Response to Referee #1 Comments

Validation of ceilometer aerosol profile retrievals using sun–sky photometer and balloon-borne in situ measurements

General comments:

This manuscript presents an evaluation of two ceilometer-based aerosol retrieval approaches, i.e. the Klett–Fernald backward inversion and a forward iterative method, using a combination of AERONET, GRASPpac, and COBALD balloon-borne observations. The study addresses an important topic in atmospheric remote sensing, particularly in the context of operational ceilometer networks, and provides a valuable multi-instrument validation across three European sites.

The work is generally well structured and demonstrates that the forward iterative method improves retrieval performance, especially under high aerosol load and coarse-mode conditions. The inclusion of independent in situ measurements (COBALD) is a notable strength and enhances the credibility of the validation.

However, some methodological limitations and missing analyses currently weaken the robustness of the conclusions. In particular, the manuscript lacks a rigorous uncertainty assessment, relies on strong assumptions (e.g., fixed lidar ratio), and occasionally overstates its findings. For these reasons, I recommend major revision before the manuscript can be considered for publication in ACP

We sincerely thank the referee for the time and suggestions which contributed to improving the quality of this work. All the points raised have been carefully considered and the revisions have been introduced in the manuscript. All answers in detailed are written in blue.

Uncertainty Quantification

A major shortcoming is the absence of a quantitative uncertainty analysis. Key sources of uncertainty, such as the assumed lidar ratio, calibration constant (CL), and spectral conversion using Ångström exponent, are only discussed qualitatively.

The forward method's performance is strongly dependent on calibration stability, yet no uncertainty propagation is provided. Similarly, the spectral harmonization between 1064 nm and 940 nm introduces additional uncertainty that is not quantified.

I suggest to Include a sensitivity or uncertainty analysis (e.g., varying lidar ratio, calibration constant, and Ångström exponent) and assess their impact on retrieval accuracy.

We thank the referee for this valuable suggestion. To address the main sources of uncertainty identified in the review, we have expanded the uncertainty discussion in the revised manuscript and included additional sensitivity analyses.

First, to address the impact of the assumed lidar ratio on the retrieval accuracy, we have added a new methodology (new Sect 3.2) a dedicated sensitivity analysis (new Sect. 4.3.2 and Figs. 8 and 9), in which the backward and forward AOD retrievals are re-evaluated against AERONET using four lidar ratio configurations: fixed values of 30, 50, and 70 sr, and a daily variable estimate derived from IFS–CAMS aerosol composition forecasts (S_a IFS). The analysis is performed across the same aerosol regime classification used throughout the study and at all three sites, allowing the propagation of the lidar ratio assumption into the retrieved AOD to be quantified in terms of mean deviation, standard deviation, and RMSE.

Second, we assessed the uncertainty associated with the Ångström exponent conversion by comparing the altitude-resolved transformation used in this study with an alternative approach based on a single profile-mean Ångström exponent (Line 361). This comparison provides an estimate of the uncertainty introduced by the spectral harmonization procedure used to compare ceilometer and COBALD observations.

Finally, regarding the calibration constant (CL), the E-Profile framework provides an uncertainty estimate of 25% for the attenuated backscatter product, derived from previous network-wide intercomparison exercises. Since CL is the factor relating the raw signal to the attenuated backscatter, this uncertainty effectively includes the contribution of the calibration constant rather than treating it as a separate term; for a more detailed discussion of these uncertainties, we refer to Li et al. (2021), referenced in the manuscript. Although a dedicated propagation of calibration uncertainty into the retrieved backscatter and AOD products is not currently available within the E-Profile processing chain, the calibration constant is operationally determined within the E-Profile processing framework and is therefore not available as an independent retrieval parameter for systematic perturbation tests. A sensitivity test of the calibration constant therefore lies beyond the scope of the present work and will be addressed in future studies. We have explicitly clarified this point in the revised manuscript and acknowledge calibration uncertainty as one of the dominant sources of uncertainty affecting ceilometer aerosol retrievals.

To clarify all these topics, we added the following lines to the manuscript.

New Sect 3.2:

Line 158: "Section 3.2 describes the methodology used to derive lidar-ratio estimates from the Copernicus Atmosphere Monitoring Service (CAMS) forecasts, based on the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF).

Regarding the lidar ratio sensitivity analysis from (from 284 to 327 line):

Line 284: “The aerosol lidar ratio is a critical parameter in elastic-backscatter aerosol retrievals, as it links aerosol backscatter and extinction coefficients. Since single-wavelength ceilometers cannot independently retrieve Sa, its value must be prescribed from external information. Although a fixed lidar ratio of 50 sr was adopted as the baseline configuration in this study, a variable lidar ratio derived from CAMS aerosol composition forecasts was also estimated to evaluate the sensitivity of the retrievals to this assumption and to assess its potential impact under different aerosol regimes.”

Line 327: “This approach yields a daily, site-specific lidar ratio that reflects the prevailing aerosol composition as represented in the CAMS forecast across all three stations (Granada, Payerne, and Lindenberg). This lidar ratio is subsequently used in a sensitivity analysis together with fixed lidar-ratio values applied throughout the study period, particularly given the contrasting aerosol regimes across the three sites, ranging from mineral dust and anthropogenic pollution at Granada to predominantly fine-mode continental aerosol at Payerne and Lindenberg. Nevertheless, several limitations should be noted. The CAMS AOD at 550 nm is used as a proxy for the optical weighting, whereas the ceilometer operates at 1064 nm; spectral differences in species-specific AOD contributions may therefore introduce a bias in the weighted lidar ratio. Additionally, a single near-surface RH value at 1000 hPa is used to represent the column-mean hygroscopic state, whereas the vertical RH profile can vary substantially, particularly in the presence of elevated aerosol layers such as Saharan dust plumes.

Regarding the Angström exponent assumption:

Line 361: “To estimate the magnitude of the uncertainty associated with the Ångström exponent, we compare profiles converted with the altitude-resolved Ångström exponent against those converted with a single profile-mean value. For the two COBALD case studies, this comparison yields a column-mean difference of approximately 5.6 ±3.5,% for the Payerne dust case (25 June 2019) and of about -2.6 ± 6.2 % for the Lindenberg multi-layered case (27 August 2020); in the latter, the stronger vertical variability of the Ångström exponent produces a larger dispersion of the differences (relative standard deviation ~6 %) around this small column-mean value. These values indicate that the uncertainty associated with the Ångström exponent conversion should be accounted for when interpreting the profile comparisons.”

Regarding lidar ratio (lines 887 – 938):

Line 887: " The column-integrated evaluation in Sect. 4.3.1 was carried out using a single fixed lidar ratio of 50 sr to convert backscatter into extinction. Although this value produced good overall agreement with AERONET, particularly for the forward retrieval, the residual deviations identified under high aerosol load, and most markedly under coarse-mode-dominated conditions, were attributed to the limitation of applying a single fixed lidar ratio across the contrasting aerosol regimes sampled at the three sites.

Line 946"In summary, the fixed lidar ratio of 50 sr provides robust results across across all sites and aerosol regimes, with the lowest overall RMSE for the full dataset.Under coarse-mode dominated conditions, the fixed lidar ratio of 50 sr tends to underestimate AOD while under fine-mode dominated conditions it shows a small positive bias, reflecting the inherent limitation of a single fixed value across diverse aerosol regimes. The IFS-CAMS derived lidar ratio reduces these regime-dependent deviations, particularly under coarse-mode dominated conditions at Granada and Payerne, where it achieves near-zero mean deviations (-0.006 ± 0.053) and RMSE reductions of up to 38% compared to the fixed lidar ratio of 50 sr. At Lindenberg under coarse-mode conditions, no clear advantage of IFS-CAMS-derived lidar ratio over the fixed lidar ratio of 50 sr is observed, though this result should be interpreted with caution given the limited number of cases (N = 63). For the backward method, the lidar ratio choice has a secondary impact compared to the reference altitude selection, which remains the dominant source of uncertainty under high aerosol load conditions regardless of the assumed lidar ratio value."

Assumption of Constant Lidar Ratio

The use of a fixed lidar ratio (50–52 sr) throughout the analysis is a critical limitation. Aerosol lidar ratio varies significantly with aerosol type, and this assumption directly affects extinction profiles and derived AOD.

Provide justification for the chosen value and evaluate the sensitivity of results to realistic lidar ratio variability. This is particularly important given the study's focus on different aerosol regimes.

We thank the referee for this comment. We would like to clarify an important distinction that may have been unclear in the manuscript: two different lidar ratio values are used for two fundamentally different purposes, and it is important not to conflate them.

The first lidar ratio (Sa = 52 sr) is used exclusively during the Rayleigh calibration procedure to determine the system calibration constant CL. This value is prescribed by the E-Profile network and applied uniformly across all CHM15k instruments within the network to ensure consistency and comparability of calibration constants across stations. This choice is not made by the authors but is a network-level decision, and its impact is limited to the calibration step rather than the aerosol retrieval itself.

The second lidar ratio (Sa = 50 sr) is used to solve lidar equation and retrieved backscatter and extinction profiles. The chosen value represents a reasonable central estimate for mixed continental conditions at 1064 nm and is consistent with previous ceilometer studies (Brunamonti et al., 2021). Literature values at 1064 nm range from approximately 45-70 sr for Saharan dust (Haarig et al., 2022; Shin et al., 2018) and 70-95 sr for aged biomass-burning smoke (Haarig et al., 2018). To explicitly address the referee's concern regarding lidar-ratio variability, we have now included a dedicated sensitivity analysis using (S_a = 30), 50, and 70 sr, together with a variable lidar ratio derived from IFS-CAMS aerosol composition forecasts. These results confirm that the optimal lidar ratio is regime-dependent and demonstrate the sensitivity of the retrieved AOD to the assumed lidar-ratio value. These results are now discussed explicitly in the revised manuscript, and the use of a fixed value is acknowledged as an inherent limitation of single-wavelength elastic backscatter systems, which cannot independently retrieve Sa without additional measurements.

This limitation further highlights the value of the GRASPpac approach, which retrieves (S_a) dynamically within the inversion framework (Román et al., 2018), thereby reducing the reliance on externally prescribed lidar-ratio values.

Use of GRASPpac as Reference

GRASPpac retrievals are treated as a reference dataset, yet they are themselves inversion products with inherent assumptions (e.g., vertically homogeneous aerosol properties). This limitation is not sufficiently discussed.

Clarify the role of GRASPpac as a comparative dataset rather than a ground truth and discuss its uncertainties.

We thank the referee for this comment. We agree that this distinction was insufficiently discussed in the original manuscript. To clarify the role of GRASPpac within this study, we have revised the manuscript to explicitly state that GRASPpac is used as an independent comparative dataset rather than as a ground-truth reference. We added the following paragraph in the manuscript to make a clear message.

Line 1: "This study evaluates two approaches for retrieving aerosol properties from ceilometer observations, using aerosol optical depth (AOD) from AERONET and synergistic aerosol profiles obtained by combining the AERONET sun–sky photometer and ceilometer measurements through the GRASPpac algorithm reference." (reference word was subtracted)

Line 423: "While $GRASP_{pac}$ is used here as a high-quality comparative dataset, it should be noted that it is itself an inversion product subject to uncertainties, particularly those arising from the assumption of vertically homogeneous intensive aerosol properties, which may introduce biases under vertically stratified conditions. Its retrievals also depend on the quality of sun-photometer data inputs and the ceilometer range-corrected signal, so errors in either propagate into the reference profiles. We note that the GRASP framework is designed to provide dynamic estimates of these retrieval uncertainties, accounting for the propagation of both random and systematic errors, and

Herrera et al. (2022) showed these estimates to be conservative upper bounds on the actual retrieval errors. For the extinction and backscatter coefficients at 1064 nm, Román et al. (2018) report a synthetic-test accuracy (bias) of −5 % and −11 % and an uncertainty (standard deviation) of 21 % and 31 %, respectively, and find the GRASP-estimated uncertainty to be representative of the real one across all aerosol loads and solar zenith angles. As with most remote-sensing inversions, retrieval accuracy is reduced under low aerosol loading, which supports the interpretation of $GRASP_{pac}$ as a particularly reliable reference under the high-load, coarse-mode-dominated conditions that are the focus of this study.

