

## Reply to the comments by the anonymous Referee #1

The authors gratefully acknowledge the reviewer's effort in improving the quality of the manuscript. Below, a point-to-point response is provided to all of the reviewers' major comments. The remaining minor comments have been all of them fixed in the new version of the manuscript.

### General Comments

The paper "Intercomparison of aerosol measurements performed with multi-wavelength Raman lidars, automatic lidars and ceilometers in the frame of INTERACT- II campaign" reports the results of a campaign using a variety of instruments to measure aerosol in cloud-free or clear-sky conditions. While the authors report interesting results, I think that they could make the analysis more rigorous and motivate the work more clearly. I have made recommendations below.

### Specific Comments

1. I recommend that the authors provide general motivation in the introduction for this study. Why does anyone need to measure atmospheric aerosols using these types of instruments? Why is this intercomparison needed? Is it to help design better networks for measuring pollution, for example? I would like to understand this and to make sure the audience understands how the intercomparison gives us important and useful information. Can the authors say anything specific about the aerosols that were measured (type or other properties) during the campaign?

**Though the introduction largely describes the current state-of-the-art for the use of ceilometers for aerosol profiling in the troposphere and their big potential to improve the current baseline aerosol observing capabilities at the global scale as low-cost and low-maintenance instruments, to meet the reviewer's request, in the new version of the manuscript, the two following sections have been added to the introduction at line 69: "Given the role commercial lidars and ceilometers may cover as a low-cost and low-maintenance baseline component of the aerosol non-satellite observing system at the global scale, several intercomparison experiments must be designed to assess the performances of commercial systems with respect to advanced multi-wavelength lidars and to ensure comparability between different instruments, measurements and retrieval techniques. Recommendation outcome from these experiments can also strongly support the design of current and future aerosol observing networks for measuring aerosols and pollution."**

2. Please can the authors explain, again in a general way, which of the instruments is expected to measure aerosols (of a given type) most accurately and why. For example, can you give a general sense of where (in the atmospheric column) the instruments are expected to give the best results? And why? Perhaps it would be helpful to touch on differences in wavelength here as well as other differences in hardware or firmware? I realise that none of the instruments gives us "truth", but can the authors give the reader a sense of the accuracy expected? Thus, when the differences are reported, the readers immediately understand which of the instruments is believed to be closer to the true observed quantity.

I suggest these two points in order to give the reader a better sense of why these particular instruments are important to study (as I think that they are) and to make a stronger case for why the intercomparison analysis in this paper matters to the community.

The following paragraph has been added at the lines 58-70 of the new version of the manuscript to meet the reviewer's request of clarification for the reader.:

*“With respect to the past when lidars were strictly research instruments, many modern automated lidars are available on the commercial market and can now contribute efficiently to continuous monitoring atmospheric aerosol. Automatic lidars have very different features from models equipped with diode-pumped laser or solid-state laser emitting in the UV at 355 nm or in the visible spectrum at 532 nm. Only multi-wavelength lidars emits wavelengths in the near infrared at 1064 nm. Typically, the higher is the energy emitted per laser pulse (In the order of a few  $\mu\text{J}$  to mJ) the more demanding will be the required maintenance and costs. In analogy, higher is the energy emitted per laser pulse the better will be the lidar signal to noise ratio and the lower will be the random uncertainty affecting the estimation of the estimated aerosol properties. A ceilometer generally differentiates from a one-wavelength automatic lidar because it emits a single wavelength in the near infrared between 900 and 1100 nm to avoid strong Rayleigh scattering, the pulse repetition rate is on the order of a few kilohertz, and the pulse energy of the laser is in the order of a few  $\mu\text{J}$  to allow eye-safe operations, continuously and unattendedly operations. UV and visible automatic lidars can typically cover the whole tropospheric range while ceilometer, depending on the model, may cover the boundary layer only or detect aerosol features also in the free troposphere.”*

We want to clarify that the author intentionally didn't provide any details on the system precision and accuracy because these may strongly change from an instrument to another and they prefer to provide an extensive characterization of the measurements in the section where the intercomparison with the CIAO lidars is discussed.

3. In section 5 MUSA is referred to as the reference signal in the full overlap region. Why is MUSA the reference? Is it expected to be the highest standard of measurement to which we want to compute the ceilometer observations?

