Manuscript acp-2020-294 – Author's replies to reviewers

We thank gratefully the editor and two anonymous referees for careful reading and comments. Below are the referee's comments in **black***, and replies from the authors in* ***blue****. Please note that page and line numbers given below refer to the revised version of the manuscript without tracked-changes.*

Anonymous Referee #1

The paper presents an interesting comparison between the aerosol backscatter coefficients measured by two different lidar systems (a sophisticated multiwavelength lidar with elastic and Raman channels, and a ceilometer) and those obtained by a balloon-borne instrument performing in-situ measurements. The latter is taken as reference to validate the backscatter profiles provided by the lidars.
The paper is well written and describes a sound methodology that, besides providing the validation mentioned in the paper's title, can be useful for similar verifications at other sites.
I think the paper is worth publishing (although given its scope, focusing on techniques and methods rather than on atmospheric processes, perhaps the sister journal Atmospheric Measurement Techniques would provide a more suitable forum).
The authors may wish to consider the following remarks that in my view would improve the manuscript.

Main remarks

1. In the paper it is implied that the COBALD instrument is taken as the reference against which the lidar-derived backscatter coefficients are validated, on grounds that an in-situ instrument inherently provides "higher precision and signal-to-noise ratio compared to remote sensing measurements" (line 31, page 2). For this reason, I miss a more detailed description of the instrument specifications, namely, systematic (bias) and random (noise) error.

The accuracy and precision of COBALD BSR were estimated by Vernier et al. (2015) (Section 2.1) as 5 % and 1 %, respectively, at upper tropospheric conditions (which can be regarded as an upper limit here, due to the higher absolute signal measured in the lower troposphere). This previously missing information is now included in the manuscript (Page 5, Lines 7-8).

2. The above remark is somewhat linked to a seemingly lack of explanation for the mean deviations between the lidar-derived backscatter coefficients and those provided by the COBALD instruments in the PBL (+ 6% for RALMO and +13% for the ceilometer, below 2 km, (lines 23-24, page 1)). Is this just a random effect resulting from the limited dataset? Might negative differences be obtained for other datasets? Is this an effect resulting from the wavelength conversion and the FOV correction discussed respectively in sections 3.2 and 3.3? Does it come from other reasons (see next point).

Indeed, the uncertainties related with the wavelength conversion and the FOV correction, as well as spatial and temporal variability effects, contribute to the mean and standard deviations of our statistical comparison. A more detailed and quantitative discussion of the uncertainties associated with these effects is now included in the new Section 4.3.
In particular, concerning the wavelength conversion: "From Equation 1 we can derive that an error of 0.2 in AE, which is a conservative estimate considering the small difference between the wavelengths that are compared, results in an error of 5 % in β_{aer} for the 355-to-455 nm conversion, and 2.5 % for the 1064-to-940 nm conversion" (Page 15, Lines 9-11).
Concerning the FOV correction: "From Figure 2b, we can estimate an uncertainty of up to ± 20% in βaer for the PBL (for both 455 and 940 nm), due to variability of the correction factors in the range of AE = 0.8-1.5, which is not resolved by the parameterization of the FOV correction factors in AE-space." (Page 15, Lines 12-14).

Finally, about spatial and temporal variability: ”These effects can lead to large discrepancies over small altitude layers, as in the case of strong vertical gradients in βaer (e.g., top of boundary layer: Figure 4a-d), as well as potentially over larger altitude regions, due to the horizontal gradient of the βaer field around the station (e.g., in the case of the strongly outlying profiles of the statistical comparison: see Figures 5, 8). The lack of information on the aerosol size distribution and the high spatial and temporal variability of atmospheric aerosols prevent an accurate quantification of these artifacts, which inevitably affect the standard deviations of our statistical comparison" (Page 15, Lines 16-21).
Hence, we conclude: "Despite of these limitations, the comparison of individual profiles (Section 4.1) shows that both RALMO and CHM15K are able to achieve an excellent agreement with COBALD measurements, including the correct representation of fine and complex structures in the βaer vertical profiles (Figures 3-4, S2-S3). In particular, the case study of 12 July 2018 (Figures 4a-d) shows differences between the lidars and COBALD which are smaller than expected statistical uncertainty associated with the remote sensing measurements alone (10-15 %, see Section 2.1). This suggests that, under optimal conditions (such as, no wind shear, uniform βaer field, mono-modal aerosol size distribution), the deviations between the two lidars and COBALD are typically smaller than the average σ of our statistical comparison.” (Page 15, Lines 28 - Page 16, Line 1).

