISS sAOD (classes with at least five matchups), with the interquartile range as error bars.

Above about 0.02, few matchups remain and the interquartile range widens. For L2, about half of the matchups (321 of 652) have sAOD_L2 = 0, so the class medians in Fig. 7a stay close to zero over most of the sAOD range. When these cases are excluded (Fig. 7b, N = 331), the class medians fall on both sides of the 1:1 line with wide interquartile ranges, and no sAOD range of consistent agreement emerges. This is now stated in Sects. 3.3 and 3.6.

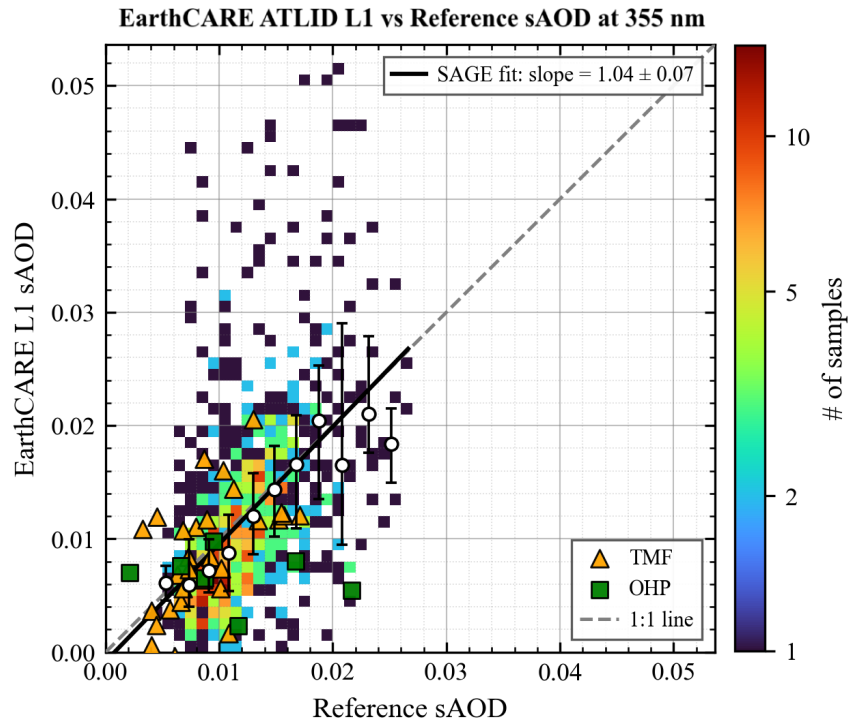


Figure 4. ATLID L1 sAOD versus reference sAOD, integrated from the ATLID-derived tropopause +1 km to 30 km. SAGE III/ISS matchups ($N = 895$) are shown as a 2D histogram (number of matchups per 0.001×0.001 bin, logarithmic colour scale); TMF (triangles) and OHP (squares) matchups are shown individually. White circles give the median ATLID sAOD in 0.002-wide classes of SAGE III/ISS sAOD (classes with at least five matchups), with the interquartile range as error bars. The dashed grey line is the 1:1 reference. The solid black line shows the linear regression fit to the SAGE III/ISS matchups: slope = 1.04 ± 0.07 (± 1 standard error), intercept = $(-0.8 \pm 0.9) \times 10^{-3}$, $R = 0.67$.