Line 980: "an important comparative dataset for evaluating ceilometer-based algorithm".

Statistical Analysis

The statistical evaluation is limited to mean and standard deviation. No confidence intervals, significance testing, or correlation analyses are presented. Given the variability in the results (e.g., Tables 2–3), stronger statistical support is required.

Include additional statistical metrics (e.g., confidence intervals, correlation coefficients) and assess the significance of differences between methods.

We thank the referee for this valuable suggestion. We agree that the inclusion of additional statistical metrics strengthens the comparison between retrieval methods and provides a more comprehensive assessment of their performance. To address this point, we have expanded the statistical analysis throughout the Results section by incorporating correlation coefficients, regression slopes, intercepts, and root-mean-square errors (RMSE) in addition to the previously reported metrics. These additional statistics allow a more rigorous evaluation of the agreement between datasets and a clearer quantification of the differences between the backward and forward retrieval approaches.

Line 525: "The corresponding scatter plot provided in the Supplement (Fig. S1) further supports these findings. Although both ceilometer retrievals correlate strongly with the COBALD observations (r = 0.95 and 0.94 for the backward and forward methods, respectively), the similar correlation coefficients conceal important differences in retrieval performance. The backward retrieval yields RMSE = 0.29 Mm−1sr−1, a slope of 0.66, and an intercept of 0.01 Mm−1sr−1; the slope below unity suggests an underestimation of the aerosol backscatter signal that becomes increasingly pronounced at higher backscatter values. This behaviour is consistent with the increasing difficulty of identifying a suitable aerosol-free reference altitude under optically thick dust conditions, which can affect the stability of the backward integration. In contrast, the forward retrieval achieves RMSE = 0.15 Mm−1sr−1, a slope of 0.92, and an intercept of

0.01 Mm−1sr−1, reducing the RMSE by nearly 50 % and showing closer agreement with the COBALD reference across the full range of observed backscatter values."

Line 558: "This contrast is further quantified in the scatter analysis (Fig. S2 in the Supplement). The backward retrieval yields r = 0.51, RMSE = 2.11 Mm−1sr−1, a slope of 3.21, and an intercept of 1.29 Mm−1sr−1. The positive intercept and steep slope suggest that the backward method produces strongly overestimated backscatter even where COBALD measures near-zero aerosol loading, consistent with the noise-driven downward error propagation of the Klett–Fernald integration when a reliable aerosol-free reference altitude is unavailable (Li et al., 2021). The forward retrieval, in contrast, achieves r = 0.97, RMSE = 0.06 Mm−1sr−1, a slope of 1.00, and an intercept of −0.02 Mm−1sr−1, suggesting strong agreement with COBALD throughout the aerosol layer and showing that the forward method remains robust even under complex, low-SNR conditions. The corresponding scatter plot is provided in the Supplement (Fig. S2).

The scatter analysis for the backscatter subsection can be found between lines 700 - 719 for backscatter.

Line 619: "To complement the profile deviation analysis, scatter statistics were computed by pairing all coincident ceilometer and GRASPpac values at matching altitudes within the 0.25–5 km range. To ensure consistency with the deviation profile analysis described above (Sect 4.2.1 and 4.2.2), the same percentile filter (10th–90th percentile of the pairwise difference distribution) was applied prior to computing statistics, so that both analyses operate on an identical subset of data points. Pearson correlation coefficients (r), root-mean-square errors (RMSE), linear regression slopes, and intercepts were then computed from the filtered pairs and used to quantify the overall agreement between each retrieval method and the GRASPpac reference across all aerosol regimes and stations. The corresponding scatter plots for backscatter and extinction at the three sites and aerosol regimes are provided in the Supplement (Figs. S3–S4).".

Line 700: "The profile-level results are confirmed by the scatter analysis. Across all three sites the forward retrieval consistently raises the correlation coefficient, reduces the RMSE, and brings the regression slope closer to unity relative to the backward approach, with the largest contrast emerging under coarse-mode-dominated conditions.

Line 719: "Lindenberg is consistent with both other sites, with the lower aerosol loading of this continental station reflected in generally smaller RMSE values across all regimes. The forward retrieval achieves close agreement with no appreciable bias (r = 0.96, RMSE = 0.07 Mm−1 sr−1, slope near unity), while the backward method (r = 0.82, RMSE = 0.16 Mm−1 sr−1) reproduces the Granada crossover pattern: overestimation at low backscatter and under-estimation at higher values, with the transition occurring near 0.2 Mm−1 sr−1. The coarse-mode subset again shows the strongest divergence between the two methods, though the small number of profiles available at this site means this result should be interpreted with caution."

For the extinction analysis the scatter analysis can be found from line 769 - 789

Line 769: "The scatter analysis for extinction shows a similar behaviour to that found for backscatter: at every station the forward retrieval improves the correlation, reduces the RMSE, and keeps the slope closer to unity than the backward approach, with the contrast sharpening under coarse-mode dominance.

Line 789: "Lindenberg reproduces the Granada pattern most closely. The forward retrieval agrees well with GRASPpac over the full dataset (r = 0.94, RMSE = 4.86 Mm−1) while the backward method (r = 0.80, RMSE = 8.23 Mm−1) again overestimates at low extinction and underestimates at high, crossing over near 7 Mm−1. The coarse-mode subset shows the steepest backward overestimation of all three sites (slope 1.33, intercept 23.14 Mm−1), against which the forward retrieval reduces the RMSE by nearly 70% (7.47 Mm−1); as noted previously, this result should be interpreted with caution given the limited number of coarse-mode profiles available at this site."

Between lines 856 and 878can be seen the new scatter analysis for the AOD comparison

Line 856: "The corresponding scatter plots are provided in the Supplement (Figs. S7–S8). The scatter analysis against AERONET AOD reinforces these findings across all three sites, and the contrast between the two methods is the largest of the three quantities examined. For the full dataset the forward retrieval achieves high correlations, near-zero intercepts, and slopes close to unity, whereas the backward retrieval shows much lower correlations and slopes well below unity, indicating that it fails to reproduce AOD variability across the dynamic range. Because AOD is a column-integrated quantity, the reference-altitude error of the backward formulation accumulates vertically, which plausibly explains why the gap is wider here than for the backscatter and extinction profiles. The scatter plots for all sites and regimes are shown in the Supplement (Figs. S7–S8).

Line 878: Under fine-mode-dominated conditions the forward retrieval correlates well at all three sites (r = 0.77–0.85, RMSE ≤ 0.02), far above the backward method (r = 0.11–0.43, slopes as low as 0.05), which reproduces fine-mode AOD variability very poorly. The forward slopes, however, remain below unity (0.56–0.83, lowest at Granada), indicating that even the forward retrieval underestimates fine-mode AOD at 1064 nm; this residual bias is most plausibly linked to the spectral conversion and Ångström exponent assumptions used to bring the retrieval to the AERONET reference wavelength. Overall, the lower dispersion and smaller deviations of the forward retrieval translate into stronger column-integrated agreement across aerosol regimes and sites, with its main remaining limitation associated with fine-mode scaling."

Line 975: “with Pearson correlation coefficients of r ≈ 0.96–0.97 against $GRASP_{pac}$ backscatter compared to r ≈ 0.72–0.82 for the backward method, and RMSE values approximately half those obtained with the backward retrieval across all three sites”

Line 985: “with RMSE values reduced by up to a factor of two with respect to the backward method (0.149 versus 0.287 Mm−1sr−1 under dust conditions at Payerne) and close agreement under complex multi-layered conditions at Lindenberg (r = 0.973, RMSE = 0.06 Mm−1sr−1)”

Line 1001: “as demonstrated by correlation coefficients of r = 0.88–0.94 and regression slopes close to 1 for the forward retrieval against AERONET AOD, compared to r = 0.50–0.54 and slopes as low as 0.36–0.47 for the backward method across the three sites.”

Selection Bias in Aerosol Classification

The exclusion of cases with intermediate Ångström exponent (0.8 < AE < 1.2) removes mixed aerosol conditions, which are common in real atmospheric scenarios. This may bias the conclusions.

Justify this choice or include mixed cases and assess their impact on the results.

We thank the referee for this comment. We agree that mixed aerosol conditions are common in real atmospheric scenarios; however, their exclusion from the fine/coarse classification does not introduce a significant bias in our conclusions for two reasons.

First, mixed aerosol cases are not excluded from the overall analysis presented in the manuscript. They are only excluded from the specific fine-mode and coarse-mode subsets used to investigate the influence of aerosol type on retrieval performance. Consequently, all mixed cases remain represented in the full-dataset statistics, which constitute the primary basis for the comparison between retrieval methods.

Second, mixed profiles represent a small fraction of the high aerosol load cases in our dataset: at Granada, only 22 out of 263 high-load profiles (8.4%) fall into the mixed category; at Payerne, 8 out of 468 (1.7%); and at Lindenberg, 21 out of 348 (6.0%). Therefore, the exclusion of these cases has only a limited impact on the statistics derived for the fine-mode and coarse-mode subsets.

Third, most mixed-condition profiles occur under low aerosol load (AOD_{440} < 0.2), where retrieval differences between the forward and backward methods are small and well-documented in the full-dataset analysis. The purpose of the fine/coarse classification is to isolate the aerosol regimes where the two retrieval methods respond most differently, and the statistical robustness of this separation is supported by the dominance of clearly classified cases within the high-load subset.

Nevertheless, we acknowledge this limitation and have added a clarifying sentence to the manuscript to make the representativeness of the classification explicit.

Line 471: "It should be noted that the fine/coarse classification is not intended to be exhaustive but rather to isolate the aerosol regimes where retrieval differences are most pronounced; mixed aerosol conditions are therefore excluded only from this specific classification analysis and not from the overall dataset evaluation. The full-dataset analysis remains the primary statistical reference and encompasses all aerosol conditions including mixed cases"

Overinterpretation of Results

The manuscript occasionally uses strong statements such as "systematically outperforms," despite limited sample sizes in some cases (e.g., coarse-mode events at Lindenberg). These claims should be moderated.

We thank the referee for this comment. We agree that statements such as "systematically outperforms" may overstate the robustness of the comparison, particularly for subsets with limited sample size. Accordingly, we have revised the manuscript to adopt a more cautious interpretation of the results and to avoid overly general conclusions where the statistical basis is limited. Specifically:

Line 8: "The results show that the forward iterative approach generally provides better agreement with the reference datasets than the Klett--Fernald backward approach."

Line 519: "The backward method can produce large negative backscatter deviations"

Line 573: "This case study indicates the forward method's greater robustness for nighttime aerosol characterization."

Line 732: "Across all stations and conditions, the forward retrieval generally reduces the dispersion of extinction differences relative to GRASPpac."

Line 829: "Similar but smaller reductions in deviation and dispersion"

Line 844: "indicating agreement comparable to or better than, that reported in previous studies using the same photometric reference"

Line 968: "The results indicate that the forward iterative approach generally provides close agreement with the reference datasets than the conventional backward method. These differences are primarily linked to the calibration strategy of the forward method,"

Line 978: "The differences between the two approaches are most evident"

Line 980: "In these cases, the forward retrieval exhibits smaller deviations and a more vertically consistent representation of aerosol layers"

We wish to clarify that, throughout the manuscript, we have consistently acknowledged the limited robustness of the coarse-mode results due to the limited availability of cases.

Minor Comments

The Introduction is somewhat lengthy and could be streamlined to better emphasize the research gap and novelty.

We thank the referee for this comment**.** We have shortened the Introduction by removing background information that was not essential to the objectives of the study and by strengthening the focus on the research gap and the novelty of the proposed analysis. Specifically, we modified the following passages:

Line 20: "Aerosol–Radiation Interactions (ARI) and Aerosol–Cloud Interactions (ACI) (Forster et al., 2021). The combined Effective Radiative Forcing (ERF) from these processes results in a net cooling of approximately −1.3 W m−2 ± 0.7 W m−2 (Forster et al., 2021)."