In the new version of the manuscript the following paragraph has been added to explain why MUSA is considered the “reference” system for the intercomparison campaign: *MUSA is routinely tested with respect to several systematic quality-assurance tests developed in order to harmonize the lidar measurements, to set up quality standards, and to improve the lidar data evaluation (Pappalardo et al., 2014). MUSA signals are also routinely evaluated using the Rayleigh fit test, and signal-to-noise analysis described in Baars et al. (2016). Additionally, the telecover test (Freudenthaler, 2008) is performed regularly and especially after transportation of the system. The system is aligned using a CCD camera to reduce the effect of misalignment between the telescope and laser axis, being MUSA a bistatic lidar. Finally, the multi-wavelength detection capability enables to so called "3+2" lidar data analysis which, taking advantage of the simultaneous retrieval of lidar extensive (aerosol extinction at 355 nm and 532 nm; backscattering coefficients at 355 nm, 532 nm and 1064 nm) and intensive properties (lidar ratios at 355 nm and 532 nm and color ratios) at different wavelengths permits to check the physical consistency of the retrieved aerosol properties.*

The authors also clarified when describing PEARL lidar that “PEARL has been extensively intercompared with MUSA to have a redundant aerosol profiling capability at CIAO.”.

4. Please define the "fractional difference". For example, in section 4 Paragraph 5, "average fractional difference" is not defined and later in the paragraph (line 327) an "average difference" is increasing. Are these the same metric? The authors need to define clearly the measure or measures of difference applied to the results.

**The concept of fractional difference is now explained at lines 268 and 269 using the following sentence: "Fractional difference is defined as the difference between CIAO lidar and MiniMPL RCS values normalized to CIAO lidar RCS".**

5. There are a few places where the authors discuss "random uncertainty" (section 4 for example in line 322). Please could the authors define how they determine the random uncertainty? Also, if there are some statistical tests being performed to assess differences then please state which tests are being used. For example, is there a null hypothesis of random white noise?

**In the new version of the manuscript a reference has been added at the corresponding lines to clarify the processing applied to the CIAO lidar signals and the retrieval of the corresponding uncertainties.**

**Random uncertainty is the contribution to the total uncertainty budget typically named by lidar experts as "statistical error." According to the GUM and metrology, the term "error" is less appropriate than uncertainty when an estimation of the error is provided. According the GUM (Guide to the Expression of Uncertainty in Measurement):**

***"Whereas the exact values of the contributions to the error of a result of a measurement are unknown and unknowable, the uncertainties associated with the random and systematic effects that give rise to the error can be evaluated. But, even if the evaluated uncertainties are small, there is still no guarantee that the error in the measurement result is small; for in the determination of a correction or in the assessment of incomplete knowledge, a systematic effect may have been overlooked because it is unrecognized. Thus, the uncertainty of a result of a measurement is not necessarily an indication of the likelihood that the measurement result is near the value of the measurand; it is simply an estimate of the likelihood of nearness to the best value that is consistent with presently available knowledge.***

***Uncertainty of measurement is thus an expression of the fact that, for a given measurand and a given result of measurement of it, there is not one value but an infinite number of values dispersed about the result that are consistent with all of the observations and data and one's knowledge of the physical world, and that with varying degrees of credibility can be attributed to the measurand."***

**The random uncertainty for raw lidar signals is evaluated as the standard deviation of the Poisson distribution of counts (square root of the counts), because the backscattered radiation is acquired in photon-counting mode and a Poisson distribution is assumed for the detected signals. For CIAO lidars, the raw signals are pre-processed to apply instrumental corrections and, optionally, a vertical smoothing or temporal averaging. This stage is commonly known as "pre-processing" of raw signals. The pre-processed signals, with time and vertical resolutions depending, respectively, on temporal and vertical integration performed by the pre-processing module, are the input of the second part of the processing algorithm, known as "processing" of the pre-processed signals, providing the profiles of aerosol optical properties. These profiles have a time sampling which is the integration time used in pre-processing stage and effective vertical resolution depending on the vertical smoothing performed in pre-processing and processing modules.**

The random or statistical uncertainties of pre-processed signals are calculated starting from random uncertainties of raw lidar signals, by using the standard formula of statistical uncertainty propagation at each step of the pre-processing stage. Random uncertainties in the aerosol extinction or backscattering profiles are calculated starting from random uncertainties of pre-processed lidar signals, by using the Monte Carlo simulation for all applied signal handling procedures in the processing stage.