3. Related to the previous point, the authors put forward the possibility (lines 1-2 of page 13) that the 15% positive bias below 2.5 km (line 30 of page 12; by the way, shouldn’t it be rather 13%, cf. line 24 of page 1 and fig. 9f) be related to “minor unsolved geometric overlap issues in the ceilometer’s retrieval algorithm, or (more likely) related to the assumption of a constant lidar ratio (50 sr)”. Could this also be the cause for the (smaller (+6.5%)) positive bias in the RALMO vs. COBALD comparison below 2 km? The influence of an assumed lidar ratio could checked with relative ease. Have the authors done it?

The uncertainty associated with the assumption of a constant lidar ratio (50 sr) in the retrieval algorithm for CHM15K, which adds up to the uncertainties discussed in the previous comment, is now also discussed more in detail in Section 4.3.
"Using a similar ceilometer (Jenoptik CHM15kx), Wiegner and Geiss (2012) estimate that an error of ± 10 sr in lidar ratio leads to an error in βaer smaller than 2 % in the boundary layer. Ackermann (1998) shows that 50 ± 10 sr represents well the expected range of variability of the lidar ratio of continental aerosol in the infrared spectrum, for all RH conditions between 0-90 %. Therefore, this uncertainty conceivably plays a minor role compared to the effects discussed above.” (Page 15, Lines 24-27).
For RALMO, no a-priori assumption on lidar ratio is required for the backscatter retrieval (see Section 2.1). Therefore, this uncertainty only affects the CHM15K comparison. The sentence at Page 13, Lines 1-2 of the original manuscript was removed within the revision of the statistical comparison (see answer to comment 5 below).

4. It would also be advisable that the authors provide some indication on the statistical error in the measurements (estimated error bars), not only for the COBALD sondes, but also for the lidar-derived backscatter coefficients. That would help clarifying how much of the standard deviation found in the comparisons presented is due to the uncertainty of the measurements of each instrument, which must set a lower limit to that standard deviation affected as well, as the authors point out, by the differences between the atmosphere volumes measured by the sonde and by the lidars.

Information on the estimated mean statistical uncertainty of RALMO and CHM15K is now provided in Section 2.1. For RALMO “The mean statistical uncertainties associated with the retrieval of βaer at 355 nm from Raman inversion techniques are typically estimated as 15 % in the PBL (Pappalardo et al., 2004)." (Page 4, Lines 2-4). For CHM15K: “using a similar ceilometer (CHM15kx by Jenoptik, Germany), Wiegner and Geiss (2012) report a relative error of 10 % on βaer at 1064 nm retrieved by this method" (Page 4, Lines 14-15). The observed standard deviations of the statistical comparison, as well as the comparison of individual profiles, are now discussed in the context of these statistical uncertainties in Section 4.3 and Conclusions (see in particular Page 15 Line 30 - Page 16 Line 1, Page 17 Lines 4-7).

5. I would suggest restricting the use of relative differences in the comparisons of the backscatter coefficients to the layers with a medium to high aerosol content. I think using it in zones of low aerosol content or in the free troposphere is misleading, as small absolute differences will yield large figures when they are divided by a very small backscatter coefficient, which in turn is probably driven by statistical noise. In this respect, the authors should probably follow the criteria stated in section 4D of their reference Matthais et al., 2004. If the authors want to highlight something important coming out from these comparisons in terms of relative errors at those altitude ranges with little or no aerosol content, they should be more explicit.