Line 28: "Beyond climate, aerosols significantly impact air quality by increasing fine particulate matter (PM2.5) concentrations, degrading visibility, and exacerbating respiratory and cardiovascular diseases (Pai et al., 2022)"

Some methodological choices (e.g., convergence criteria, 30-minute averaging window, cloud filtering threshold) require clearer justification.

We thank the referee for this comment. We agree that a clearer justification of these methodological choices is warranted and have expanded the corresponding explanations in the revised manuscript.

Regarding the convergence criteria for the forward iterative retrieval, the threshold of 0.01% relative change in $\alpha a(z_i)$ with a maximum of 30 iterations follows directly from Li et al. (2021), where this criterion was shown to ensure numerical stability while keeping computational cost low.

Regarding the 30-minute averaging window, this choice reflects a balance between two competing requirements: sufficient signal averaging to achieve adequate SNR in the attenuated backscatter profiles, particularly at higher altitudes, and sufficient temporal resolution to capture aerosol variability. A 30-minute window is standard practice in ceilometer aerosol studies (Wiegner et al., 2014; Brunamonti et al., 2021) and is consistent with the temporal sampling of AERONET Level 2.0 almucantar retrievals used as reference. For the COBALD comparison, the 30-minute window encompasses the full balloon ascent period through the troposphere, ensuring spatial and temporal representativeness.

Regarding the cloud-filtering threshold, profiles with a detected cloud base below 5 km were removed to avoid integration artefacts in the AOD comparison. This threshold was chosen to retain the maximum number of aerosol-only profiles while excluding cases where cloud contamination would compromise the extinction integration. A 5 km upper integration limit was selected because most of the tropospheric aerosol mass is concentrated below this altitude (Amiridis et al., 2005), and ceilometer SNR becomes insufficient for reliable retrievals above this range under typical conditions

We have added clarifying sentences to the manuscript to make these justifications explicit.

Line 242: " The convergence criterion of 0.01% relative change in $\alpha a(zi)$ with a maximum of 30 iterations, follows Li et al. (2021) and ensures numerical stability, throughout the retrieval."

Line 441: "Aerosol backscatter coefficient profiles were retrieved using the backward and forward inversion methods described in Sect. 3.1. Following the retrieval, a cloud-screening procedure was applied to remove profiles affected by low-level clouds. Specifically, profiles with a detected cloud base below 5 km were excluded to avoid artefacts in the extinction integration. This threshold retains the maximum number of aerosol-only profiles while ensuring that cloud contamination does not affect the analysis, which is restricted to the lower troposphere where most aerosol mass is concentrated (Amiridis et al., 2005)

Line 447: "The cloud-screened profiles were subsequently averaged over 30-minute intervals to reduce measurement noise while preserving sufficient temporal resolution to capture aerosol variability. This averaging period is consistent with previous CHM15k–COBALD validation studies (Brunamonti et al., 2021) and provides an appropriate balance between signal-to-noise ratio requirements at higher altitudes and temporal representativeness."

Figures 5 and 6 are informative but overly dense; readability could be improved by splitting panels or simplifying presentation.

We thank the referee for this comment. We agree that Figures 5 and 6 are dense given the number of sites, aerosol regimes, and retrieval methods displayed simultaneously. However, after careful consideration, we maintain the current layout for the following reason: the primary purpose of these figures is to enable a direct visual comparison between the forward and backward retrieval methods across all aerosol regimes and sites. Splitting the panels, for example by separating forward and backward results into distinct figures, would substantially hinder this side-by-side comparison, which is central to the conclusions of the study. We believe that the current layout, while dense, provides the most efficient representation for this comparative analysis.

Terms such as “favorable atmospheric conditions” should be explicitly defined.

We thank the referee for this comment. We agree that "favourable atmospheric conditions" was insufficiently precise. We have replaced this expression at all relevant locations in the manuscript to explicitly state the conditions required for a reliable Rayleigh calibration: cloud-free signal integration with negligible aerosol loading at the reference altitude. The specific changes are as follows:

Line 18: “made under nighttime cloud-free conditions with negligible aerosol loading at the reference altitude.”

Line 209: “This procedure requires at least three hours of signal integration during nighttime under cloud-free conditions with negligible aerosol loading at the reference altitude”.

Line 817: “These results indicate the greater stability of the forward approach, which benefits from an independent calibration performed under cloud-free conditions with negligible aerosol loading at the reference altitude.”

Additional corrections:

Additional change. During the final revision we identified a typographical inconsistency in Eq. (1) and its accompanying text (former line 181). The backscatter term was written as the attenuated backscatter coefficient β^*while the equation simultaneously retained the explicit two-way transmission term $\exp\left(-2\int_0^z \alpha\, d\, z'\right)$. By definition, the attenuated backscatter coefficient already incorporates this transmission, $\beta^*(z) = \beta(z)\exp\left(-2\int_0^z \alpha(z')\, dz'\right)$, so keeping both would apply the atmospheric transmission twice. Since the intention was to begin the calibration derivation from the lidar equation itself, we have removed the asterisk, so that Eq. (1) is now correctly expressed in terms of the total backscatter coefficient $\beta(z)$together with the explicit two-way transmission term.

Line 183: “The lidar equation can be expressed in terms of the backscatter coefficient, β (Eq. 1). In this formulation, P (z) denotes the received signal as a function of atmospheric optical properties and measurement range z, the exponential is the transmission term, while CL represents the system-dependent calibration constant accounting for instrumental characteristics.”

New Eq 1: $$\frac{P_{(z)}\, z^2}{C_L} = \beta\,(z)\,\exp\left\{-2\int_0^z \alpha(z')\, d_{z'}\right\}$$

Additional change. We found that first time term $GRASP_{pac}$ appears the pac part is not correctly defined.

Line 70: $GRASP_{pac}$ (Photometer and Ceilometer)

Additional change. To address the referees' comments, the first paragraphs of Sect. 3.5 (lines 438–456) was revised.

Line 438: “To evaluate the performance of the backward and forward ceilometer retrievals, aerosol optical depth (AOD) was derived from the retrieved aerosol profiles and compared with independent AERONET sun–sky photometer observations. This comparison provides an assessment of the ability of both retrieval approaches to reproduce the total aerosol optical load within the atmospheric column.”

Line 456:” AOD was then calculated by vertically integrating the extinction profiles from the surface to 5 km a.g.l. using the trapezoidal rule. Finally, the resulting ceilometer-derived AOD values were compared with AERONET Level 2.0 observations. “

Additional change. On our own revision, we found that one profile had been described as noise-limited above ~2 km, when in fact it does not converge towards the molecular signal at high altitude — indicating a calibration error rather than low signal-to-noise. The text (line 545) has been corrected to present this as a failure case of the method rather than a profile reliable below 2 km. The revised sentence now reads:

Line 556: "In this case the retrieved profile does not converge to the molecular signal, pointing to a calibration error rather than low signal-to-noise; the profile is thus erroneous and is presented as an example of the method failing when a reference altitude cannot be found."

Additional change. In the course of our revision we noted a number of statements that were not formulated with sufficient precision, as well as points requiring clarification, which have been revised as follows:

Line 634: **“**The forward method consistently shows smaller deviations and reduced variability than the backward approach, reducing the near-surface mean deviation by approximately a factor of three under coarse-mode-dominated conditions. This behaviour is consistent with the greater robustness of the forward retrieval in situations where the aerosol-free reference altitude required by the backward inversion is difficult to identify or may be affected by residual aerosol contamination.”

Line 649: “The substantially larger deviations observed for the backward retrieval are consistent with the increased difficulty of identifying an aerosol-free reference region under high aerosol loading, which enhances the propagation of errors throughout the profile. The forward approach is less affected by this limitation, maintaining similar deviations at both altitude ranges and closer agreement with the GRASPpac reference under dust-rich and low-signal conditions (Fig. 5b).”

Line 731: “Comparison of extinction profiles retrieved by the backward and forward methods against $GRASP_{pac}$”

Line 795: “Overall, the comparison with $GRASP_{pac}$ indicates that the forward iterative retrieval generally provides a more stable and vertically consistent representation of aerosol extinction than the traditional backward approach. The differences between the two methods are most pronounced under high aerosol load and coarse-mode-dominated conditions, where the forward retrieval exhibits smaller deviations and lower variability. These results support the use of the forward approach for operational aerosol profiling within ceilometer network”

Line 801: “The column-level consistency of the ceilometer retrievals is assessed in two steps. First, AOD derived from the backward and forward retrievals is compared with AERONET observations using the standard fixed lidar ratio of 50 sr across the different aerosol regimes (Sect. 4.3.1). Second, the sensitivity of these AOD estimates to the assumed lidar ratio is evaluated by testing alternative fixed lidar ratios and a composition-based variable estimate (Sect. 4.3.2)”

Line 972: “whereas the forward method relies on an independent calibration derived from nighttime observations.”

Line 974: “shows lower dispersion in the retrieved backscatter and extinction profiles”

Line 984: “lower variability and smaller dispersion”

Line 992: “many of the main features captured by more complex multi-instrument solutions.”

Line 1008: “In summary, this study indicates that the forward iterative retrieval generally provides a more vertically consistent representation of aerosol backscatter and extinction than the traditional Klett–Fernald backward inversion, with smaller deviations and lower dispersion relative to the reference datasets considered here. While extinction retrievals remain sensitive to the assumed lidar ratio, backscatter profiles are less affected by this assumption and therefore exhibit lower uncertainty. Together, these results support the use of operational ceilometers, when carefully calibrated and quality-controlled, as practical and sustainable sources of vertically resolved aerosol information, contributing meaningfully to long-term air quality monitoring and climate-related atmospheric research.”

Response to Referee #2 Comments

Validation of ceilometer aerosol profile retrievals using sun–sky photometer and balloon-borne in situ measurements

General comments:

Paper deals with analysis of how well the widespread ceilometers could be used fro aerosol profile characterisation. Taken into account that we have quite some ceilometers as compared to lidars this study can potentially significantly increase amount of available aerosol observations. The paper is generally well structured, and has its messages clear, and within the scope of ACP. However, it avoids many significant "elephants in the room", doesn't make it clear why this specific methodology of comparison were selected, and full of inconsistencies in little details, the ones where the "devil is usually is". I must admit it took me much longer than I anticipated to find at least some of them, that in my opinion should be addressed before the paper could be published. Below I give several insites how the paper should be improved. My overall recommendation is acceptance after major revision

Response

We sincerely thank the referee for the time and the suggestions which contributed to improving the quality of this work. All the points raised have been carefully considered and the revisions have been introduced in the manuscript. All answers in detailed are written in blue.

Major Comments:

Statements of superiority of one method over another are not well suported, without a proper error analysis, i.e. quantifiable details of possible retrieval/estimation errors of different methods. It seems that whole study either overlooks significantly many nuances of the retrievals which make direct comparisons questionable, either specifically designed to skew for one of the methods.

We thank the referee for this comment, which raises two distinct concerns: first, that comparative statements were not supported by a quantifiable error analysis, and second, that the comparison framework might favour one method by construction. We address both, and we note that the revisions introduced in response to the more specific comments below were designed precisely to resolve them.

Regarding the error analysis, the revised manuscript now characterises the retrieval differences with a stronger statistical analysis. For every site and aerosol regime we report mean deviation, standard deviation, and RMSE against GRASPpac on a percentile-filtered dataset (Sect. 4.2; Tables 2 and 3), and we have added scatter analyses reporting Pearson correlation, RMSE, regression slope and intercept against the independent

AERONET AOD measurements, as well as against both extinction and backscatter from GRASPpac (Lines 700, 789, 856) and COBALD (Lines 525 and 558). In addition, the principal sources of retrieval/estimation error are now quantified per method: the lidar-ratio assumption through a dedicated sensitivity analysis comparing S_a = 30, 50, 70 sr and a composition-aware variable estimate (Sect. 4.3.2, Figs. 8–9; Lines 284, 887); and the Ångström-exponent treatment in the spectral conversion, estimated to contribute differences of -2.6–5.6% (Line 361), together with the calibration constant uncertainty, which is now explicitly discussed. Comparative statements in the revised text are now tied to these metrics, and we have reworded passages that asserted superiority in general terms so that each comparison is expressed as the degree of agreement with the reference datasets and the independent validation measurements, with the supporting statistics cited.