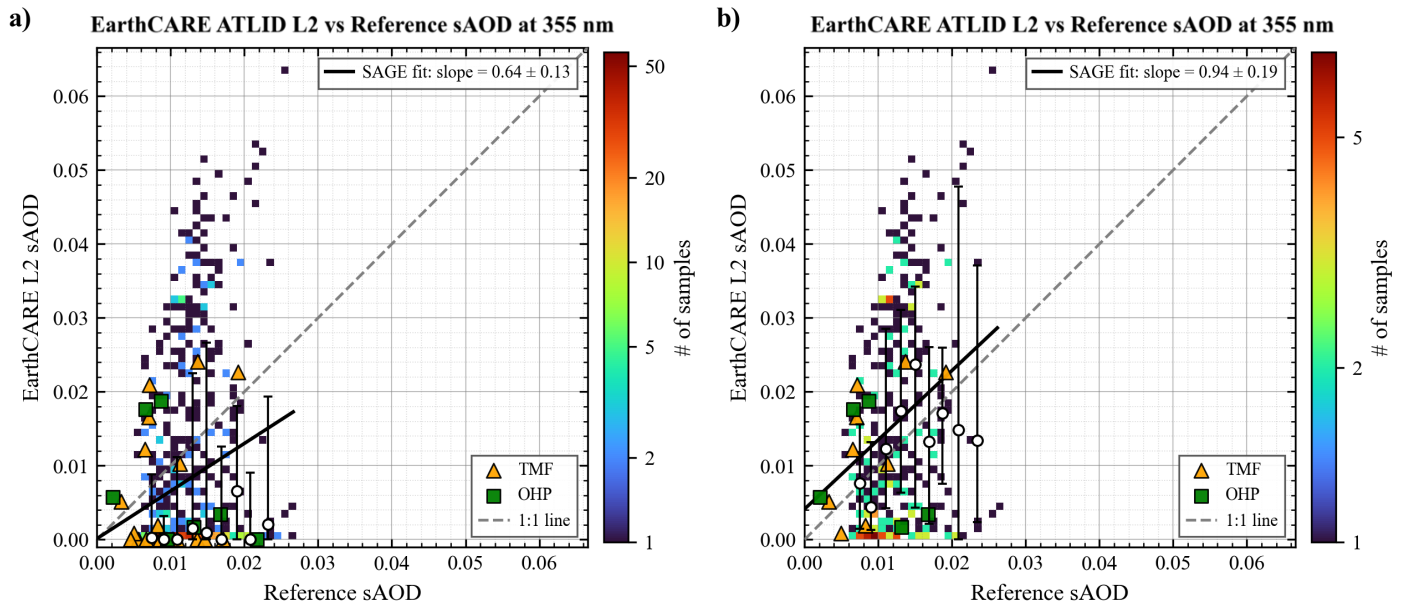


Figure 7. Native ATLID L2 sAOD versus reference sAOD, integrated from the ATLID-derived tropopause +1 km to 30 km. SAGE III/ISS matchups are shown as a 2D histogram (number of matchups per 0.001×0.001 sAOD bin, logarithmic colour scale), and TMF (triangles) and OHP (squares) matchups are shown individually. White circles give the median ATLID L2 sAOD in 0.002-wide classes of SAGE III/ISS sAOD (classes with at least five matchups), with the interquartile range as error bars. **(a)** All colocated matchups, including cases for which the L2 processor returns no stratospheric retrieval ($\text{sAOD_L2} = 0$, in the bottom row of the histogram, along the horizontal axis). The linear regression fit to the SAGE III/ISS matchups ($N = 652$; solid black line) yields a slope of 0.64 ± 0.13 and an intercept of $(0.1 \pm 1.7) \times 10^{-3}$ ($R = 0.19$). **(b)** Same comparison but restricted to actual retrievals ($\text{sAOD_L2} > 0$); the null retrievals ($\text{sAOD_L2} = 0$) shown along the horizontal axis in panel (a) are here excluded from the display, the class medians and the fit. The regression to the SAGE III/ISS matchups ($N = 331$) then yields a slope of 0.94 ± 0.19 with a positive intercept of $(4.1 \pm 2.6) \times 10^{-3}$ ($R = 0.26$). In both panels, the dashed grey line is the 1:1 reference, the slope of the fit is given in the upper-right corner, and quoted uncertainties are one standard error. The L2 comparison includes fewer SAGE III/ISS matchups than the L1 comparison of Fig. 4 because the daily L2 files contain fewer profiles than the L1 files (about 6,400 vs. 22,750 per day).

- Line 267: it looks like the gray shaded region is the broadest in the NH tropics/midlatitude (10N-30N) and narrowest at higher latitudes. Is this sentence phrased backwards?

The referee is right, the sentence was phrased backwards. The ± 2 SE envelope is narrowest where the collocations are dense (mid-latitudes) and broader where sampling is sparse. The text has been corrected to accurately reflect the figure.

Specific comments:

- Line 148: Change “the following metrics” to “the metrics in Table 1.”
- Line 154: remove comma after “independent”
- Section 3.2.2: Include “=” sign between “lat” and the value in all cases.

Done. We have implemented all three technical corrections in the revised manuscript.