Following this comment and one remark of Reviewer #2, the statistical comparison (including mainly Section 4 and Figures 5-10) has been strongly revised and improved. In particular, the following changes were made:

- Low and medium-high aerosol content measurements are separated according to an empirical threshold, and shown by different colors in Figures 5, 7, 8, 10.
- Mean and standard deviations profiles (Figures 5, 8) are calculated from medium-high aerosol content only, and the use of relative differences in the discussion is mainly restricted to medium-high aerosol content measurements.
- The probability density functions (Figures 6, 9) are calculated for medium-high aerosol content and all data separately.
- The Pearson correlation coefficient is now also calculated for medium-high aerosol content data, allowing to quantitatively evaluate the linearity of the correlation between the lidar and COBALD measurements.
- A new Section 3.3 (Page 9 Line 10 - Page 10 Line 17) was added to the manuscript, where we introduce the sorting of the data according to aerosol content, and formally define the compared quantities (Equations 3-6).

This approach allows to meaningfully quantify the deviations of medium-high aerosol content measurements at all altitudes, which is the main focus of this study, while at the same time not fully neglecting low aerosol content measurement at higher altitudes, which are important because a good agreement in the free troposphere ensures that all profiles are well calibrated (see Page 9, Lines 24-26).
We believe the revised statistical comparison is strongly improved compared to the previous version of the paper, both in terms of scientific content and clarity, and we thank the reviewer for this comment.

Minor issues

1. The statistical analyses of the comparisons of RALMO and the CHM15K ceilometer against COBALD are divided in figs. 6 and 9 in altitude zones, the first one being 0-2 km asl. However, in the text, when discussing the comparisons at the lowermost altitudes, the authors often use the 2.5 km limit (e.g. lines 25 and 28 in page 10, etc.). It would be easier for the reader to follow the discussions if the text and the figures would use the same limits.

Done. In the revised statistical comparison, we reduced the number of vertical intervals from three (0.8-2 km, 2-4 km, 4-6 km) to two (0.8-3 km, 3-6 km), and the limit value of z = 3 km is now used consistently throughout the manuscript.

2. Page 2, line 26: “the atmospheric number density”. I would suggest “the atmospheric number density of molecules”. Note that the pressure-to-temperature ratio would also do.

Done (page 2 line 26).

3. While the minimum height of measurements for the CHM15K instrument is indirectly given through its full overlap range, this information seems to be missing for the RALMO system. Even though the Raman technique employed in RALMO to derive the aerosol backscatter coefficient allows compensating incomplete overlap effects to some extent, I think RALMO’s minimum usable altitude should be stated for completeness.

Indeed, thanks to its Raman retrieval technique, the RALMO backscatter is unaffected by incomplete overlap issues. Nevertheless, the signal-to-noise is typically very low in the first 200 m above the station, hence this can be considered as a 'minimum usable altitude'. This information is now included in the manuscript (Page 3, Lines 28-30).

4. Note a possible inconsistence in the full overlap range of CHM15K. In line 4 of page 4 it is stated as “800 m above the station”, while in lines 2-3 of page 6 is it is said that “We only select measurements from ≈ 300 m above the ground station in order to minimize the effect a possible incomplete overlap of the lidar systems in the lower part of the profiles”. Left aside the already mentioned fact that no overlap information seems to be given for RALMO, do the authors use CHM15K data obtained below its full overlap range? This deserves some calrification.

Below the full overlap altitude of CHM15K (800 m above station), the backscatter profiles are corrected for the incomplete overlap between the incoming beam and the receiver’s field of view, as described in Hervo et al. (2016). This (previously missing) information is now included in the manuscript (Page 4, Line 9).
The rejection of all measurements below 300 m above the station is aimed to avoid the region of maximum incomplete overlap of CHM15K, as well as to avoid the region of low signal-to-noise ratio of RALMO at low altitudes (see answer to previous comment). This statement is now clarified in the manuscript (Page 6, Lines 23-25).

5. I suggest that a logarithmic scale be used for the horizontal axes in figs. 2a, b, and c. 6. In the caption of fig. 2 it should be stated that the graph in panel a is obtained for 3 -3 N=10 cm .Currently that information is found only in the main text. 7. I suggest trying to find a symbol (and give a name) for the ratio / β β aer mol . Otherwise the authors have to use the rather awkward notation BSR-1 to refer to that ratio and the text may even fall in ambiguities, for example in line 3 of page 8 when they say “Figure 2a shows the simulated aerosol backscatter ratio (i.e. BSR − 1)”. But the backscatter ratio is BSR, not BSR-1. Perhaps just / β β aer mol would do.