Regarding the concern that the comparison could be skewed towards one method, we emphasise that the two ceilometer retrievals are evaluated under identical conditions: they use the same E-Profile Level 2 attenuated backscatter, the same lower boundary (z0 = 250 m a.g.l), and the same fixed lidar ratio of 50 sr (Lines 166, 258, 402, 444). No method-specific tuning is applied to either retrieval. Therefore, the experimental setup minimizes method-dependent biases and allows differences associated with the inversion approaches themselves to be assessed under equivalent conditions. Additionally, the relative behaviour of the methods is compared against GRASPpac, which is used as a comparative reference dataset (line 418) rather than as absolute truth but is corroborated against two independent references: column-integrated AERONET AOD and balloon-borne COBALD in-situ measurements.

We believe these revisions provide the quantifiable error analysis the referee requested and demonstrate that the comparison is conducted on an equal footing across methods, with the main conclusions supported by independent validation datasets.

Whole methodology of comparison is not very well justified, averaging thousands of profiles by splitting them into specified groups won't give any particular incentives on what is actually happening. My strongest belief is that scatter plots and correlation coefficients as well as root mean square errors and linear should be provided to support the conclusions of the paper (at least for COBALD cases and AERONET, since only these could be taken as a real validation).

We thank the referee for this comment. We agree that the original version of the manuscript relied too heavily on aggregated statistics and that additional statistical diagnostics are necessary to better characterise the agreement between retrieval methods and independent reference datasets. To address this concern, we have expanded the statistical analysis throughout the Results section by incorporating scatter plots, Pearson correlation coefficients, RMSE values, regression slopes, and intercepts.

These metrics provide a more rigorous assessment of retrieval performance and allow the conclusions to be supported by quantitative evidence rather than by averaged profile statistics alone.

These new diagnostics have been incorporated in all validation exercises, including the comparisons against COBALD, GRASPpac, and AERONET. The most relevant additions are summarised below.

For the COBALD case studies, scatter plots and regression statistics have been added to complement the profile comparisons and to quantify the agreement between the retrieved and in situ backscatter coefficients.

Line 525: "The corresponding scatter plot provided in the Supplement (Fig. S1) further supports these findings. Although both ceilometer retrievals correlate strongly with the COBALD observations (r = 0.95 and 0.94 for the backward and forward methods, respectively), the similar correlation coefficients conceal important differences in retrieval performance. The backward retrieval yields RMSE = 0.29 $Mm^{-1}sr^{-1}$, a slope of 0.66, and an intercept of 0.01 $Mm^{-1}sr^{-1}$; the slope below unity suggests an underestimation of the aerosol backscatter signal that becomes increasingly pronounced at higher backscatter values. This behaviour is consistent with the increasing difficulty of identifying a suitable aerosol-free reference altitude under optically thick dust conditions, which can affect the stability of the backward integration. In contrast, the forward retrieval achieves RMSE = 0.15 $Mm^{-1}sr^{-1}$, a slope of 0.92, and an intercept of 0.01 $Mm^{-1}sr^{-1}$, reducing the RMSE by nearly 50 % and showing closer agreement with the COBALD reference across the full range of observed backscatter values."

Line 558: "This contrast is further quantified in the scatter analysis (Fig. S2 in the Supplement). The backward retrieval yields r = 0.51, RMSE = 2.11 $Mm^{-1}sr^{-1}$, a slope of 3.21, and an intercept of 1.29 $Mm^{-1}sr^{-1}$. The positive intercept and steep slope suggest that the backward method produces strongly overestimated backscatter even where COBALD measures near-zero aerosol loading, consistent with the noise-driven downward error propagation of the Klett–Fernald integration when a reliable aerosol-free reference altitude is unavailable (Li et al., 2021). The forward retrieval, in contrast, achieves r = 0.97, RMSE = 0.06 $Mm^{-1}sr^{-1}$, a slope of 1.00, and an intercept of −0.02 $Mm^{-1}sr^{-1}$, suggesting strong agreement with COBALD throughout the aerosol layer and showing that the forward method remains robust even under complex, low-SNR conditions. The corresponding scatter plot is provided in the Supplement (Fig. S2)."

Line 619: "To complement the profile deviation analysis, scatter statistics were computed by pairing all coincident ceilometer and GRASPpac values at matching altitudes within the 0.25–5 km range. To ensure consistency with the deviation profile analysis described above (Sect 4.2.1 and 4.2.2), the same percentile filter (10th–90th percentile of the pairwise difference distribution) was applied prior to computing statistics, so that both analyses operate on an identical subset of data points. Pearson correlation coefficients (r), root-mean-square errors (RMSE), linear regression slopes, and intercepts were then computed from the filtered pairs and used to quantify the overall agreement between each retrieval method and the GRASPpac reference across all aerosol regimes and stations. The corresponding scatter plots for backscatter and extinction at the three sites and aerosol regimes are provided in the Supplement (Figs. S3–S4).".

For the comparison with GRASPpac backscatter profiles, the discussion has been expanded to include correlation coefficients, RMSE values, and regression parameters for each aerosol regime and station.

The scatter analysis for the backscatter subsection can be found between lines 700 – 719 for backscatter.

Line 700: "The profile-level results are confirmed by the scatter analysis. Across all three sites the forward retrieval consistently raises the correlation coefficient, reduces the RMSE, and brings the regression slope closer to unity relative to the backward approach, with the largest contrast emerging under coarse-mode-dominated conditions.

Line 719: "Lindenberg is consistent with both other sites, with the lower aerosol loading of this continental station reflected in generally smaller RMSE values across all regimes. The forward retrieval achieves close agreement with no appreciable bias (r = 0.96, RMSE = 0.07 Mm−1 sr−1, slope near unity), while the backward method (r = 0.82, RMSE = 0.16 Mm−1 sr−1) reproduces the Granada crossover pattern: overestimation at low backscatter and underestimation at higher values, with the transition occurring near 0.2 Mm−1 sr−1. The coarse-mode subset again shows the strongest divergence between the two methods, though the small number of profiles available at this site means this result should be interpreted with caution.

An analogous statistical analysis has also been incorporated for the extinction-profile comparison (Lines 769 – 789).

Line 769: "The scatter analysis for extinction shows a similar behaviour to that found for backscatter: at every station the forward retrieval improves the correlation, reduces the RMSE, and keeps the slope closer to unity than the backward approach, with the contrast sharpening under coarse-mode dominance.

Line 789: "Lindenberg reproduces the Granada pattern most closely. The forward retrieval agrees well with $GRASP_{pac}$ over the full dataset (r = 0.94, RMSE = 4.86 Mm−1) while the backward method (r = 0.80, RMSE = 8.23 Mm−1) again overestimates at low extinction and underestimates at high, crossing over near 7 Mm−1. The coarse-mode subset shows the steepest backward overestimation of all three sites (slope 1.33, intercept 23.14 Mm−1), against which the forward retrieval reduces the RMSE by nearly 70% (7.47 Mm−1); as noted previously, this result should be interpreted with caution given the limited number of coarse-mode profiles available at this site."

Most importantly, for the independent validation against AERONET AOD, the manuscript now includes a detailed scatter analysis that directly addresses the referee's request for quantitative validation metrics (Lines 856 – 878).

Line 856: "The corresponding scatter plots are provided in the Supplement (Figs. S7–S8). The scatter analysis against AERONET AOD reinforces these findings across all three sites, and the contrast between the two methods is the largest of the three quantities examined. For the full dataset the forward retrieval achieves high correlations, near-zero intercepts, and slopes close to unity, whereas the backward retrieval shows much lower correlations and slopes well below unity, indicating that it fails to reproduce AOD variability across the dynamic range. Because AOD is a column-integrated quantity, the reference-altitude error of the backward formulation accumulates vertically, which plausibly explains why the gap is wider here than for the backscatter and extinction profiles. The scatter plots for all sites and regimes are shown in the Supplement (Figs. S7–S8).

Line 878: "Under fine-mode-dominated conditions the forward retrieval correlates well at all three sites ($r = 0.77$–0.85, RMSE ≤ 0.02), far above the backward method ($r = 0.11$–0.43, slopes as low as 0.05), which reproduces fine-mode AOD variability very poorly. The forward slopes, however, remain below unity (0.56–0.83, lowest at Granada), indicating that even the forward retrieval underestimates fine-mode AOD at 1064 nm; this residual bias is most plausibly linked to the spectral conversion and Ångström exponent assumptions used to bring the retrieval to the AERONET reference wavelength. Overall, the lower dispersion and smaller deviations of the forward retrieval translate into stronger column-integrated agreement across aerosol regimes and sites, with its main remaining limitation associated with fine-mode scaling."

These statistical metrics are now explicitly reflected in the Discussion and Conclusions sections, where the interpretation of the results is supported by the corresponding quantitative indicators.

Line 975: "with Pearson correlation coefficients of $r \approx 0.96$–0.97 against $GRASP_{pac}$ backscatter compared to $r \approx 0.72$–0.82 for the backward method, and RMSE values approximately half those obtained with the backward retrieval across all three sites"

Line 985: "with RMSE values reduced by up to a factor of two with respect to the backward method (0.149 versus 0.287 $Mm^{-1}sr^{-1}$ under dust conditions at Payerne) and close agreement under complex multi-layered conditions at Lindenberg ($r = 0.973$, RMSE $= 0.06$ $Mm^{-1}sr^{-1}$)"

Line 1001: "as demonstrated by correlation coefficients of $r = 0.88$–0.94 and regression slopes close to 1 for the forward retrieval against AERONET AOD, compared to $r = 0.50$–0.54 and slopes as low as 0.36–0.47 for the backward method across the three sites."

Overall, the revised manuscript now provides the scatter plots, correlation coefficients, RMSE values, regression slopes, and intercepts requested by the referee. These additions strengthen the validation framework and allow the relative performance of the retrieval methods to be assessed using quantitative and independently verifiable metrics.

Methods compared have multiple parameters that affect their performance, section of altitudes, correction of overlap, and most importantly lidar ratio selection. This should be discussed and quantified, it is not clear if 50 sr goes well for all the cases, GARSPpac has this parameter estimated, a comparison would be nice way to show that selection was correct, or grouping results into cases when retrieved and assumed SAs are close or very different to see the impact on the performance.

We thank the referee for this valuable suggestion. We agree that the performance of both retrieval methods is influenced by several methodological choices, including the selected altitude range, overlap treatment, calibration procedure, and the assumed aerosol lidar ratio. Among these factors, the lidar ratio is generally recognised as one of the dominant sources of uncertainty in elastic-backscatter aerosol retrievals because it directly controls the conversion from backscatter to extinction and therefore affects the retrieved AOD. For this reason, we focused our additional analysis on quantifying the impact of the lidar-ratio assumption.

To address this issue, we have added a dedicated sensitivity analysis (new Sects. 3.2 and 4.3.2 and Figs. 8 and 9), in which daily site-specific lidar-ratio estimates derived from IFS-CAMS aerosol composition forecasts are compared with three fixed lidar-ratio values (30, 50, and 70 sr). The backward and forward AOD retrievals are re-evaluated against AERONET observations using each lidar-ratio configuration The analysis is performed across the same aerosol regime classification used throughout the study and at all three sites, allowing the propagation of the lidar ratio assumption into the retrieved AOD to be quantified in terms of mean deviation, standard deviation, and RMSE.