For the MiniMPL, though this is a polarized elastic backscatter lidar operating only at 532 nm, the applied processing follows a similar logic in the pre-processing of the lidar signals.

6. At the end of the technical corrections, I have placed a number of comments on the figures which need to be addressed.

**The authors fully addressed all of the technical corrections recommended by the Reviewer #2 and provide below comments to the most relevant.**

Technical Corrections

Title: Please change "frame" to "framework". Text:

1. Many acronyms are undefined in the main body of the paper. To aid the reader, please explicitly define the following: CNR-IMAA, EARLINET, FOV, FWHM, GRUAN, RAOB, HYSPLIT and APD in line 149 (is it Avalanche Photo Diode?) used before line 161 Avalanche Photo Detector are these the same "APD"?,

**The listed acronyms have been defined in the new version of the manuscript.**

2. Please put units on the RCS. I believe that the authors are using "arbitrary units" (a.u.) throughout. Is this correct? Can a.u. be placed next to all the measurements please?

**Ok. A.u. has been reported next to all the measurements only when absolute values of RCS are reported.**

Line 22 Is average difference a root mean squared difference? Absolute difference? Or something else?

**This is "average fractional difference" and has been appropriately modified throughout the manuscript.**

Line 29 Rewrite to something more like: "Some tests performed during this campaign using the CHM15k ceilometer made it clear that the CHM15k historical dataset (2010- 2016) available at CIAO should be reviewed in order to evaluate the potential effect of...."

**Modified accordingly.**

L239 Please could you briefly (in a sentence) say why the assumption of < 1% is a good one? I can see there is a reference, but a quick explanation would be helpful, if a brief one is possible.

**Upon the basis of additional calculations, the authors have modified the text as follows “The uncertainty contribution for the spectral dependence of  $\beta'$  and, therefore, of the aerosol backscattering coefficient and of molecular and aerosol extinction coefficients has been estimated within a few percent.”.**

L 352 Is the output profile from Raman PEARL lidar? If so, is it interpolated to the same resolution as the RCS from which instrument?

**Figure 7 embeds the measurements performed with both MUSA and PEARL, all of them interpolated at the same output resolution.**

L377-378 "...because MUSA is considered the reference signal only in the full overlap...." Has this been stated before? Has MUSA been the reference all along?

**Please see the authors reply to the reviewer’s general comments.**

L394 I would suggest the wording should be changed to "Dark current measurements or profiles", not just "dark currents"

**The authors have discussed the use terminology and it seems that the community working with ceilometers prefers this terminology. This is the reason why kept it also because we believe this is not confusing.**

Figures:

1. All sub-panels within all figures should be labelled with letters a,b,c, etc. 2. In the text and captions all of the sub-panels in the figures should be referred to using the figure number and letter together. Please do not use left/right, top/bottom. The letters make the text concise and precise. For example, caption for figure 8 should read more like: "Panel a shows attenuated backscatter retrieved from ... Similarly, panel b shows the same comparison but for 01 December ..."

Figure 3, 4, 6, 8, 9, 12 have a red line (or red bar) labelled "Lidar" but MUSA is in the caption. Lidar is not specific enough. Please make the legend consistent and more precise. Is it MUSA Lidar? In contrast, for example, Figure 10 has a red line called PEARL which is also a Raman lidar like MUSA.

Figure 4. Caption is confusing. Can authors please explain what they mean by "using NOAA HYSPLIT model started at the three levels from the ground the top layer observed by MUSA and MiniMPL lidars"? Are we talking about model levels? What is "the top layer observed"?

Figure 5. Should read "Blue line is the same as the black line but..." Also, captions usually put the line colour or line style in parentheses like this: "Profiles of the average fractional difference (black line)..."

Fig 9 End of caption: " Panel b shows the attenuated backscatter vertical profiles taken using the MUSA/PEARL lidar which operates at wavelength 1064 nm during the same time period as was used to create the average profiles in panel a."

Fig 10 Change to "Comparison between" not among. Also the line colour is "green" not "dark". This is the line with the dark current measurement subtracted away but the line is green.

Fig 11 Change "calculated on" to "calculated for"

Figure 14: What time does each square represent? Can't be 30 s resolution?! There are 7 years on the x-axis. How were the laser pulses averaged?

All the Figures have been modified according to the reviewers' suggestions. For the last question about Figure 14 each square represents the number of pulses emitted per hour.