Done. The X-axis of Figure 2a-2b-2c (mode radius) was changed to logarithmic scale, and the information on number concentration ($N = 10^3$ cm^{-3}) was added to the caption of Figure 2. The notation 'BSR-1' was replaced with the more compact 'β_{aer}/β_{mol}' throughout Section 3.2 and in Figure 2.

Anonymous Referee #2

The manuscript shows an intercomparison between a Raman lidar, a commercial ceilometer and an optical in-situ sonde takes as reference. This contribution doesn't represent a substantial contribution to scientific progress and I agree that it is more indicated for Atmospheric Measurement Techniques than ACP.

Major Issues:

The FOV correction factor has been computed with a well defined aerosol distribution. How the results change if a more likely bi-modal distribution is used instead? Or changing the distribution width and/or the refractive index? As it is implemented, the correction is depending on a particular type of aerosol.

The single-lognormal size distribution assumed for the calculation of the FOV correction is to be interpreted as an *average* size distribution of boundary layer aerosols, rather than that of a well-defined population. This assumption has the advantage that the correction factors can be described as functions of a single parameter (R_m), which can be constrained through the observed AE, as discussed in Section 3.2. Assuming a more complex (e.g. bi-modal) size distribution, as well as relaxing one or more parameters of the size distribution (width, refractive index), would inevitably result in increased number of degrees of freedom of the correction factors, for which insufficient observational constraints are available. Therefore, the correction would be practically not applicable. In particular, for a bi-modal size distribution the correction factors would also depend on the number concentration (N) ratio of the two modes, whereas the correction for a single-lognormal distribution are independent of N.
Since furthermore, previous studies show that a mono-modal distribution represents well the average size distribution of continental aerosols in the Northern mid-latitudes (e.g., Watson-Perris et al., 2019), we believe the assumption of a single lognormal size distribution is justified. This choice is now motivated more carefully in the manuscript (Page 8, Lines 12-15), and a quantitative discussion of the uncertainty introduced by the FOV correction in the statistical comparison is now also provided in Section 4.3 (Page 15, Lines 12-15).

The statistical intercomparison is unclear and counter-intuitive. I would suggest to the authors to use the Pearson Cross Correlation coefficient paired with the Root Mean Square Error, on the whole atmospheric profile and at different altitude ranges, e.g. into the PBL, free troposphere...

Following this comment and one remark of Reviewer #2, the statistical comparison (including mainly Section 4 and Figures 5-10) has been strongly revised and improved. In particular, the following changes were made:

- Low and medium-high aerosol content measurements are separated according to an empirical threshold, and shown by different colors in Figures 5, 7, 8, 10.
- Mean and standard deviations profiles (Figures 5, 8) are calculated from medium-high aerosol content only, and the use of relative differences in the discussion is mainly restricted to medium-high aerosol content measurements.
- The probability density functions (Figures 6, 9) are calculated for medium-high aerosol content and all data separately.
- The Pearson correlation coefficient is now also calculated for medium-high aerosol content data, allowing to quantitatively evaluate the linearity of the correlation between the lidars and COBALD measurements.
- A new Section 3.3 (Page 9 Line 10 - Page 10 Line 17) was added to the manuscript, where we introduce the sorting of the data according to aerosol content, and formally define the compared quantities (Equations 3-6).

This approach allows to meaningfully quantify the deviations of medium-high aerosol content measurements at all altitudes, while at the same time not fully neglecting low aerosol content measurement in the free troposphere, and maintaining a clear and systematic structure. The Pearson correlation coefficient, evaluated for medium-high aerosol content data in three altitude intervals (0.8-3 km, 3-6 km, 0.8-6 km), adds a useful further insight to the characterization of the performances of the lidar instruments with respect to COBALD.
Given the similarity between the definitions of RSME and standard deviation (see Equation 5 in the manuscript), and the fact that standard deviation is predominantly used in previous aerosol backscatter intercomparison in the context of EARLINET (e.g., Matthais et al., 2004; Pappalardo et al., 2004), here we decide to keep standard deviation as a measure of variability. This choice also aims to avoid redundancy of information, which might turn out confusing for a reader (note that already three statistical parameters are used in this paper: mean deviation, standard deviation, and Pearson correlation coefficient, each of them quantified for different datasets and different altitude regions).
We believe the revised statistical comparison is strongly improved compared to the previous version of the paper, both in terms of scientific content and clarity, and we thank the reviewer for this comment.