Although the referee also suggests comparing the assumed lidar ratio with the lidar ratio retrieved by GRASPpac, we note that GRASPpac retrieves an effective lidar ratio within its inversion framework under assumptions that differ from those adopted by the ceilometer-only retrievals. Therefore, rather than using GRASPpac as an absolute reference for lidar-ratio validation, we opted to evaluate the practical impact of different lidar-ratio assumptions directly through comparison with the independent AERONET AOD observations.

The details of this methodology are described in Lines 284–321 of the revised manuscript.

Line 284: “The aerosol lidar ratio is a critical parameter in elastic-backscatter aerosol retrievals, as it links aerosol backscatter and extinction coefficients. Since single-wavelength ceilometers cannot independently retrieve Sa, its value must be prescribed from external information. Although a fixed lidar ratio of 50 sr was adopted as the baseline configuration in this study, a variable lidar ratio derived from CAMS aerosol composition forecasts was also estimated to evaluate the sensitivity of the retrievals to this assumption and to assess its potential impact under different aerosol regimes."

Line 327 "This approach yields a daily, site-specific lidar ratio that reflects the prevailing aerosol composition as represented in the CAMS forecast across all three stations (Granada, Payerne, and Lindenberg). This lidar ratio is subsequently used in a sensitivity analysis together with fixed lidar-ratio values applied throughout the study period, particularly given the contrasting aerosol regimes across the three sites, ranging from mineral dust and anthropogenic pollution at Granada to predominantly fine-mode continental aerosol at Payerne and Lindenberg. Nevertheless, several limitations should be noted. The CAMS AOD at 550 nm is used as a proxy for the optical weighting, whereas the ceilometer operates at 1064 nm; spectral differences in species-specific AOD contributions may therefore introduce a bias in the weighted lidar ratio. Additionally, a single near-surface RH value at 1000 hPa is used to represent the column-mean hygroscopic state, whereas the vertical RH profile can vary substantially, particularly in the presence of elevated aerosol layers such as Saharan dust plumes."

The results of this sensitivity analysis are discussed in Sect. 4.3.2 (The details of this methodology are given in Lines 887 - 923 of the revised manuscript) and demonstrate that the optimal lidar ratio depends on the aerosol regime. However, they also show that a fixed value of 50 sr provides a reasonable compromise across the three study sites and aerosol conditions.

The results of the sensitivity analysis can be found from lines 887 to 946.

Line 887: " The column-integrated evaluation in Sect. 4.3.1 was carried out using a single fixed lidar ratio of 50 sr to convert backscatter into extinction. Although this value produced good overall agreement with AERONET, particularly for the forward retrieval, the residual deviations identified under high aerosol load, and most markedly under coarse-mode-dominated conditions, were attributed to the limitation of applying a single fixed lidar ratio across the contrasting aerosol regimes sampled at the three sites.

Line 946: "In summary, the fixed lidar ratio of 50 sr provides robust results across across all sites and aerosol regimes, with the lowest overall RMSE for the full dataset. Under coarse-mode dominated conditions, the fixed lidar ratio of 50 sr tends to underestimate AOD while under fine-mode dominated conditions it shows a small positive bias, reflecting the inherent limitation of a single fixed value across diverse aerosol regimes. The IFS-CAMS derived lidar ratio reduces these regime-dependent deviations, particularly under coarse-mode dominated conditions at Granada and Payerne, where it achieves near-zero mean deviations (-0.006 ± 0.053 and 0.018 ± 0.045 respectively) and RMSE reductions of up to 38% compared to the fixed lidar ratio of 50 sr. At Lindenberg under coarse-mode conditions, no clear advantage of IFS-CAMS-derived lidar ratio over the fixed lidar ratio of 50 sr is observed, though this result should be interpreted with caution given the limited number of cases (N = 63). For the backward method, the lidar ratio choice has a secondary impact compared to the reference altitude selection, which remains the dominant source of uncertainty under high aerosol load conditions regardless of the assumed lidar ratio value."

It is not very clear if methods or their realisations are authors developments or some other product, and which inputs were used, these have to be clearly stated! All the codes and products should be stated, referenced and versioned. Ideally with repository links.

We thank the referee for this comment, which has helped us improve the transparency and reproducibility of the study. We agree that the original manuscript did not clearly distinguish between the retrieval algorithms, processing tools, and observational products used in this work, nor did it consistently indicate whether these were author-developed implementations or externally maintained software packages.

To address this issue, we have revised the manuscript to explicitly identify the software used for each retrieval, provide the corresponding references, clarify the origin of the processing chains, and include repository or access links whenever available. These clarifications have been incorporated into Sects. 2 and 3, as well as the Data Availability statement. The main additions are summarised below.

Regarding the ceilometer retrievals, neither the Klett–Fernald backward inversion nor the forward iterative retrieval were implemented through custom software developed by the authors. Instead, both retrievals were performed using the openly available A-Profiles package, which implements the corresponding algorithms described in the literature.

Line 166: "Both the Klett–Fernald backward inversion and the forward iterative method were implemented using the the A-Profiles package (https://github.com/AugustinMortier/a-profiles) in this work, following the implementations of Wiegner and Geiß (2012) and Wiegner et al. (2014) for the backward retrieval and Li et al. (2021) for the forward retrieval. Ceilometer calibration was performed operationally within the E-Profile framework, which provided the calibration constant CL used throughout this study."

Regarding GRASPpac, we clarified that it is not an independent software package or repository but rather a specific inversion strategy implemented within the broader GRASP framework.

GRASPpac is not an independent code or a separate repository from GRASP. Rather, it is an inversion strategy implemented within the general GRASP framework. Access to GRASPpac is therefore obtained in the same way as access to GRASP itself, through the official GRASP distribution channels, and only requires configuring the GRASP settings file to execute the corresponding inversion strategy. The GRASPpac methodology and configuration are described in detail in Román et al. (2018), where the retrieval setup and the specific inversion approach are fully documented. In the present work, the GRASPpac configuration values were obtained directly through the CAECENET tool, described comprehensively in Herrero del Barrio et al. (2024), which provides detailed information about the methodology, implementation, and generated products; interested users may contact the authors for further information regarding CAECENET products and their use.

Line 418: "GRASPpac is not a standalone code but an inversion strategy implemented within the general GRASP framework (Dubovik et al., 2021); it is accessed through the official GRASP distribution and configured via the GRASP settings file to execute the corresponding strategy, as documented in Román et al. (2018)."

Finally, we expanded the Data Availability section to provide direct access information for the software and datasets used in this study, thereby facilitating reproducibility of the analysis.

Line 1015: "The A-Profiles package used for the backward and forward ceilometer retrievals is openly available at https://github.com/AugustinMortier/a-profiles. GRASP is distributed through its official channels (https://www.grasp-open.com) the GRASPpac strategy is described in Román et al. (2018) and was run via the CAECENET system (Herrero del Barrio et al., 2024). E-Profile Level 2 products are available through CEDA (https://catalogue.ceda.ac.uk/)."

Particular details of retrievals i.e., inputs, wavelengths, reference and ground altitudes selections, calibrations, sensitivity of methods to these decisions should be clearly stated, quantified and discussed.

We thank the referee for this summary comment. We agree that the retrieval details and their associated sensitivities must be explicitly stated, quantified, and discussed. In the revised manuscript, we have expanded the methodological descriptions and clarified the principal assumptions underlying each retrieval approach. Rather than repeat material here, we note that each of the listed items has been clarified in the revised manuscript in response to the corresponding major and minor comments, and we summarise below where each is now addressed:

- **Inputs and products.** The inputs to each method, and whether each method/product is an authors' development or an external product, are now stated with references, versions, and links in Sects. 2 and 3 and in the Data Availability statement (Lines 166, 414, 1004). The backward and forward retrievals use the A-Profiles package on E-Profile Level 2 attenuated backscatter; GRASPpac uses spectral AOD and sky radiance together with the cloud-screened ceilometer range-corrected signal.

- **Wavelengths.** The photometer channels used in the analysis (440, 675, 870, and 1020 nm) are now explicitly specified (Lines 135 and 414). The operating wavelength of the CHM15k ceilometer (1064 nm), as well as the 1064-to-940 nm spectral conversion applied for the COBALD comparison, are also described in detail (Line 414).
- **Reference and ground altitudes.** The lower boundary (z0 = 250 m a.g.l.) and the reference-altitude search window (1500–5000 m a.g.l.), the method used to select zref automatically via the a-profiles package, and the vertical limitations and representativeness issues of all three methods are now described in Lines 259 and 402.
- **Calibration procedure.** The ceilometer calibration (calibration constant, CL, obtained operationally by E-Profile through Rayleigh calibration under nighttime, cloud-free, and low-aerosol periods for the CHM15k instruments) is now more detailed in line 166.
- **Sensitivity analysis.** The sensitivity of the retrievals to the principal assumptions is now quantified and discussed: the lidar ratio assumption via the dedicated sensitivity analysis comparing S_a = 30, 50, and 70 sr, together with the IFS–CAMS variable estimate (new Sects. 3.2 and Sects 3.3 and 4.3.2, Figs. 8 and 9; Lines 284 and 887), the unity-transmittance assumption at z0 and the temperature-dependent overlap correction (Line 222 and 224), and the spectral-conversion sensitivity to the Ångström-exponent treatment (Line 361).

Taken together, these revisions provide a more complete description of the retrieval inputs, methodological choices, calibration procedures, and sensitivity analyses. We trust that, taken together, these revisions provide the clear statement, quantification, and discussion of retrieval details and sensitivities that the referee requested.

Authors contribution is vague and lacks detail. Only 8 authors out of 18 (!sic) mentioned in dedicated paragraph. Provide detailed description of contributions or shorten the co-authors list.

We thank the referee for this comment. We would like to clarify that all listed authors fulfil the authorship criteria and made substantive contributions to the work. Therefore, rather than reducing the author list, we have revised and expanded the Author Contributions section to explicitly identify the role of each contributor and ensure that no author appears without a defined contribution. The revised statement reads:

Line 1018: "JMR, AC, and FNG conceptualized the study and developed the methodology. JMR, AC, FNG, ES, AH, CHB, RR, PLMM performed the formal analysis and data curation. FNG, AC, and AH provided supervision. JMR wrote the original manuscript draft, which was reviewed and edited by all authors (ADZ, VMNH, ORN, MJGM, JABA, MH, DPR, LD, LAA). FNG administered the project and acquired funding."

Minor Comments:

Line 125: solar irradiance and sky radiance at multiple wavelengths covering the 340-1640 nm range. It is not clear which channels where used, GRASPpac should follow AERONET methodology, hence using different range, please correct or provide more details.

We thank the referee for this comment. The previous paragraph was indeed unclear regarding the specific wavelengths used. We have revised the text to explicitly indicate the channels employed in the GRASPpac retrieval. We added the following sentence to correct it.

Line 135: "Although the instrument is capable of operating over its full nominal spectral range, only the channels centred at 440, 675, 870, and 1020 nm were utilized in this study, as these wavelengths are available for the majority of AERONET sun–sky photometers and are consistent with the standard configuration of the GRASP inversion procedure (Román et al., 2018).

Line 160: *"obtained from numerical models."* Can authors precise, which? Are different models used, if yes, why, and under which conditions they are selected? Estimates of beta may vary a lot depending which model was used, please provide details.

We thank the referee for this comment. The original sentence was indeed misleading. No numerical weather prediction model is used in this step. Instead, the molecular backscatter coefficient in the E-Profile chain processing is calculated using Rayleigh scattering theory with pressure and temperature profiles from the US Standard Atmosphere (United States Committee on Extension to the Standard Atmosphere, 1976). We have clarified this point in the revised manuscript.

Line 180: "with standard atmospheric pressure and temperature profiles (United States Committee on Extension to the Standard Atmosphere, 1976) interpolated to the ceilometer altitude grid."

Line 198: *"high SNR"*, please provide exact number or range of values.

We thank the referee for this comment. The original phrasing lacked specificity. Following clarification with the E-Profile team, no fixed SNR threshold is applied during the operational Rayleigh calibration procedure. Rather, the method inherently requires sufficiently high SNR to yield reliable results this is why it is applicable to the CHM15k but not to lower-power instruments such as the CL31 or CL51. Calibration quality is ensured through the combination of cloud screening, extended signal averaging, and internal quality-control procedures implemented within the E-Profile processing chain.

Accordingly, the manuscript has been revised as follows:

Line 218: "requiring several hours of signal averaging to achieve sufficient SNR and is therefore being applicable to higher-power instruments such as the CHM15k (Li et al., 2021; Wiegner et al., 2014) (further details in Sect.3.1.1)."

Line 229: *"A reference altitude z_ref is selected in the upper troposphere, where βa(zref) = 0 is assumed"*. How this altitude is selected, is there a range, specific value or a method to define?

We thank the referee for this comment. We have added the following paragraph to clarify how z_{ref} is selected.

Line 259: "The reference altitude (z_{ref}) was determined automatically within a fixed search window between 1500 and 5000 m a.g.l using the a-profiles package (https://github.com/AugustinMortier/a-profiles). Within this window, the algorithm identifies the vicinity of the minimum of a 3-bin running-average smoothed attenuated backscatter profile, selecting specifically the point where the raw signal most closely matches the smoothed signal. At this altitude, the aerosol backscatter is assumed negligible ($\beta_a(z_{ref}) = 0$) (Wiegner and Geiß, 2012; Wiegner et al., 2014)."

Line 271: *"The use of altitude-resolved Ångström exponents allows the spectral conversion to account for vertical variability in Aerosol microphysical properties."* And assumption of constant S_a doesn't. Discuss the possible consequences of such methodological differences in the methods.

We thank the referee for raising this important point. The referee correctly identifies an asymmetry in the methodology: the spectral conversion from 1064 to 940 nm employs an altitude-resolved Ångström exponent and therefore accounts for vertical variability in aerosol microphysics, whereas the lidar inversion assumes a vertically constant lidar ratio of 50 sr. We agree that the implications of this difference were not sufficiently discussed in the original manuscript and have now clarified this point in the revised version.

The two choices have very different consequences. As quantified in the sensitivity test added in response to the following comment (Sect. 3.3), the spectral conversion is only weakly sensitive to the treatment of the Ångström exponent over the narrow near-infrared interval involved (940–1064 nm), with column-mean differences of approximately 2.6–5.6 % between the altitude-resolved and profile-mean approaches.

In contrast, the assumption of a vertically constant lidar ratio is the more consequential simplification for representing vertical microphysical variability, since the extinction retrieval scales linearly with the assumed lidar ratio. Consequently, errors in the lidar-ratio assumption propagate directly into both the extinction profiles and the integrated AOD. In layers dominated by fine-mode aerosol the fixed value of 50 sr tends to underestimate extinction, whereas in coarse-mode-dominated layers such as Saharan dust it tends to overestimate it (Sect. 3.3, Line 361). The magnitude of this effect, and the extent to which a composition-aware variable lidar ratio (S_a IFS) can mitigate it, are quantified in the lidar ratio sensitivity analysis presented in Sect. 4.3.2 (line 887).

We have therefore clarified in the revised text that, although the altitude-resolved spectral conversion is methodologically more rigorous in accounting for vertical variability, this rigour cannot be fully transferred to the inversion: single-wavelength elastic ceilometers cannot independently retrieve a vertically resolved lidar ratio without additional measurements, so the lidar ratio remains the dominant constraint on representing vertical microphysical variability. This limitation is now explicitly stated and linked to the sensitivity analysis in Sect. 4.3.2.

Line 272: “additional uncertainty”, can this uncertainty be estimated somehow?

We thank the referee for this suggestion. To quantify this source of uncertainty, we performed an additional sensitivity analysis for the two COBALD case studies by comparing profiles converted using the altitude-resolved Ångström exponent employed in this work with profiles converted using a single profile-mean Ångström exponent.

The resulting column-mean differences are small: approximately 5.6% for the Payerne dust case (25 June 2019) and 2.6% for the Lindenberg multi-layered case (27 August 2020). For the Lindenberg case, the stronger vertical variability of the Ångström exponent leads to a larger dispersion of the profile differences, with a relative standard deviation of approximately 6 %.

These results indicate that the spectral conversion is only weakly sensitive to the Ångström exponent treatment over the narrow 940–1064 nm interval, and the estimate has been added to the revised manuscript.

Line 361: “To estimate the magnitude of the uncertainty associated with the Ångström exponent, we compare profiles converted with the altitude-resolved Ångström exponent against those converted with a single profile-mean value. For the two COBALD case studies, this comparison yields a column-mean difference of approximately 5.6 ± 3.5% for the Payerne dust case (25 June 2019) and of about -2.6 ± 6.2 % for the Lindenberg multi-layered case (27 August 2020); in the latter, the stronger vertical variability of the Ångström exponent produces a larger dispersion of the differences around this small column-mean value. These values indicate that the uncertainty associated with the Ångström exponent conversion should be accounted for when interpreting the profile comparisons."

Line 291: *"In the GRASPpac (Photometer And Ceilometer) configuration"* Authors continuously refer to GRASPpac, at the same time it is completely not clear how this code/configuration can be accessed, it is not a part of GRASP open repository, which is by the way not provided in the text. It is also not clear how ceilometer differs from lidars (GRASP-wise), publishing a repo link and providing brief explanations should clarify all that.

We thank the referee for this comment, which has helped us clarify the nature and accessibility of GRASPpac and its relationship to the GRASP framework.

Regarding accessibility, GRASPpac is not an independent code or a separate repository from GRASP. Rather, it is an inversion strategy implemented within the general GRASP framework. Access to GRASPpac is therefore obtained in the same way as access to GRASP itself, through the official GRASP distribution channels, and only requires configuring the GRASP settings file to execute the corresponding inversion strategy. The GRASPpac methodology and configuration are described in detail in Román et al. (2018), where the retrieval setup and the specific inversion approach are fully documented.

To improve reproducibility, we have now explicitly added the GRASP access information and repository link to both the manuscript and the Data Availability section.

In the present work, the GRASPpac configuration values were obtained directly through the CAECENET tool, described comprehensively in Herrero del Barrio et al. (2024), which provides detailed information about the methodology, implementation, and generated products; interested users may contact the authors for further information regarding CAECENET products and their use.

Line 418: "GRASPpac is not a standalone code but an inversion strategy implemented within the general GRASP framework (Dubovik et al., 2021); it is accessed through the official GRASP distribution and configured via the GRASP settings file to execute the corresponding strategy, as documented in Román et al. (2018). This synergistic approach is particularly valuable for operational networks lacking advanced multi-wavelength research lidars. GRASPpac retrievals used in this work were generated automatically in near real time by the CAECENET system (Herrero del Barrio et al., 2024)."

Line 300: "*is discretised into 60 logarithmic altitude layers between 250 and 7000 m."* Does forward and backward used 250 and z_0 and 700m as z_ref? Theoretically if methods are compared these have to be the same, since I presume their selection may significantly impact all the retrievals. Please discuss, for e.g. Hervo et al. (2016, cited in the manuscript) demonstrated that the CHM15k overlap function is highly temperature-dependent and often incomplete below 600–1000 m. It may introduce significant error, assuming transmittance of one in high aerosol load cases.

We thank the referee for this comment. We agree that clarifying the vertical boundaries of each retrieval method is important, particularly given their potential impact on the comparison. We also agree that the effects of overlap correction and the unity-transmittance assumption in the forward method deserve further discussion.

Regarding the overlap correction, the empirical temperature-dependent correction of Hervo et al. (2016) is applied within the E-Profile processing chain, which substantially reduces overlap-related artefacts above 250 m a.g.l. and therefore justifies this as the effective lower boundary for all three methods.

The assumption of unity transmittance at z0 in the forward method introduces an uncertainty that is expected to be negligible at 1064 nm under typical conditions (two-way aerosol transmittance below 200 m deviating from unity by less than ~3% even for moderate aerosol loads; Li et al., 2021). Nevertheless, we acknowledge that this assumption may become less valid during intense aerosol episodes, particularly at Granada during Saharan dust intrusions or severe pollution events, where near-surface aerosol optical depths can be substantially larger.

To clarify the vertical domains used by each retrieval approach and to facilitate comparison between methods, we have added the following explanation to the manuscript:

Line 222: “The algorithm initializes the aerosol backscatter coefficient βa(z0) at the lowest reliable measurement height (z_0 = 250 m a.g.l.) and assumes that aerosol transmittance near the surface is approximately unity at infrared wavelengths.”

Line 224: Consequently, the initial aerosol backscatter is taken directly from the measured attenuated backscatter (Eq. 5). Although the CHM15k reaches full overlap only at approximately 800 m, the empirical temperature-dependent225 overlap correction described by (Hervo et al., 2016) substantially reduces artefacts above 250 m a.g.l. This correction is applied within the E-Profile processing chain and establishes 250 m a.g.l. the effective lower boundary for reliable aerosol retrievals in this study, consistent with the exclusion criteria applied throughout the analysis (Sects. 3.3 and4.2).

Line 402: “For the profile comparisons presented in this study, it is important to consider the different vertical domains used by the retrieval approaches (backward, forward, and GRASPpac). The three retrieval approaches do not share identical vertical boundaries. Both the forward and backward methods share the same lower boundary (z0 = 250 m a.g.l.) and the same reference-altitude search window of 1500–5000 m a.g.l. The backward method selects zref dynamically for each profile, whereas the forward method sets ztop to the upper boundary of this same window. GRASPpac uses a fixed vertical domain between 250 and 7000 m a.g.l. (Román et al., 2018), however, for the statistical comparison presented in this study profiles were evaluated between 250 and 5000 m a.g.l, since a number of GRASPpac retrievals did not extend above approximately 4500–5000 m a.g.l..”

Line 303: “*In this study, AODs processed by the CÆLIS system (González et al., 2020) were combined with cloud-screened, Level 2 ceilometer profiles*”, I'm confused, were radiances used too? Also provide wl set, because sun-photometer was stated to cover 340-1640 nm range in line 125. It is not clear which channels were used.

We thank the referee for this comment. The original text was indeed ambiguous regarding the input measurements used in the $GRASP_{pac}$ retrieval. We have revised the sentence to explicitly state that both spectral AOD and sky radiance measurements at 440, 675, 870, and 1020 nm were used as inputs, consistent with the standard AERONET inversion configuration (Román et al., 2018), together with cloud-screened ceilometer range-corrected signal profiles at 1064 nm.

To clarify this point, the following text has been added to the manuscript:

Line 414: “In this study, spectral AOD and sky radiance measurements at 440, 675, 870 and 1020 nm processed by the CÆLIS system (González et al., 2020) were combined with cloud-screened, ceilometer range-corrected signal profiles at 1064 nm as inputs to the GRASPpac retrieval (Román et al., 2018). This wavelength subset follows the standard AERONET inversion configuration. GRASPpac is not a standalone code but an inversion strategy implemented within the general GRASP framework (Dubovik et al., 2021); it is accessed through the official GRASP distribution (https://www.grasp-open.com) and configured via the GRASP settings file to execute the corresponding strategy, as documented in Román et al. (2018). This synergistic approach is particularly valuable for operational networks lacking advanced multi-wavelength research lidars. GRASPpac retrievals used in this work were generated automatically in near real time by the CAECENET system (Herrero del Barrio et al., 2024)

Section 3.4 How did backscatter became extinction? Description here is a bit messy, please revise. What SAs were used, and why? I mean we can select different value and get different aod, please discuss sensitivity to that selection. Were S_a estimated by AERONET in the vicinity of observation checked? How fare are they from 50? It would be great and very interesting to see such a comparison.

We thank the referee for this valuable comment, and we agree that the original description did not sufficiently explain how extinction profiles were derived from the retrieved backscatter profiles. To improve clarity, we revised the manuscript to explicitly describe the backscatter-to-extinction conversion.