Figure 5 and 8 are unnecessary.

Unfortunately, we are unable to understand properly this comment. Figures 5 and 8 show the vertical profiles of the RALMO – COBALD (Figure 5) and CHM15K – COBALD (Figure 8) difference in aerosol backscatter coefficient, which are a fundamental component of our statistical comparison. As discussed in the previous answer, Figures 5 and 8 were strongly improved in the revised version of the manuscript.

CHM15K inversion should be explained more in detail, as the molecular signal can be very low at 1064nm.

Using a similar ceilometer (CHM15kx by Jenoptik, Germany), Wiegner and Geiss (2012) show that a Klett inversion algorithm can provide accurate aerosol backscatter profiles, despite the low molecular backscatter at infrared wavelengths and the low signal-to-noise ratio in the free troposphere. This reference is now included in the manuscript (Page 4, Lines 11-13).

Optical measurements are strongly affected by water vapor absorption at 940nm. Usually, the ceilometers at this wavelength use a radiative transfer computation to correct the profile. What about COBALD sonde? In the text it is not mentioned.

The effect of water vapor absorption on COBALD BSR at 940 nm is negligible, due to the short optical path length of this instrument (≈ 10 m). COBALD uses two LEDs emitting 250 mW optical power and a detector with FOV of 6°. A good overlap between the emitted light beams and the detector FOV is established at ≈ 0.5 m distance from the sonde, and the backscattered signal from a distant layer decreases by the inverse distance squared. This means that, assuming uniform scattering conditions, the signal contribution at 10 m distance from the sonde falls below 0.25 % compared to that in the vicinity of the detector. In this sense, COBALD is considered as an *in-situ* instrument in this comparison.
Therefore, the uncertainty contribution associated with water vapor absorption at 940 nm on COBALD BSR is to be considered as included in the estimate of 5 % accuracy and 1 % precision provided by Vernier et al. (2015), now also given in the manuscript (Page 5, Lines 7-8).

Moreover, some equations are needed to better explain lines 25-28 (Pag. 4).

A new equation was added to define BSR (Equation 1), and the explanation has been broken in multiple sentences for more clarity (Page 5, lines 3-8).

It would be more interesting to intercompare the two instruments vs. COBALD for different meteorological conditions and aerosol loading.

We agree only in part with this comment. The performances of the analyzed instruments, both the lidars and COBALD, are to our best knowledge independent of meteorological conditions, hence we do not expect any systematic bias associated with meteorological parameters (such as temperature or specific humidity). Furthermore, our comparison already avoids all in-cloud and high RH measurements. This is done partly through the selection of the dataset (discussed in Section 2.3), where profiles for which a precise calibration of the lidar signal cannot be achieved are rejected (most frequently due to fog or low clouds), and subsequently through the rejection of all data points with RH > 90% in the statistical comparison (according to the radiosonde measurements) (see Page 7, Lines 1-3). Finally, the limited available dataset (17 soundings for RALMO vs. COBALD, 31 for CHM15K vs. COBALD), together with the irregular periodicity of COBALD soundings (see Table S1 in Supplementary material), and the strong day-to-day variability of boundary layer aerosols, do not allow for any robust investigation of seasonal patterns or other systematic weather-related behaviors.

On the other hand, our comparison does already take different aerosol loadings into account. This is done through the scatter plots of RALMO vs. COBALD β_{aer} (Figure 7) and CHM15K vs. COBALD β_{aer} (Figure 10), where their deviations are displayed and discussed as functions of the absolute COBALD β_{aer} signal, i.e. the aerosol loading (see in particular Page 13 Lines 15-18, Page 14 Lines 26-30). In the revised version of the manuscirpt, this aspect is further emphasized by the distinction between low and medium-high aerosol content measurements, now shown in different colors in Figures 5, 7, 8, 10, and evaluated separately in the statistical comparison (see Sections 3.3, 4.2, 4.3, and replies to comments above).