Line 451: “After the retrieval and filtering steps, aerosol extinction profiles were obtained by converting the retrieved aerosol backscatter coefficients using a fixed lidar ratio of 50 sr. This value was selected as a representative estimate for mixed continental aerosol conditions at 1064 nm and is consistent with previous ceilometer studies (Brunamonti et al., 2021). The processing was performed using the A-Profiles package."

With this correction, we believe that now it is clear that a fixed aerosol lidar ratio of 50sr was used to transform backscatter into extinction and that this value was not derived from AERONET observations.

Regarding the sensitivity to lidar ratio, we agree that different lidar-ratio assumptions can lead to substantially different extinction profiles and AOD values. For this reason, in the revised manuscript we have included a dedicated sensitivity analysis (Sect. 4.3.2), in which AOD retrievals are evaluated using fixed lidar-ratio values of 30, 50, and 70 sr, as well as a daily variable estimate derived from IFS–CAMS aerosol composition forecasts. This analysis quantifies the impact of the lidar-ratio assumption on the retrieved AOD in terms of mean deviation, standard deviation, and RMSE.

Lines 333-338: "*High-load conditions were further categorised according to the AE as a proxy for dominant particle size. Days with AE ≥ 1.2 were classified as fine-mode dominated, typically associated with anthropogenic or biomass-burning aerosols (Reid et al., 2005), whereas days with AE ≤ 0.8 were classified as coarse-mode dominated, generally linked to mineral dust. Days with intermediate AE values (0.8 <AE <1.2) were excluded from this classification, as they represent mixed aerosol conditions without a clearly dominant particle mode.*" Please provide typical S_a values for these 3 groups, I have serious concerns that 50 Sr will fit them all.

We thank the referee for this important remark. We fully agree that a single lidar ratio of 50 sr is not universally optimal across all aerosol regimes, and we would like to clarify both the reasoning behind this choice and its implications for each aerosol category.

Typical lidar ratios reported in the literature differ substantially between aerosol types: fine-mode dominated aerosols (AE ≥ 1.2), associated with urban/industrial pollution or biomass burning, typically exhibit lidar ratios in the range of 50–70 sr, while coarse-mode dominated aerosols (AE ≤ 0.8), linked to mineral dust, tend toward lower values of 40–55 sr (e.g., Müller et al., 2007; Groß et al., 2011). Mixed conditions (0.8 < AE < 1.2) are excluded from the aerosol-type classification precisely because their optical properties cannot be attributed to a single dominant mode, making any single lidar ratio assumption particularly uncertain; this exclusion is therefore also methodologically motivated, not only pragmatic.

The choice of 50 sr reflects a compromise representative of background continental aerosol conditions and is consistent with the operational framework of the E-Profile network, where real-time processing without ancillary AERONET constraints is required. We do not claim that 50 sr is optimal for each individual case, but rather that it constitutes a reasonable central estimate for continuous, unsupervised operation.

To explicitly assess the impact of this assumption, the revised manuscript now includes a dedicated lidar-ratio sensitivity analysis (Sect. 4.3.2), in which retrievals using fixed values of 30, 50, and 70 sr, together with a composition-dependent estimate derived from IFS–CAMS forecasts (SaIFS), are evaluated against AERONET observations. This analysis demonstrates that the optimal lidar ratio is indeed aerosol-regime dependent and confirms the referee's concern that a single value cannot perfectly represent all aerosol types. Nevertheless, it also shows that 50 sr provides a reasonable compromise across the range of conditions sampled in this study, particularly for operational applications.

Line 354: "*altitudes above 5 km were not considered"*, but they were included into the GRASPpac retrieval. Please discuss how avoiding 2km of the profile may affect the comparison, notably total AOD.

We thank the referee for this comment. The exclusion of altitudes above 5 km a.g.l. from the comparison is not expected to significantly affect the results. As indicated in line 380 of the revised manuscript:

Line 383: "The comparison analysis was restricted to altitudes between 0.25 and 5 km above ground level. Heights below 0.25 km were excluded to avoid the incomplete-overlap region of the ceilometer, while altitudes above 5 km were excluded due to the reduced signal-to-noise ratio of the ceilometer measurement".

Although GRASPpac retrievals are performed over a larger vertical domain (250–7000 m a.g.l.), the statistical comparison presented in this study was restricted to the common altitude range between 0.25 and 5 km a.g.l. in order to ensure consistency between the retrieval approaches.

Aerosol concentrations above 5 km a.g.l. are generally negligible under typical conditions at our sites. Raman lidar climatologies at Mediterranean sites show that the bulk of aerosol extinction is confined below 4–5 km, with only occasional elevated layers from long-range dust transport reaching above this altitude (Amiridis et al., 2005). Furthermore, the ceilometer signal-to-noise ratio decreases substantially at these altitudes, such that the integrated signal becomes increasingly dominated by measurement noise rather than aerosol backscatter. Restricting the analysis to the lower 5 km therefore represents a compromise between maximising aerosol coverage and maintaining retrieval reliability.

We acknowledge, however, that under specific atmospheric conditions such as long-range transport of dust or biomass-burning plumes at elevated altitudes aerosol layers may extend above 5 km a.g.l. In these situations, the adopted upper integration limit could lead to a slight underestimation of the total column aerosol optical depth relative to a full-column retrieval. Nevertheless, such cases represent a small fraction of the dataset and are not considered representative of the prevailing aerosol conditions at

Granada, Payerne, and Lindenberg, where most of the aerosol load is confined to the lower troposphere. Consequently, the contribution from altitudes above 5 km is expected to remain comparatively small and should not materially affect the conclusions of this study.

Line 391: "*This case represents a well-documented episode of long-range Saharan dust transport"* please provide references.

We thank and appreciate this comment because this sentence lacked the corresponding citation to the statement for that the following references are added in line 525: "(Sousa et al., 2019; Córdoba-Jabonero et al., 2021; Sicard et al., 2022"

Line 433:" lidar *signal and can typically be retrieved with good accuracy under standard conditions". Define "good accuracy" and "standard conditions."* The whole paragraph is not fully correct, elastic lidar signal depends both on backscatter and extinction, theres' no way to separate them, assume another lidar ratio and backscatter retrieval will change too, this should be properly discussed.

We thank the referee for the comment. We agree that the original paragraph was not sufficiently precise and could incorrectly suggest that backscatter can be retrieved independently of the lidar-ratio assumption. We have therefore rewritten this section to clarify both the role of the lidar ratio in the inversion and the meaning of the atmospheric conditions under which the retrieval is expected to perform reliably.

Specifically, we removed the qualitative expressions "good accuracy" and "standard conditions" and replaced them with a more rigorous discussion of the sensitivity of backscatter and extinction retrievals to the assumed lidar ratio.

Line 592: " Although the elastic lidar signal depends simultaneously on both backscatter and extinction, and neither can be retrieved independently from a single-wavelength elastic lidar without assuming a lidar ratio (Sa) (Ansmann and Müller 2005), whereas Raman lidar systems can retrieve both quantities independently, the two quantities respond differently to uncertainties in that assumption. The backscatter coefficient is generally less sensitive to Sa uncertainties than the extinction coefficient, particularly at near-infrared wavelengths under cloud-free conditions with adequate signal-to-noise ratio. This reduced sensitivity arises because, in the forward iterative method, the backscatter retrieval depends on Sa through the accumulated two-way aerosol transmittance term, which remains close to unity within the boundary layer for typical aerosol loads at 1064 nm, thereby limiting the propagation of Sa errors into the backscatter solution (Li et al., 2021). Extinction profiles, by contrast, are derived by directly scaling the retrieved backscatter coefficient by the assumed Sa and therefore inherit both the backscatter retrieval uncertainty and the full proportional uncertainty associated with that assumption."

Line 435: *"outliers, only data within the 10th–90th percentile range were retained."* What data was retained, profile as a whole or specific altitudes? Please specify. Absolutely not clear what percentiles authors are talking about, were there multiple retrievals/profiles?

We thank the referee for this comment. We agree that in a multi-instrument comparison study such as this one, the criteria used to exclude extreme values are critical for ensuring the robustness and reproducibility of the results, and the original manuscript did not describe this procedure with sufficient clarity. For this, we rewrote the paragraph:

Line 607: "Deviation profiles were computed between ceilometer retrievals and $GRASP_{pac}$ for each simultaneous measurement. To reduce the influence of extreme deviations arising from occasional retrieval artefacts, residual cloud contamination, or episodic aerosol events not representative of the general dataset a symmetric percentile filter was applied, retaining only data points within the 10th-90th percentile range of the full distribution of differences. This filter excludes the 10% largest positive and 10% largest negative deviations, removing a total of 20% of the dataset while preserving the central bulk of the distribution on which the statistical analysis is based. The mean deviation and standard deviation were subsequently calculated for each altitude bin, following the aerosol scenarios described in Sect 3.4

Section 4.2: please, provide at least brief description of datasets, i.e. date ranges, number of profiles and how these were selected.

We thank the referee for the suggestion. We agree that a description of the dataset prior the results enrich the discussion. For that purpose, we added:

Line 586: "For this evaluation, the comparison covers the 2019–2020 period across the three European stations. A total of 2064 simultaneous ceilometer–GRASPpac profile pairs were available at Granada, 1524 at Payerne, and 1000 at Lindenberg. These profiles were selected based on the concurrent availability of Level 2 ceilometer attenuated backscatter and GRASPpac retrievals, under cloud-free conditions and within a 30-minute temporal averaging window.

Section 4.2.1: how the error margin was estimated? Is it variation, is it error propagation? Please, provide explanations.

We thank this commentary and we agree that the stats used in the discussion to make the evaluation must be clarified better. However, we added the explanation in the sect 4.2 because this is used on both comparisons Backscatter and Extinction.

The uncertainty of each retrieval method relative to GRASPpac is characterised statistically using the mean deviation, standard deviation, and root mean square error (RMSE) computed from the percentile-filtered dataset. The mean deviation quantifies the average bias of each retrieval method with respect to GRASPpac, the standard deviation reflects the dispersion of individual differences around that mean and is therefore indicative of the consistency of each retrieval approach, and the RMSE provides a combined measure of both bias and dispersion. These statistics are computed independently for each altitude bin and for each aerosol scenario defined in Sect. 3.4

Line 604: "For each altitude bin, the mean deviation represents the average difference between the ceilometer retrievals and the GRASPpac reference, while the standard deviation quantifies the dispersion of the individual differences around that mean and therefore reflects the consistency of each retrieval approach. These statistics are computed from the filtered dataset."

Line 454: "*relative dispersion below 35 %"*, why relative dispersion is provided/discussed only for this case, please provide them for all other cases too.

We thank the referee for this comment. We would like to clarify that relative dispersion values for all cases, sites and aerosol regimes are already provided comprehensively in Tables 3 and 4, which report mean deviation and standard deviation with relative deviation in percentage for both backscatter and extinction comparisons against GRASPpac respectively. The values mentioned across the text, including the one at Line 454 that prompted this comment, are illustrative examples to support the narrative and the complete set can be found in the corresponding table entries.

Line 489: *"site-specific microphysical differences."* Please discuss these conditions, it is not clear for reader why stations are analysed separately. Provide some explanations why they have particularly different aerosol conditions, and why these sites where chosen out of many.

We thank this suggestion because one of the strong points of the manuscript is that the evaluation is carried out in very different places of the continent. To clarify this, we added a more detailed explanation in the section 2.1 Stations and datasets.

Line 87: "These sites were selected to represent contrasting aerosol regimes across Europe, enabling a comprehensive evaluation of retrieval performance under a wide range of aerosol and meteorological conditions. Granada is characterised by urban aerosol conditions frequently influenced by Saharan dust intrusions, Payerne by clean fine-mode dominated background conditions, and Lindenberg by mixed continental aerosol typical of central Europe."

Line 99: "These influences result in a highly variable aerosol regime, dominated by coarse-mode mineral dust during Saharan outbreak events in spring and summer, and by fine-mode anthropogenic aerosol from urban and regional sources during winter."

Line 104: "The aerosol regime at Payerne is predominantly fine-mode dominated, associated with regional pollution and long-range transport of anthropogenic aerosol

(Dada et al., 2025) with occasional coarse-mode events linked to Saharan dust outbreaks (Collaud Coen et al., 2004)."

Line 111: "The aerosol conditions at Lindenberg are representative of central European continental background, with a mixture of fine-mode anthropogenic aerosol from regional sources and occasional long-range transport events (Ansmann et al., 2002). Its location in the North European Plain results in a variable air-mass history, with aerosol properties influenced by both marine and continental origins (Wandinger et al., 2004)."

Line 543: "*Comparison against GRASPpac reveals smaller dispersion for backscatter than extinction coefficients (Tables 2 and 3)."* Is it about relative or absolute dispersions?

We thank the referee for this clarification request. The statement refers to relative differences. We have added "relative" before "dispersion" in line 543 to make this explicit:

Line 764: "Comparison against GRASPpac reveals smaller relative dispersion for backscatter than extinction coefficients (Tables 2 and 3)."

In general if theres so much discussion about S_A, why not to compare them too, GRASP pac should provide some. Ideally use the S_a from GRASP pac for both forward and backward methods.

We thank the referee for this suggestion. While we agree that a direct comparison of lidar ratios would be of scientific interest, the aim of this study is to evaluate the three retrieval methods in their operational configuration, as they would be applied within the E-Profile network without ancillary constraints. Using the GRASPpac-derived lidar ratio as input for the forward and backward methods would improve the agreement by construction but would not reflect the actual operational conditions under which these methods are deployed. For this reason, we prefer to maintain the fixed lidar ratio assumption and focus the evaluation on the performance achievable in a fully automatic, unsupervised processing chain. We acknowledge this as a limitation of the study and note it explicitly in the revised manuscript.

Line 764: "Although the GRASPpac-retrieved lidar ratio could in principle be prescribed as input to the backward and forward inversions, we deliberately retain the fixed operational value of Sa = 50 sr for both methods. Adopting the GRASPpac lidar ratio would improve the agreement by construction, but it would not represent the fully automatic, unsupervised E-Profile processing chain that this study sets out to evaluate; the impact of the fixed-ratio assumption is instead quantified through the sensitivity analysis in Sect. 4.3.2."

Section 4.2: in general, why cases are split into such specific groups: all, high load, and high fine/coarse. It is evident that klett will struggle in these cases, provide same stats for low loads of fine/coarse too.

We thank the referee for this insightful comment. The division into aerosol load categories was motivated by the following reasoning. During the initial evaluation of the full dataset, we observed that low aerosol load cases showed very small mean deviations and dispersion across all three retrieval methods, with behaviour closely resembling that of the all-data statistics. This is physically expected: under low aerosol load conditions, the backscatter signal is dominated by molecular scattering, and the sensitivity of the elastic retrieval to the assumed lidar ratio is reduced, as the aerosol transmittance term remains close to unity and errors in S_a propagate weakly into the retrieved profiles.

For high aerosol load cases, by contrast, the choice of lidar ratio has a much stronger impact on the retrieved profiles, and the aerosol type as proxied by the Ångström Exponent determines the appropriate lidar ratio value. Since the lidar ratio is itself size- and composition-dependent, separating high-load cases into fine- and coarse-mode dominated regimes allows us to assess whether the fixed lidar ratio assumption of 50 sr introduces systematic biases that differ between aerosol types. This is particularly relevant at Granada, where Saharan dust events (coarse mode, lower lidar ratio) and urban/pollution episodes (fine mode, higher lidar ratio) represent physically distinct retrieval challenges.

Following the referee's suggestion, we have computed the same statistics for the low-load fine- and coarse-mode regimes. The fine-mode low-load cases exhibit the smallest mean deviations and dispersion of the entire dataset (forward mean deviation between −0.001 and +0.005, RMSE 0.007–0.009 across the three sites), confirming that when the aerosol type is close to the assumed 50 sr lidar ratio and the load is low, all retrievals converge and the agreement is strong. The coarse-mode low-load cases, by contrast, show an underestimation (forward mean deviation between −0.013 and −0.018, RMSE 2–3 times larger than the fine-mode case), although still about a factor of five below the corresponding high-load coarse-mode values. Since the load is identical in both sub-regimes, this contrast isolates the size-dependent component of the fixed-lidar-ratio deviation from its load-dependent amplification: the bias originates in the aerosol type and grows with load. This is why the fine/coarse separation is most diagnostically relevant under high load, while the low-load fine-mode regime serves as the reference for optimal retrieval agreement. Since the low-load cases closely track the all-data statistics and contribute little additional diagnostic information beyond establishing this reference behaviour, we restrict the main-text discussion to the all-data and high-load (fine/coarse) regimes.

Line 607: *"covering a wide range of aerosol types, loading conditions, and meteorological regimes."* None of these were discussed.

We appreciate the suggestion and we will be more rigorous on this sentence. For that we substituted the previous sentence with this one:

Line 962: "The analysis was conducted at three European sites with contrasting aerosol regimes: Granada, characterised by Saharan dust outbreaks and local anthropogenic pollution; Payerne, representative of clean background and fine-mode dominated conditions; and Lindenberg, representative of mixed continental aerosol typical of central Europe. Together, these sites encompass a wide range of aerosol loads and particle types, from coarse-mode dust-dominated to fine-mode anthropogenic conditions."

Line 630: *"At the column-integrated level, forward-derived AOD values show good agreement with AERONET observations at all sites, with mean biases close to zero and standard deviations comparable to the intrinsic uncertainty of AERONET AOD."* Would it be still the case if S_a assumption will be different? I seriously doubt.

We thank the referee for this pertinent comment. The referee is correct that the statement holds only under the assumed lidar ratio of 50. To address this, we have added a sensitivity analysis of the AOD retrievals to the assumed lidar ratio (Sect. 4.3.2), which directly quantifies the impact of different lidar ratio assumptions on the agreement with AERONET. The results show that the good agreement with near-zero mean deviations and low RMSE is specific to S_a 50 sr and SA~IFS, which perform consistently well across all sites and aerosol regimes. Under S_a 30 sr, AOD is systematically underestimated at all sites, while S_a 70 sr leads to systematic overestimation, with mean deviations reaching -0.126 and $+0.060$respectively under high aerosol load conditions at Granada. We have revised the sentence accordingly to make the dependence on the lidar ratio assumption explicit.

Line 994: "At the column-integrated level, forward-derived AOD values show good agreement with AERONET observations at all sites when a lidar ratio of 50 sr is assumed, with mean biases close to zero and standard deviations comparable to the intrinsic uncertainty of AERONET AOD. The sensitivity of this agreement to the assumed lidar ratio is discussed in Sect. 4.3.2, where it is shown that the choice of lidar ratio has a significant impact on retrieval accuracy, particularly under high aerosol load and coarse-mode dominated conditions."

Additional corrections:

Additional change. During the final revision we identified a typographical inconsistency in Eq. (1) and its accompanying text (former line 181). The backscatter term was written as the attenuated backscatter coefficient β^*while the equation simultaneously retained the explicit two-way transmission term $\exp\left(-2\int_0^z \alpha\, d\, z'\right)$. By definition, the attenuated backscatter coefficient already incorporates this transmission, $\beta^*(z) =$

$\beta(z)\exp\left(-2\int_0^z \alpha(z')\,dz'\right)$, so keeping both would apply the atmospheric transmission twice. Since the intention was to begin the calibration derivation from the lidar equation itself, we have removed the asterisk, so that Eq. (1) is now correctly expressed in terms of the total backscatter coefficient $\beta(z)$together with the explicit two-way transmission term.

Line 183: “The lidar equation can be expressed in terms of the backscatter coefficient, β (Eq. 1). In this formulation, P (z) denotes the received signal as a function of atmospheric optical properties and measurement range z, the exponential is the transmission term, while CL represents the system-dependent calibration constant accounting for instrumental characteristics.”

New Eq 1: $\frac{P_{(z)}\,z^2}{C_L} = \beta\,(z)\,\exp\left\{-2\int_0^z \alpha(z')\,d_{z'}\right\}$

Additional change. We found that first time term $GRASP_{pac}$ appears the pac part is not correctly defined.

Line 70: $GRASP_{pac}$ (Photometer and Ceilometer)

Additional change. To address the referees' comments, the first paragraphs of Sect. 3.5 (lines 438–456) was revised.

Line 438: “To evaluate the performance of the backward and forward ceilometer retrievals, aerosol optical depth (AOD) was derived from the retrieved aerosol profiles and compared with independent AERONET sun–sky photometer observations. This comparison provides an assessment of the ability of both retrieval approaches to reproduce the total aerosol optical load within the atmospheric column.”

Line 456:” AOD was then calculated by vertically integrating the extinction profiles from the surface to 5 km a.g.l. using the trapezoidal rule. Finally, the resulting ceilometer-derived AOD values were compared with AERONET Level 2.0 observations. “

Additional change. On our own revision, we found that one profile had been described as noise-limited above ~2 km, when in fact it does not converge towards the molecular signal at high altitude — indicating a calibration error rather than low signal-to-noise. The text (line 545) has been corrected to present this as a failure case of the method rather than a profile reliable below 2 km. The revised sentence now reads:

Line 556: "In this case the retrieved profile does not converge to the molecular signal, pointing to a calibration error rather than low signal-to-noise; the profile is thus erroneous and is presented as an example of the method failing when a reference altitude cannot be found."

Additional change. In the course of our revision we noted a number of statements that were not formulated with sufficient precision, as well as points requiring clarification, which have been revised as follows:

Line 634: **"**The forward method consistently shows smaller deviations and reduced variability than the backward approach, reducing the near-surface mean deviation by approximately a factor of three under coarse-mode-dominated conditions. This behaviour is consistent with the greater robustness of the forward retrieval in situations where the aerosol-free reference altitude required by the backward inversion is difficult to identify or may be affected by residual aerosol contamination."

Line 649: "The substantially larger deviations observed for the backward retrieval are consistent with the increased difficulty of identifying an aerosol-free reference region under high aerosol loading, which enhances the propagation of errors throughout the profile. The forward approach is less affected by this limitation, maintaining similar deviations at both altitude ranges and closer agreement with the GRASPpac reference under dust-rich and low-signal conditions (Fig. 5b)."

Line 731: "Comparison of extinction profiles retrieved by the backward and forward methods against $GRASP_{pac}$"

Line 795: "Overall, the comparison with $GRASP_{pac}$ indicates that the forward iterative retrieval generally provides a more stable and vertically consistent representation of aerosol extinction than the traditional backward approach. The differences between the two methods are most pronounced under high aerosol load and coarse-mode-dominated conditions, where the forward retrieval exhibits smaller deviations and lower variability. These results support the use of the forward approach for operational aerosol profiling within ceilometer network"

Line 801: "The column-level consistency of the ceilometer retrievals is assessed in two steps. First, AOD derived from the backward and forward retrievals is compared with AERONET observations using the standard fixed lidar ratio of 50 sr across the different aerosol regimes (Sect. 4.3.1). Second, the sensitivity of these AOD estimates to the assumed lidar ratio is evaluated by testing alternative fixed lidar ratios and a composition-based variable estimate (Sect. 4.3.2)"

Line 972: "whereas the forward method relies on an independent calibration derived from nighttime observations."

Line 974: "shows lower dispersion in the retrieved backscatter and extinction profiles"

Line 984: "lower variability and smaller dispersion"

Line 992: "many of the main features captured by more complex multi-instrument solutions."

Line 1008: “In summary, this study indicates that the forward iterative retrieval generally provides a more vertically consistent representation of aerosol backscatter and extinction than the traditional Klett–Fernald backward inversion, with smaller deviations and lower dispersion relative to the reference datasets considered here. While extinction retrievals remain sensitive to the assumed lidar ratio, backscatter profiles are less affected by this assumption and therefore exhibit lower uncertainty. Together, these results support the use of operational ceilometers, when carefully calibrated and quality-controlled, as practical and sustainable sources of vertically resolved aerosol information, contributing meaningfully to long-term air quality monitoring and climate-related atmospheric research.”